Descriptive Analysis in Sensory Evaluation

Editors
Sarah E. Kemp, Joanne Hort
and Tracey Hollowood

WILEY Blackwell

Descriptive Analysis in Sensory Evaluation

A series of books on selected topics in the field of Sensory Evaluation

The first book in the Sensory Evaluation series is *Sensory Evaluation: A Practical Handbook*, published in May 2009. It focuses on the practical aspects of sensory testing, presented in a simple, 'how to' style for use by industry and academia as a step-by-step guide to carrying out a basic range of sensory tests. In-depth coverage was deliberately kept to a minimum. Subsequent books in the series cover selected topics in sensory evaluation. They are intended to give theoretical background, more complex techniques and in-depth discussion on application of sensory evaluation that were not covered in the *Practical Handbook*. However, they will seek to maintain the practical approach of the handbook and chapters will include a clear case study with sufficient detail to enable practitioners to carry out the techniques presented.

Descriptive Analysis in Sensory Evaluation

EDITED BY

Sarah E. Kemp

Consultant and formerly Head of Global Sensory and Consumer Guidance, Cadbury Schweppes, UK

Joanne Hort

Massey Institute of Food Science and Technology
Massey University
New Zealand

Tracey Hollowood

Sensory Dimensions Ltd.
Nottingham, UK

WILEY Blackwell

This edition first published 2018

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Kemp, Sarah E., editor. | Hort, Joanne, editor. | Hollowood, Tracey, editor.
Title: Descriptive analysis in sensory evaluation / [edited] by Sarah E. Kemp, Joanne Hort, Tracey Hollowood.
Description: Hoboken, NJ : John Wiley & Sons, 2018. | Includes bibliographical references and index. |
Identifiers: LCCN 2017028435 (print) | LCCN 2017043923 (ebook) | ISBN 9781118991671 (pdf) | ISBN 9781118991664 (epub) | ISBN 9780470671399 (cloth)
Subjects: LCSH: Sensory evaluation.
Classification: LCC TA418.5 (ebook) | LCC TA418.5 .D47 2018 (print) | DDC 660.072–dc23
LC record available at https://lccn.loc.gov/2017028435

Cover Design: Wiley
Cover Image: © nepstock/Gettyimages

Set in 10/12pt Meridien by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To George, Elizabeth, George and William
To Mike, Holly and Socks
To Campbell, Emma and Lara
In memory of Pieter Punter

Contents

Editor Biographies

Sarah E. Kemp, BSc (Hons), PhD, CSci, FIFST, RSensSci, is a chartered sensory and consumer science professional with more than 30 years of experience in academia and industry. Dr Kemp gained a BSc in Food Technology in 1986 and a PhD in Taste Chemistry in 1989 from the Food Science and Technology Department at the University of Reading, UK. In 1990, she did a postdoctoral research fellowship in sensory science at the Monell Chemical Senses Center in Philadelphia, USA. Dr Kemp has held many positions in industry, including Manager of Sensory Psychology (US) and Director of European Consumer and Marketing Research (France) in the Fragrance Division at Givaudan, Product Area Leader and Sensory Science Leader in Foods Consumer Science at Unilever Research, Colworth, UK, Head of Global Sensory and Consumer Guidance at Cadbury Schweppes, UK, and Director of Sensory and Consumer Services at Reading Scientific Services Limited, UK. Dr Kemp has also set up and run her own consultancy service and catering company. She has written numerous scientific articles in the field of sensory evaluation, has provided sensory training courses, including lecturing on the European Masters Course in Food Science, and has worked on bodies developing standards in sensory evaluation, including the British Standards Institution and ASTM International. She is a fellow of the Institute of Food Science and Technology and a founder member, past Chair and examiner for the IFST's Sensory Science Group, as well as being a member of other professional sensory societies. Her other activities include Governor of East Kent College, UK.

Tracey Hollowood, BSc (Hons), PhD, MIFST, is currently Managing Director of Sensory and Consumer Research for Sensory Dimensions (Nottingham) Ltd in the UK. She has over 25 years' experience in academia and industry; she worked at Nottingham University for 10 years during which time she achieved her doctorate investigating perceptual taste-texture-aroma interactions. She established the UK's first Postgraduate Certificate in Sensory Science and designed and managed the University's prestigious Sensory Science Centre. Her research focused on psychophysical studies, interactions in sensory modalities and fundamental method development. She has over 30 peer-reviewed publications, has run numerous workshops and delivered oral presentations to many international audiences including at the Pangborn Sensory Science Symposia 2015 in Gothenburg. She has participated in the organization of seven international symposia, including the International Symposium of Taste 2000 and Pangborn 2005 in Harrogate.

Tracey is a previous chair of the Institute of Food Science and Technology (IFST) Midland branch and the Professional Food Sensory Group (PFSG), now the Sensory Science Group (SSG).

Joanne Hort, BEd (Hons), PhD, CSci, FIFST, RSensSci, is the Fonterra-Riddet Chair of Consumer and Sensory Science at Massey University in New Zealand following on from her various academic roles, latterly SABMiller Chair of Sensory Science at the University of Nottingham. Initially, Professor Hort studied food technology and began her career in teaching. However, she returned to university to receive her doctorate concerning the modelling of the sensory attributes of cheese from analytical and instrumental measures in 1998. As a lecturer at Sheffield Hallam University, she carried out sensory consultancy for local industry, developed a sensory programme at undergraduate level and oversaw the installation of new sensory facilities before being appointed as Lecturer in Sensory Science at the University of Nottingham in 2002. There she established the University of Nottingham Sensory Science Centre, which is internationally renowned for both its sensory training and research into flavour perception. She obtained her Chair in 2013 and her multidisciplinary approach combining analytical, brain imaging and sensory techniques provides rich insight into multisensory interactions, individual variation and temporal changes in flavour perception, and the emotional response to sensory properties, leading to over 90 publications. Joanne sits on the editorial board for Food Quality and Preference and Chemosensory Perception. She is a Fellow of the Institute of Food Science and Technology. She is a founder member and past Chair of the European Sensory Science Society and a founder member, past Chair and examiner for the IFST's Sensory Science Group.

List of Contributors

Hervé Abdi
School of Brain and Behavioral Sciences, The University of Texas at Dallas, Richardson, TX, USA

Christel Adam
Givaudan, Argenteuil, France

Lucía Antúnez
Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Pando, Uruguay

Gastón Ares
Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Pando, Uruguay

Cindy Beeren
London, UK

Rebecca N. Bleibaum
Dragonfly SCI, Inc., Santa Rosa, CA, USA

Wender L.P. Bredie
Department of Food Science, University of Copenhagen, Frederiksberg, Denmark

Dominic Buck
Sensometrica, London, UK

Edgar Chambers IV
Center for Sensory Analysis and Consumer Behavior, Kansas State University, Manhattan, KS, USA

Sylvie Chollet
Institut Charles Viollette, Laboratoire régional de recherche en agroalimentaire et biotechnologies, Equipe Qualité et Sécurité des Aliments – QSA, Lille, France

Anne Churchill
Givaudan UK Ltd, Ashford, UK

Gail Vance Civille
Sensory Spectrum, New Providence, NJ, USA

Graham Cleaver (Retired)
Unilever Research & Development, Wirral, UK

Sophie Davodeau
Givaudan, Naarden,The Netherlands

Christian Dehlholm
Department of Food Science, University of Copenhagen, Frederiksberg, Denmark

Luis de Saldamando
Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Pando, Uruguay

Clare Dus
Sensory Spectrum, New Providence, NJ, USA

Margaret A. Everitt (Retired)
Margaret Everitt Ltd, Cheltenham, UK

Ana Giménez
Sensometrics & Consumer Science, Instituto Polo Tecnológico de Pando, Facultad de Química, Universidad de la República, Pando, Uruguay

Ruth Greenaway
Sensory Dimensions, Nottingham, UK

Anne Hasted
Qi Statistics Ltd, Reading, UK

Hildegarde Heymann
Department of Viticulture and Enology, University of California, Davis, CA, USA

Tracey Hollowood
Sensory Dimensions, Nottingham, UK

Joanne Hort
Massey Institute of Food Science and Technology, Massey University, Palmerston North, New Zealand

Sylvie Issanchou
Centre des Sciences du Goût et de l'Alimentation, AgroSup Dijon, CNRS, INRA, Université de Bourgogne Franche-Comté, Dijon, France.

Patricia A. Keane (Retired Principal)
Arthur D. Little, Inc., Cambridge, MA, USA

Sarah E. Kemp
Consultant and formerly Head of Global Sensory and Consumer Guidance, Cadbury Schweppes, UK

Annlyse Retiveau Krogmann
Sensory Spectrum, New Providence, NJ, USA

Jing Liu
Department of Food Science, University of Copenhagen, Frederiksberg, Denmark

Alejandra M. Muñoz
IRIS: International Resources for Insights and Solutions, LLC, Mountainside, NJ, USA

Michael Nestrud
Ocean Spray Cranberries, Lakeville-Middleboro, MA, USA

May Ng
PepsiCo R&D, Leicester, UK

Pieter H. Punter (Deceased)
OP&P Product Research, Utrecht, The Netherlands

Carol Raithatha
Carol Raithatha Ltd, Norwich, UK

Lauren Rogers
Consultant, Stoke-on-Trent, UK

Joel L. Sidel
Joel L. Sidel Consulting, Los Altos, CA, USA

Lee Stapleton
Sensory Spectrum, New Providence, NJ, USA

K.W. Clara Tao
QRS-Tragon LLC, Redwood City, CA, USA

Amy Trail
Sensory Spectrum, New Providence, NJ, USA

Dominique Valentin
AgroSup Dijon, Université de Bourgogne, Dijon, France

Preface to the Series

Sensory evaluation is a scientific discipline used to evoke, measure, analyse and interpret responses to products perceived through the senses of sight, smell, touch, taste and hearing (Anonymous, 1975). It is used to reveal insights into the way in which sensory properties drive consumer acceptance and behaviour, and to design products that best deliver what consumers want. It is also used at a more fundamental level to provide a wider understanding of the mechanisms involved in sensory perception and consumer behaviour.

Sensory evaluation emerged as a field in the 1940s. It began as simple 'taste testing' typically used in the food industry for judging the quality of products such as tea, cheese, beer, and so on. From the 1950s to the 1970s, it evolved into a series of techniques to objectively and reliably measure sensory properties of products, and was typically used to service quality assurance and product development. Through the 1980s and 1990s, the use of computers for data collection and statistical analysis increased the speed and sophistication of the field, so that sensory, consumer and physicochemical data could be combined to design products that delivered to consumer needs.

Today, sensory evaluation is a sophisticated, decision-making tool that is used in partnership with marketing, research and development and quality assessment and control throughout the product lifecycle to enable consumer-led product design and decision making. Its application has spread from the food industry to many others, such as personal care, household care, cosmetic, flavours, fragrances and even the automotive industry. Although it is already widely used by major companies in the developed market, its use continues to grow in emerging markets, smaller companies and new product categories, as sensory evaluation is increasingly recognised as a necessary tool for competitive advantage.

The field of sensory evaluation will continue to evolve and it is expected that faster, more flexible and more sophisticated techniques will be developed. Social networking tools are transforming the way research is undertaken, enabling direct and real-time engagement with consumers. The use of sensory evaluation by marketing departments will continue to grow, particularly in leveraging the link between product sensory properties and emotional benefits for use in branding and advertising. Advances in other fields, such as genomics, brain imaging, and instrumental analysis, will be coupled with sensory evaluation to provide a greater understanding of perception.

Owing to the rapid growth and sophistication of the field of sensory evaluation in recent years, it is no longer possible to give anything but a brief overview of individual topics in a single general sensory science textbook. The trend is towards more specialised sensory books that focus on one specific topic, and to date, these have been produced in an ad-hoc fashion by different authors/editors. Many areas remain uncovered.

We, the editors, wanted to share our passion for sensory evaluation by producing a comprehensive series of detailed books on individual topics in sensory evaluation. We are enthusiastic devotees of sensory evaluation, who are excited to act as editors to promote sensory science. Between us, we have over 70 years of industrial and academic experience in sensory science, covering food, household and personal care products in manufacturing, food service, consultancy and provision of sensory analysis services at local, regional and global levels. We have published and presented widely in the field; taught workshops, short courses and lecture series; and acted as reviewers, research supervisors, thesis advisors, project managers and examiners. We have been active in many sensory-related professional bodies, including the Institute of Food Science and Technology Sensory Science Group, of which we are all past Chairs, the European Sensory Science Society, of which one of us is a past Chair, the Institute of Food Technologists, the British Standards Institute and ASTM International, to name but a few. As such, we are well placed to have a broad perspective of sensory evaluation, and pleased to be able to call on our network of sensory evaluation colleagues to collaborate with us.

The book series *Sensory Evaluation* covers the field of sensory evaluation at an advanced level and aims to:

- be a comprehensive, in-depth series on sensory evaluation
- cover traditional and cutting-edge techniques and applications in sensory evaluation using the world's foremost experts
- reach a broad audience of sensory scientists, practitioners and students by balancing theory, methodology and practical application
- reach industry practitioners by illustrating how sensory can be applied throughout the product life cycle, including development, manufacture, supply chain and marketing
- cover a broad range of product applications, including food, beverages, personal care and household products.

Our philosophy is to include cutting-edge theory and methodology, as well as illustrating the practical application of sensory evaluation. As sensory practitioners, we are always interested in how methods are actually carried out in the laboratory. Often, key details of the practicalities are omitted in journal papers and other scientific texts. We have encouraged authors to include such details in the hope that readers will be able to replicate methods themselves. The focus of sensory texts often tends to be food and beverage products assessed using olfaction and taste. We have asked authors to take a broad perspective to include non-food products and all the senses.

The book series is aimed at sensory professionals working in academia and industry, including sensory scientists, practitioners, trainers and students; and industry-based professionals in marketing, research and development and quality assurance/control, who need to understand sensory evaluation and how it can benefit them. The series is suitable as:

- reference texts for sensory scientists, from industry to academia
- teaching aids for senior staff with responsibility for training in an academic or industrial setting
- course books, some of which to be personally owned by students undertaking academic study or industrial training
- reference texts suitable across a broad range of industries; for example, food, beverages, personal care products, household products, flavours, fragrances.

The first book in the series, *Sensory Evaluation: A Practical Handbook* was published in May 2009 (Kemp et al. 2009). This book focuses on the practical aspects of sensory testing, presented in a simple, 'how to' style for use by industry and academia as a step-by- step guide to carrying out a basic range of sensory tests. In-depth coverage was deliberately kept to a minimum. Further books in the series cover the basic methodologies used in the field of sensory evaluation: discrimination testing, descriptive analysis, time-dependent measures of perception and consumer research. They give theoretical background, more complex techniques and in-depth discussion on application of sensory evaluation, whilst seeking to maintain the practical approach of the handbook. Chapters include clear case studies with sufficient detail to enable practitioners to carry out the techniques presented. Later books will cover a broad range of sensory topics, including applications and emerging trends.

The contributors we have selected are world-renowned scientists and leading experts in their field. Where possible, we have used originators of techniques. We have learned a lot from them as we have worked with them to shape each book. We wish to thank them for accepting our invitation to write chapters and for the time and effort they have put in to making their chapters useful and enjoyable for readers.

We would also like to thank our publisher, Wiley Blackwell, and particularly extend our thanks to David McDade, Andrew Harrison and their team for seeing the potential in this series and helping us bring it to fruition. We would also like to thank the anonymous reviewers of the series for their constructive comments.

We hope you will find the *Sensory Evaluation* book series both interesting and beneficial, and enjoy reading it as much as we have producing it.

Sarah E. Kemp
Joanne Hort
Tracey Hollowood

References

Anonymous (1975) *Minutes of Division Business Meeting*. Institute of Food Technologists – Sensory Evaluation Division, IFT, Chicago, IL.

Kemp, S., Hollowood, T. & Hort, J. (2009) *Sensory Evaluation: A Practical Handbook*. Oxford: Wiley-Blackwell.

Preface

Descriptive analysis is one of the cornerstone techniques in sensory evaluation. The aim of this book is to provide a comprehensive and up-to-date overview of the technique.

Descriptive analysis is covered in classic general sensory science texts, including Meilgaard et al. (2007), Lawless and Heymann (2010) and Stone et al. (2012). These have limited space to give to the topic, which makes it difficult to strike a balance between theory and practical application. To the editors' knowledge, there are four previous publications devoted to descriptive analysis. ASTM (1992) produced a manual that gives a brief comparison of different descriptive methodologies. Gacula (1997) is a textbook on descriptive analysis, and although it was a good source of information for its time, it is now a relatively old text, written prior to the introduction of newer methods. Delarue et al. (2014) and Varela and Ares (2014) are books that focus on newer methods.

The editors saw a need for a book devoted to descriptive analysis that would provide in-depth theoretical and practical coverage of traditional and recently developed descriptive techniques. The scope of this book includes history, theory, techniques and applications of descriptive analysis. It does not include time intensity descriptive techniques, which are covered in a separate book in the Sensory Evaluation series (Hort et al. 2017).

The book is structured in four sections. Section 1 is an introduction covering general topics in descriptive analysis, including panel training, panel monitoring and statistical analysis. Section 2 covers different techniques in descriptive analysis, ordered approximately according to historical development. Section 3 covers applications of descriptive analysis. Section 4 provides a summary that compares different methods.

Each chapter includes theory, psychological aspects, methodology, statistical analysis, applications, practical considerations, including hints/tips and dos/don'ts for carrying out methodology, case studies and examples, future developments and a reference list. The aim is to give a balance between theory and practice, with enough theory for readers to fully understand the background and underlying mechanisms of the technique, and in many instances enough detail to enable the reader to carry out the methodology.

Wherever possible, the authors invited to write chapters on particular techniques are the originators or early users of that technique and have extensive expertise and experience in its application. We wish to thank all authors for giving their time and effort to their chapter despite their busy schedules, and for their patience with the process. We would particularly like to thank Alejandra Muñoz for providing additional guidance.

We hope you find this book as interesting and beneficial to read as we did to produce.

Dr Sarah E. Kemp
Professor Joanne Hort
Dr Tracey Hollowood

References

ASTM (1992) *E-18 Manual on Descriptive Analysis Testing for Sensory Evaluation*. West Conshohocken: American Society of Testing and Materials.

Gacula, M.C.J. (1997) *Descriptive Sensory Anlaysis in Practice.* Washington, DC: Food and Nutrition Press.

Delarue, J., Lawlor, B. & Rogeaux, M. (2014) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research.* Cambridge: Woodhead Publishing.

Hort, J., Kemp, S.E. & Hollowood, T. (2017) *Time-Dependent Measures of Perception in Sensory Evaluation.* Oxford: Wiley-Blackwell.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evalution of Food: Principles and Practices*, 2nd edn. New York: New York.

Meilgaard, M.C., Civille, G.V. & Carr, B.T. (2007) *Sensory Evaluations Techniques*, 4th edn. Boca Raton: CRC Press.

Stone, H., Bleibaum, R.N. & Thomas, H.A. (2012) *Sensory Evaluation Practices*, 4th edn. London: Elsevier.

Varela, P. & Ares, G. (2014) *Novel Techniques in Sensory Characterization and Consumer Profiling.* Boca Raton: CRC Press.

SECTION I
Introduction

CHAPTER 1

Introduction to Descriptive Analysis

Sarah E. Kemp, May Ng, Tracey Hollowood and Joanne Hort

1.1 Introduction

Descriptive analysis is a method used to objectively describe the nature and magnitude of sensory characteristics. It was a pioneering development for its day, and represented a major step forward that gave sensory evaluation a scientific footing through the ability to produce objective, statistically reliable and statistically analysable data. Today, it remains a cornerstone method in sensory analysis.

A wide range of descriptive analysis techniques have been developed since its inception. Traditional descriptive techniques, such as profiling-based methods and quantitative descriptive analysis, involve a panel of trained assessors objectively measuring the quality and strength of the sensory attributes of samples. More recently, faster descriptive techniques, such as sorting, projective mapping and polarized sensory positioning, involve untrained consumers grouping samples based on holistic similarities and differences in sensory characteristics. Over the years, descriptive analysis has proved itself to be flexible and customizable, which has contributed to its usefulness and hence its longevity.

As descriptive analysis enables objective, comprehensive and informative sensory data to be obtained, it acts as a versatile source of product information in industry, government and research settings. Descriptive analysis was first applied to foods and beverages, but is now applied to a broad range of products including home, personal care, cars, environmental odours, plants, etc. It is used throughout the product lifecycle, including market mapping, product development, value optimization, and quality control and assurance. Descriptive analysis is particularly useful in product design, when sensory data are linked to consumer hedonic data and physico-chemical data produced using instrumental measures. This allows product developers and marketing professionals to understand and identify sensory drivers of product liking in order to design products with optimal liking. Sensory descriptive information can also be linked to other types of

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

consumer data to enhance brand elements, emotional benefits, functional benefits and marketing communication.

There are many general texts and reviews on descriptive analysis and the reader is directed to the following: ASTM (1992), Gacula (1997), Murray et al. (2001), Meilgaard et al. (2006), Kemp et al. (2009), Lawless and Heymann (2010a,b), Varela and Ares (2012, 2014), Stone et al. (2012) and Delarue et al. (2014).

1.2 Development of Descriptive Analysis

1.2.1 Evolution

Descriptive analysis grew from the need to assess products in a reliable fashion. Originally, product sensory quality relied on assessment by experts, such as brewers, wine tasters, tea tasters and cheese makers, who judged quality on key product attributes and made recommendations on how ingredients and process variables affected production and the finished product, which might often have a very fixed, invariable specification over a long period of time. The expert, sometimes called the 'golden tongue', was often a single person, who had product experience or had been trained by other experts. Businesses relied heavily on a few key individuals, which could be problematic if they left, particularly if they were the prime expert on the unique sensory characteristics of a company's product. Attributes were often important to the manufacturing process, rather than the consumer, and might comprise defects or complex terms that were difficult to understand. Attributes were often assessed using grading on quality scales that might be idiosyncratic to a company, an industry or a country. Indeed, experts could also be idiosyncratic and subjective in their judgements. Data often comprised a single value, which could not be interrogated statistically, making it difficult to compare scores in a meaningful way. In many cases, only the expert could interpret differences in scores between products.

As the market became more complex and fast-paced, with increasing numbers of ingredients, processing technologies, products, competition and consumer choice, the need arose for a more robust system for assessing product quality. The introduction of descriptive analysis moved away from a single expert to a trained panel of assessors, removing the reliance on a single person and making the data more reliable. Controls were introduced, such as experimentally verified scales, physical sensory references rather than descriptive words, consistent assessment methodology and thorough training. As sensory evaluation became recognized as a scientific discipline, good experimental design as used in other scientific areas was introduced, such as elimination of variability and bias, and use of experimental design and replication. This enabled the production of robust, objective data that could be analysed statistically. In a similar fashion, food production had moved from a craft to a science, and data produced from descriptive analysis now became available for food scientists and technologists to use in conjunction with physico-chemical instrumental measures to understand food quality in a science-based, rigorous manner.

The market continued to grow, and became increasingly international and global. Companies began to manufacture greater volumes, often at many national and international sites, and the rigorous nature of descriptive analysis now made it easier to compare data across studies and across panels, for example, to check that product quality was consistent across manufacturing sites. At this point, descriptive analysis was a key tool for quality assurance and control, and the sensory department was essentially providing a service based on routine testing. Traditional methodologies continued to be honed. In the US, several dominant descriptive analysis methods emerged driven by sensory agencies. In Europe, where the market for sensory agencies was more fragmented, the trend was towards customizing descriptive methodology to suit the needs of individual companies.

With globalization, the marketplace has evolved to be highly competitive. Consumers have become increasingly sophisticated and demanding, with a wide range of choices. To gain a competitive advantage, it is important to deliver consumers' needs, wants and desires. Product push has given way to consumer pull, and it is now consumers who are the ultimate judges of product nature and quality (Kemp 2013). The applications of descriptive analysis have evolved to become a key tool for use in product design and development, in order to interpret and deliver consumers' sensory requirements. New product development can be guided to create products based on consumer likes and dislikes. Descriptive data are now routinely combined with consumer data to determine sensory attributes that drive consumer liking, aided by the advances in technology outlined below that have enabled sophisticated, rapid statistical modelling and analysis. Physico-chemical and process data can also be combined in these models to enable manipulation of product characteristics to optimize consumer liking. Sensory attributes of key importance to the consumer can be comprehensively understood, and are now routinely used in quality control and assessment.

As the marketplace has become complex and sophisticated, so has the means of marketing products. There are many ways in which product sensory characteristics play a role in marketing, as described in section 1.4.3, including sensory pleasantness leading to repeat purchase, as an essential brand characteristic, as a functional benefit or indicator of a functional benefit, and as part of the brand/product experience, which is increasingly highlighting emotional aspects. Statistical modelling using descriptive data has been able to illuminate and design sensory characteristics linked to brand elements, functional benefits and emotional benefits. Hence, descriptive analysis is now an important tool for marketing and can be used across the product life cycle. As a result, the sensory department itself has now evolved to become a full partner to marketing and technical functions, rather than a service provider in the quality department.

As factors related to the commercial environment have influenced the evolution of descriptive analysis, and indeed sensory evaluation in general, so have advances in technology. Methods of data collection have changed considerably. In the early days, all data had to be collected using pen and paper, and then transcribed into raw data tables by hand. The chance of error was higher and data entry was usually double checked, further slowing progress. Preparing paper questionnaires was time-consuming, and could be complex given the experimental design. Transcribing data from a continuous line scale involved measuring the distance from the end of the scale to the assessment mark with a ruler, which was a daunting task made exponentially larger by the number of attributes, samples, assessors and replicates. The size and complexity of descriptive analysis studies were limited, as was the statistical analysis that was feasible.

The introduction of computers in the 1980s considerably speeded up operations. Initially, computers were expensive and one computer might be used in a conjunction with an optical reader to carry out data input and analysis. As computers became faster and cheaper, the process of descriptive analysis became increasingly more automated. Computers were introduced into sensory booths for direct data entry. Bigger studies, more complex experimental designs and faster, more comprehensive data analysis were possible. At the same time, computerized systems were developed to design, manage and run sensory testing, making descriptive analysis easier and more streamlined to perform.

Much more complex and sophisticated data analysis, such as multidimensional scaling (MDS) and generalized Procrustes analysis (GPA), became feasible and routine, leading to the symbiotic development of descriptive methods that relied on this analysis, such as free choice profiling, sorting and other techniques. This also enhanced the application of descriptive data, as complex statistical modelling linking descriptive data to consumer and physico-chemical, instrumental data became possible, using techniques such as preference mapping and response surface methodology (RSM). This enabled the sensory drivers of liking to be identified for consumer-led product development, so that today consumer-driven product design using this approach is the norm for larger companies with the available resources. Sophisticated graphics became possible, making it easier to illustrate results to lay audiences, and hence increase interest and use of descriptive analysis.

The introduction of wireless technology freed computers, so that they became portable, enabling descriptive testing to be carried out on the go in real-life environments. Technology has also become smaller and more robust, so that it can be used easily wherever and whenever necessary. For example, descriptive analysis of shower gels can now be carried out in consumers' home bathrooms using waterproof tablets in their showers, with data sent for analysis in real time. Mobile phone apps enable data to be collected conveniently as consumers go about their daily lives. The widespread use of the internet and social media has also had an impact, although care needs to be taken to ensure that the identity and location of the assessor has been verified. Virtual descriptive panels have been set up with group training carried out via web-based sessions, with references and products sent to consumers' homes. Central location testing still remains convenient, and advances in virtual reality environments have made it more realistic although this is not yet widespread.

In some ways, descriptive analysis has become a victim of its own success. It is now used routinely throughout the new product development cycle, as described above, but this cycle is becoming increasingly faster and shorter. Despite the gains in speed from computerization and other new technologies, traditional descriptive analysis can be perceived as slow to set up, to complete a study and to produce actionable results. Ever faster product launch cycles have lead to the development of more rapid methods for descriptive analysis, such as sorting and flash profiling, in which sensory characteristics for products are compared together rather than individually assessed. Some of these methods can be run with untrained assessors, eliminating what can be several months of set-up time. A study can be completed more rapidly, and although analysis can be complex, speed is on a par with modelling techniques used to link descriptive data with consumer and physico-chemical data. There may, however, be compromise of detail for speed.

Today, descriptive analysis remains a key sensory tool that is highly flexible, with the choice of many standard methods to suit a wide range of applications and the possibility of customization for specific applications. The history of the development of descriptive analysis methods is described in section 1.2.2.

1.2.2 History

1.2.2.1 To 1950s

The early history of descriptive analyses often relied upon 'golden tongue' experts, such as brew masters, wine tasters, perfumers, flavourists and others, to guide product development and quality assurance. It was possible for these experts to be reasonably successful when the marketplace was less competitive. From the 1910s to the 1950s, various score cards and sheets were developed by companies and government departments primarily for quality evaluation, and the need for accurate, reliable methods using the appropriate assessors and scales gradually became apparent (see Amerine et al. (1965) and Dehlholm (2012) for a review of early literature, and the latter for an overview of the history of descriptive methods to the present).

With the rapid introduction and proliferation of new products into the marketplace, a need for a formal means of describing food arose. Researchers at the Arthur D. Little laboratory were the first to take the ground-breaking step of developing a robust method called the *flavor profile method** (FPM) to meet this need (Cairncross & Sjostrum 1950). They demonstrated that it was possible for

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

trained assessors to produce actionable results without depending on individual experts and this was a key change in the philosophy of sensory science. The main features of the method involved analysing a product's perceived aroma, flavour and aftertaste characteristics, their intensities, order of appearance, aftertaste and overall impression using a panel of 4–6 assessors. However, one weakness of this method was that the data could not be statistically treated.

Several methods based on FPM have been developed. A step in FPM uses *consensus profiling*, in which a group of assessors work together to produce group intensity scores for attributes, and this is still used as a stand-alone method, although statistical analysis of the data is not possible (see Chapter 6). Other early derivations of the method include the *modified diagram method* (Cartwright & Kelly 1951) and the *dilution flavour profile* (Tilgner 1962a,b), although these have not been widely used. A later extension was *profile attribute analysis* (PAA) (Neilson et al. 1988), developed by Arthur D. Little, Inc., which involved the use of individual assessments of visual, tactile and auditory attributes on category/line scales and incorporated statistical analysis using ANOVA.

1.2.2.2 1960s

As there was a need to apply descriptive methods to food texture assessment, a new method called the *texture profile method* (TPM) was developed at the General Foods Technical Center by a team of researchers, under the leadership of Dr Alina Szczesniak in the 1960s (Brandt et al. 1963; Szczesniak 1963; Szczesniak et al. 1963). This method involved assessing the quality and intensity of a product's perceived texture and mouthfeel characteristics categorized into three groups: 'mechanical', 'geometric' and 'other' (alluding mostly to the fat and moisture content of foods). This technique used the 'order of appearance' principle from FPM and is conducted in order of first bite to complete mastication by a panel of 6–10 assessors, who must receive the same training in the principles of texture and TPM procedures. The type of scale used in TPM has expanded from a 13-point scale to category, line and magnitude estimation scales (Meilgaard et al. 2006). Similar to FPM, many reference products were not available to researchers outside the UK (Murray et al. 2001). Although data could not be statistically treated, the foundation of rheological principles upon which the method is built are still applicable. However, a few papers have suggested a solution to this by modifying TPM scales (Bourne et al. 1975; Hough et al. 1994). TPM has been applied to many specific product categories, including breakfast cereal, rice, whipped topping, cookies, meat, snack foods and many more (Lawless & Heymann 2010a).

1.2.2.3 1970s

In the mid-1970s, Tragon Corporation developed a method called *quantitative descriptive analysis* (QDA), later modified and registered under the name Tragon

QDA® (Stone et al. 1974). This method not only relied on sound sensory procedures but it was also fully amenable to statistical analysis, which was an important advancement for descriptive analysis methodology. Essential features of QDA were the use of screened and trained panels of 8–15 assessors guided by a trained panel leader, effective descriptive terms generated by the panel themselves, unstructured line scales and repeat evaluations and statistical analysis by analysis of variance (ANOVA) (Gacula 1997; Stone et al. 1974). The latter features of QDA not only enabled sensory scientists to obtain descriptions of product differences, but also facilitated assessment of panel performance and variability between products. Nevertheless, one limitation of QDA was the difficulty in comparing results between panels and between laboratories (Murray et al. 2001). In addition, similar to other conventional profiling methods, these techniques required extensive training and were costly to set up and maintain.

The *Spectrum™ Method* was developed in the 1970s by Gail Vance Civille, who presented the method at the Institute of Food Technologists Sensory Evaluation Courses in 1979. This technique was based on FPM and TPM, but unlike these methods, it evaluated all sensory modalities perceived and could be analysed statistically in a similar fashion to QDA data using ANOVA. A key feature was the use of a panel of 12–15 assessors who received in-depth and specialized training on scaling procedures using standard reference lists (Meilgaard et al. 2006). The use of reference products for anchoring attribute intensities purportedly reduced panel variability and gave the scores absolute meaning. This appealed to organizations who wished to use a descriptive technique in routine quality assurance operations (Lawless & Heymann 2010a). However, it also had a few disadvantages, one of which was associated with the difficulties in developing, training and maintaining a panel, as it was often very time-consuming (Lawless & Heymann 2010a). Another limitation of this technique included the difficulty in accessing reference products, as they were often unavailable to researchers outside the US. Substitution of local products could compromise the absolute nature of the scale and make cross-laboratory studies difficult, which may explain why the technique is more widely used in the US than in other countries. The Spectrum Method has been applied successfully to a wide variety of product categories, including meat (Johnsen & Civille 1986), catfish (Johnsen & Kelly 1990), paper and fabrics (Civille & Dus 1990) and skincare (Civille & Dus 1991), to name but a few.

The *ideal profile method* (IPM) came to the fore in the 1970s, with the need to identify the consumers' ideal product (Hoggan 1975; Moskowitz et al. 1977; Szczeniak et al. 1975) (see Cooper et al. (1989) for a review of early development). Originally, consumers rated predefined product attributes on their perceived and ideal intensities. In later derivations of the method, consumers were also asked to rate product acceptance, such as overall liking and purchase intention. Data analysis is complex, involving several steps to assess consistency, segmentation, definition of the ideal reference and guidance on optimization.

IPM provides actionable guidance for product improvement, although results need to be interpreted with care, particularly as consumer data are variable and consumers showed differences in their ideal profiles (van Trijp et al. 2007; Worch & Punter 2014a,b; Worch et al. 2010, 2012, 2013). Just-about-right scales have also been used to measure consumers' ideal profiles (Popper 2014). As this method measures consumer hedonics, it is beyond the scope of this book to cover it in detail.

Difference from control profiling (also known as deviation from reference profiling) was developed by Larson-Powers and Pangborn (1978a), who found that the deviation from reference scale improved the precision and accuracy of sensory responses. This technique uses a reference sample against which all other samples are evaluated on a range of attributes using a degree of difference scale. For example, samples that scored less than the reference for a specific attribute were indicated by a negative, whereas those that scored more were indicated by a positive (Lawless & Heymann 2010a). Stoer and Lawless (1993) felt this technique would be more effective to distinguish among difficult samples, or when the objective of the study involved comparisons to a meaningful reference. For example, Labuza and Schmidl (1985) used this technique to compare control product with product that had undergone accelerated shelf-life testing and demonstrated that it is useful for quality assurance or quality control work.

The importance of measuring sensory changes in products over time had long been recognized, but was difficult to carry out practically in the early days of sensory science. Continuous *time-intensity (TI) analysis* was presented in its modern form by Larson-Powers and Pangborn (1978b). Unlike conventional descriptive techniques, TI incorporated temporal aspects by continuously recording the evolution of a given sensory characteristic over a period of time. The result of TI measurement was typically a curve showing how the perceived intensity of the sensation increased and then decreased during consumption of a product. The measurement of temporal perceptual changes had been of interest for some time beforehand; an early example is Holway and Hurvich (1937), who asked assessors to trace a curve on paper to represent salt intensity. Other early methods involved making multiple assessments at short time intervals and constructing curves from the data (Sjostrom 1954) or plotting intensities on a paper graph, where the *x*-axis was time and the *y*-axis was perceived intensity (Neilson 1957). Larson-Powers and Pangborn were the first to gather continuous TI data, using a moving strip-chart recorder, in such a manner that assessors were required only to move a pen along a line scale to assess intensity and could not see their evolving curves to avoid bias.

As technology progressed, data were collected by computers; the first computerized system was developed by the US Army Nadick Food Laboratories in 1979 (Lawless & Heymann 2010b), which lead to a proliferation in TI studies. Statistical analysis of TI curves proved complex, and required some development. Assessors were often already trained QDA or profiling panellists, who were then trained in the TI assessment technique. TI was useful to describe a variety of ingredients and products with longer-lasting or changing sensory experiences (e.g. chewing gum, perfume) or products that changed over time through use (e.g. ice cream), and has also been used to understand how perception changes throughout consumption experience (e.g. sipping a cup of hot tea) (Kemp et al. 2009) and to investigate mechanisms of human perception (Piggott 2000). TI has the benefit of providing more detailed information than other descriptive techniques, but is time-consuming as evaluation is limited to one attribute at a time and requires a large number of assessments to cover even a small number of important product attributes. For reviews of TI, see Halpern (1991), Cliff and Heymann (1993), Dijksterhuis and Piggott (2000) and Lawless and Heymann (2010b). Temporal methods are beyond the scope of this book and will be covered elsewhere (Hort et al. 2017).

1.2.2.4 1980s

A more rapid method called *free choice profiling* (FCP) was developed in the UK during the 1980s (Williams & Langron 1984). This technique also met the demand and interest of marketing and product development teams in obtaining consumers' perception of products. Unlike other previous descriptive methods, this method allowed consumers to generate and use any number of their own attributes to describe and quantify product attributes. Therefore, as the assessors did not require any training, the process of data generation was relatively quicker and potentially cheaper compared to conventional techniques. However, one distinct challenge of the technique was the use of idiosyncratic words from consumers, such as 'cool stuff', 'mum's cooking', which made the interpretation of results difficult (Lawless & Heymann 2010a). Another factor to take into account was the different number of descriptors generated by the consumers; some used very few descriptors while some used many. Therefore, this method needed more sophisticated techniques, such as GPA, to transform each assessor's data into individual spatial configurations (Gower 1975). This technique has now been successfully applied to a range of products, such as alcoholic beverages (Beal & Mottram 1993; Gains & Thompson 1990), coffee (Williams & Arnold 1985), cheese (Jack et al. 1993), meat (Beilken et al. 1991), salmon (Morzel et al. 1999) and many more (see Tárrega & Taracón (2014) for a review).

Conventional descriptive and time-intensity techniques were not suitable to evaluate products with high individual variability in consumption speed, such as cigarettes. Gordin (1987) therefore developed the *intensity variation descriptive method*, which took account of individual consumption speed and provided information about changes in attribute intensities as samples were consumed. This technique asked assessors to evaluate products at specified locations in the product rather than at specified time intervals using standard descriptive methodology.

Sorting procedures were introduced as a descriptive technique in sensory science in the late 1980s. Assessors were asked to group samples according to their similarities and differences. Perceptual maps were created from the data. The inclusion of verbal description in the assessment enabled the dimensions of such maps to be explained (Popper & Heymann 1996). There are many variations on the exact sorting procedure applied or developed in sensory science (Chollet et al. 2014; Courcoux et al. 2014), including restricted sorting (Lawless 1989), free sorting (Lawless et al. 1995), descendant hierarchical sorting (Egoroff 2005), directed sorting (Ballester et al. 2009), ascendant hierarchical/taxonomic free sorting (Qannari et al. 2010), Sorted Napping® (Pagès et al. 2010), labelled sorting (Bécue-Bertaut & Lê 2011) and multiple sorting (Dehlholm et al. 2012, 2014b). Sorting techniques required minimal training, could be applied to a large number of samples and did not require any selection of attributes in advance, making them easier, quicker and cheaper to perform compared to other conventional techniques (Cartier et al. 2006). Lawless (1989) was probably one of the first to use this technique to profile sensory characteristics of odourants. Sorting has been applied on a variety of food products, including beers (Chollet & Valentin 2001), cheese (Lawless et al. 1995) and yoghurts (Saint Eve et al. 2004), and to evaluate different materials, such as plastic pieces (Faye et al. 2004) and fabrics (Giboreau et al. 2001). However, this technique should be limited to foods whose physico-chemical properties (temperature, structure, etc.) and resulting sensory properties remain stable throughout the sensory sessions (Cartier et al. 2006). Therefore, it is not appropriate to apply this technique in shelf-life studies.

1.2.2.5 1990s

Quantitative Flavour Profiling (QFP) was developed by Givaudan-Roure, Switzerland, as a modified version of QDA (Stampanoni 1994). Unlike QDA, this technique assessed flavour characteristics using a predefined lexicon for different product categories developed by a panel of 6–8 panellists, who were usually trained flavourists. Intensity was assessed by a trained panel using a line scale and end-of-scale intensity references were used for each study. A proposed advantage of QFP was its use of technical and non-erroneous terms from the experts (Murray et al. 2001). However, it also posed a challenge for marketing and product development teams to link the data to consumer perceptions and preferences. Nevertheless, the use of reference standards made this technique applicable for cross-laboratory and cross-cultural projects (Murray et al. 2001). QFP has been applied to profile foods, such as dairy products (Stampanoni 1994).

Projective mapping (Risvik et al. 1994) was proposed as a rapid method for sensorically mapping products. Untrained assessors were presented with all samples simultaneously and asked to physically place samples in space (on a sheet of paper or, more recently, by placing icons on a computer screen) so that perceptually similar samples are close to each other, and those that are more different are placed further apart, thus producing a physical representation of a perceptual map. GPA was applied for data analysis. *Napping®* is a variation on projective mapping (Pagès 2003, 2005a,b), which uses the same assessment procedure but has a more defined set of data analysis instructions. Several variations exist, including Napping with the addition of ultra-flash profiling, in which assessors also provide semantic description of products (Pagès 2003), Sorted Napping, in which assessors provide descriptions of product groupings (Pagès et al. 2010), Partial Napping, where assessors are guided, for example by sensory modalities (Dehlholm et al. 2012), and Consensus Napping, in which assessors give group assessment, although the latter was found to be unreliable with untrained assessors (Delholm 2014a). A major advantage of projective mapping was its spontaneity, flexibility and speed (Perrin et al. 2008). However, this technique did not characterize the product in detail and product description often had to be completed with sensory or instrumental data. Many variations of projective mapping exist which can influence results, including response surface framework, assessor instructions, assessor type and validation of product separations (Dehlholm et al. 2012) (see Dehlholm (2014a) and Lê et al. (2014) for a review).

Progressive profiling (Jack et al. 1994), which is similar to the intensity variation descriptive method discussed previously (Gordin 1987), merged the dynamic ideas from time intensity with ideas from flavour and texture profiling. This technique asked assessors to give an intensity score to an attribute at several time points, such as at each chew, chosen by the experimenter during the evaluation, and used references to allow comparison over time. However, limited correlations were found between progressive profiling, descriptive analysis and instrumental measurement when profiling textural attributes of hard cheese during mastication (Jack et al. 1994).

The *dynamic flavour profile method* (DeRovira 1996) was another extension of descriptive analysis and time-intensity methodology. The panels were trained to evaluate the perceived intensities of 14 specific aroma and taste attributes over time, including acid, bitter, brown, esters, floral, green, lactonic, salt, sour, spicy, sulfury, sweet, terpenoid and woody. The data produced a set of TI curves that characterized a sensory profile and were represented in three dimensions, whereby a cross-section of the plot yields a spider plot for a particular time point. Although the specification of 14 attributes was argued to be too restrictive, the method was deemed to have potential (Lawless & Heyman 2010a,b).

Dual-attribute TI (DATI) (Duizer et al. 1996) was developed to enable two sensory attributes to be measured simultaneously using continuous TI, thus halving the time required for single-attribute sensory evaluations. Although DATI was claimed to produce meaningful results (Zimoch & Findlay 2006), it has not been widely used, as assessors often found it difficult to assess and record two sensory characteristics at the same time, and therefore this technique requires further demonstration of its validity and value before it is widely accepted (Dijksterhuis & Piggot 2000).

1.2.2.6 2000s to the Present

This recent period of time has seen renewed interest in descriptive analysis, with a plethora of studies on new, rapid techniques with many modifications and variations. Sieffermann (2002) proposed a new technique called *flash profiling*. Untrained assessors selected their own attributes to describe and evaluate a set of products simultaneously, and then ranked the products using their own constructs. It was based on FCP but unlike FCP, which involves rating intensities, flash profiling required assessors to rank products on an ordinal scale for each attribute, and was therefore quicker than FCP. The individual maps created were then treated with GPA to create a consensus configuration. Cluster analysis could then also be performed on the descriptive terms to aid interpretation (Dairou & Sieffermann 2002; Tarea et al. 2007). The main advantages of this technique were that it was less time-consuming and more user friendly to run than conventional descriptive analysis (Sieffermann 2002), although data analysis was more complex. Flash profiling has been proven to be comparable to conventional profiling when assessing a set of red fruit jams, but this could be due to the large differences between the products evaluated (Dairou & Sieffermann 2002). Sieffermann (2002) proposed that flash profiling should be considered as a preliminary test rather than a substitute for conventional profiling. Nevertheless, this technique has shown practical feasibility in the evaluation of a variety of food products (Petit & Vanzeveren 2014), including dairy products (Delarue & Sieffermann 2004), apple and pear purées (Tarea et al. 2007), bread odourant extracts (Poinot et al. 2007), jellies (Blancher et al. 2007), etc. (see Delarue (2014a,b) for a review). *Individual vocabulary profiling* (Lorho 2005, 2010), a variant of flash profiling, gives better defined individual vocabularies and has been applied to sound quality evaluations.

Rank descriptive data (RDA) (Richter et al. 2010) is a variation on flash profiling, and was based upon an earlier method using ranking with an untrained panel (Rodrigue et al. 2000). In RDA, assessors developed an attribute list, were familiarized with ranking and developed a consensus rank ordering. It was found to give similar discrimination to QDA, whilst being quicker and using a smaller amount of product.

Another related technique, *polarized sensory positioning* (PSP), is a reference-based method for sensory characterization based on the comparison of samples with a set of fixed references, or poles (Teillet 2014a,b; Teillet et al. 2010; Varela & Ares 2012). There are several modifications, including PSP based on degree of difference scales and triadic PSP (Teillet et al. 2010), where an assessment is made about which reference product the test product is most and least similar to. Although the method is cheap and flexible, the comparison of samples and poles is again based on overall differences, without full product description, an indication of the sensory attributes that should be considered in the further evaluation or their relative importance.

Polarized projective mapping (PPM) (Ares et al. 2013) is a combination of PSP with projective mapping that enables the evaluation of samples in different sessions. Assessors are presented with three poles located on a piece of paper and asked to position sample products in relation to the poles so that perceptually similar samples are located close to each other and perceptually dissimilar samples are further away. Assessors can then be asked for product descriptions. Analysis is similar to that for projective mapping.

Another method that uses a reference is *Pivot Profile©* suggested by Thuillier in 2007 (see Valentin et al. 2012), in which free descriptions of the differences between a sample product and a single reference product (the 'pivot') are produced by asking assessors to list the attributes the product has in smaller or greater intensity than the pivot.

Temporal dominance of sensations (TDS) (Pineau & Schlich 2014; Pineau et al. 2003, 2009) was developed to evaluate product attributes simultaneously over time. TDS primarily records the sequence of the dominance of different attributes; however, it could also be used to record the intensities of each of the dominant sensations. The technique consists of presenting a panel of trained assessors with a complete list of attributes on a computer screen and asking them to identify, and sometimes rate, sensations perceived as dominant until perception ends. TDS has been shown to provide information on the dynamics of perception after product consumption that is not available using conventional sensory profiling (Labbe et al. 2009). However, Ng et al. (2012) have shown how using QDA and TDS in tandem can be more beneficial than using each alone. *Temporal order of sensations* (TOS) (Pecore et al. 2011) is a faster variation of TDS, which measures the order in which key attributes appear over the consumption experience.

Sequential profiling (Methven et al. 2010) is a modified version of progressive profiling, in which up to five attributes are scored over consecutive tastings, at set time intervals, in order to determine the perception of sensory attributes upon repeat consumption of a product over time. It has been shown that this technique generates additional information over standard techniques, such as a significant build-up of some attributes (e.g. mouthcoating) over total consumption volume. Several other methods that also make measurements at set time intervals include time-related profiling (Kostyra et al. 2008), time-scanning descriptive analysis (Seo et al. 2009) and multi-attribute time intensity (MATI) (Kuesten at al. 2013).

Conventional methods continued to be developed with the aim of reducing the time for evaluation. In 2010, *HITS profiling* (high identity traits) was proposed as a quicker method that provided more user-friendly information than traditional descriptive analysis techniques (Talavera-bianchi et al. 2010). The method used a simplified lexicon with fewer and more user-friendly attributes that could be understood by different users of the data. In 2012, the *optimized descriptive profile* (ODP) method was published (da Silva et al. 2012) with the aim of reducing the time for evaluation while estimating the magnitude of differences between samples. Assessors were familiarized rather than trained on references, and assessment was carried out on each attribute for all products, rather than for each product on all attributes. ODP was found to be 50% quicker than conventional profiling, whilst giving a similar sensory profile and discrimination power (da Silva et al. 2013).

Recently, verbally based qualitative methods have received attention in sensory science. 'All-that-apply' methods, most often called '*check-all-that-apply*' (CATA) or 'tick-all-that-apply' (TATA), involve assessors selecting all terms that apply to a product from a list of words. A variation is 'Pick-K attributes' (or *Pick K over N*), in which assessors select the K terms that are dominant or best describe the product. The CATA technique originated in the 1960s (Coombe 1964) and has been used in marketing research with consumers for decades, with ballots typically including CATA questions along with hedonic questions. In the experience of the authors, CATA lists for marketing research studies on food, beverage and fragranced products often included 'simple' sensory terms, such as 'sweet', 'citrus', 'strong', 'weak', etc., that were used for top-line product guidance. For example, at least since the 1990s, fragrance companies have used CATA to obtain sensory profiles of blinded fragrances and fragranced products using an attribute list of pure sensory terms (e.g. citrus, floral, strong), mixed with consumer terms (e.g. sporty, sophisticated). Interest in the application of CATA for more detailed sensory description was sparked in 2007 (Adams et al. 2007) and since then several variations have been proposed, including Pick K, or Pick K from N, in which assessors choose a set number of attributes (K) from the overall list (N) that best describe the product (see Valentin et al. (2012) for an overview), *forced-choice CATA/applicability testing*, in which assessors are required to answer yes/no questions to every attribute in the list (Ennis & Ennis 2013; Jaeger et al. 2014) and *rate-all-that-apply* (RATA) (Ares et al. 2014), in which assessors rate the terms they ticked as 'apply' (see Meyners & Castura (2014) and Ares & Jaeger (2014) for reviews).

An extension of CATA is *temporal check-all-that-apply* (TCATA) (Castura et al. 2016) which allowed continuous selection and deselection of multiple applicable attributes simultaneously over time. It built upon TDS, and used an approach similar to *time-quality tracking* (Zwillinger & Halpern 1991), an earlier method that also captured a sequence of attribute qualities without intensity scaling. Trained assessors indicate and continually update attributes that apply, thereby tracking sensations in the product as it changes over time. *TCATA fading* is a further development of TCATA, in which selected terms gradually and automatically become unselected over a predefined period of time (Ares et al. 2016). Results indicate that the TCATA and fading TCATA techniques have potential, but further research is needed to refine the methodology.

Open-ended questioning is another verbally based qualitative method that has recently received attention in sensory science. Assessors are asked for an opinion or comment and allowed to answer spontaneously and freely. Analysis of data may be carried out using a variety of techniques, including chi-square, chi-square per cell, correspondence analysis and multifactor analysis. Free comments are collected as supplementary information to other methods, such as sorting and Napping techniques. Open-ended questioning with subsequent comment analysis has been used to obtain product descriptions in consumer vocabulary (Ares et al. 2010) (see Symoneaux & Galmarini (2014) and Piqueras-Fiszman (2014) for reviews of methodology and analysis).

1.2.2.7 Continuing Customized Modification

The development of descriptive analysis illustrated above from the early days of 1950s to the present has given rise to many techniques, all of which have their relative merits. Since the earliest times of descriptive analysis, companies have developed their own customized methodology to meet specific project objectives or as their standard in-house methodology, which enables the most appropriate elements of different techniques to be modified and utilized. Most in-house descriptive methods are proprietary, but two examples of methods based on customization available in the public domain are QFP (see above and Chapter 10) and the *A⁵daptive Profile Method®* (see Chapter 11).

1.3 Descriptive Analysis as a Technique in Sensory Evaluation

1.3.1 Descriptive Analysis as a Tool

Descriptive analysis provides detailed, precise, reliable and objective sensory information about products. It uses humans as measuring instruments under controlled conditions to minimize bias in order to generate such data. In traditional methods, such as profiling-based methods and QDA, assessors with good sensory abilities are selected and trained for up to 6 months to rate perceived intensity and quality in a manner that is consistent within themselves and with other assessors to produce data that have been validated as acceptable (Heymann et al. 2014). Newer methods, such as FCP, flash profiling, sorting, projective mapping and PSP, can use naive consumers with no prior experience or training to group products based on overall similarities or differences, sometimes identifying and naming product differences first and then measuring them, or grouping products and then naming groups afterwards (Varela & Ares 2014).

There are some generic steps that are common across most traditional descriptive methods: assessor screening and selection; assessor training, including attribute generation, intensity calibration, development of assessment protocol and performance check; data generation using replication; and data analysis and reporting (Kemp et al. 2009). Newer, ‘rapid’ techniques have fewer generic steps: data generation, and data analysis and reporting (Dehlholm 2012). Some also include a prior familiarization step. Testing is quicker as there are fewer initial steps, so that a study can be completed in as little as one day, which reduces costs, although data analysis is more complex. However, it is noteworthy that once the panel in traditional techniques has been trained, subsequent studies on the same product/product category can also be run in a similar time-scale to newer methods, depending on the number of samples, without the inconvenience of having to recruit assessors for each study.

A key factor in the choice of descriptive analysis method is the choice of assessor, who may have no training, some familiarization or intensive training. Generally, the lower the level of training, the higher the variability of data produced and so the higher the number of assessors needed. Traditional methods use highly trained assessors, with the Spectrum Method said to use the most intensive training. Product experts have also been used, who may be more or less experienced than a trained panel. Newer techniques can use consumers with no training, but the trade-off is more variable data that are more difficult to interpret. Consumers may have differing levels of experience and expertise, ranging from naive consumers with no prior experience to category, product or brand users. Highly brand-loyal users can be more discriminating than trained panels. Newer methods give different levels of familiarization. For example, in FCP, assessors are exposed to many test samples when eliciting differences prior to the measurement phase, whereas some free sorting techniques provide no familiarization with the technique or samples.

Many studies have compared methodologies (a comparison of methods is given in Chapter 20) (see Ares & Varela (2014), Stone (2014) and Valentin et al. (2012). Often, similar results were obtained, although data from rapid methods appears less reliable and consistent (Dehlholm 2012). The most important factor when choosing a method, as for any good scientific study, is that the method selected should be appropriate to the objective of the study, and be able to produce actionable results and recommendations. Whichever method is used, good experimental controls, careful attention to practical experimentation and robust data analysis as described below will give confidence in the results obtained.

Descriptive analysis studies are typically carried out in a sensory laboratory with a controlled environment, which is neutral and has controlled lighting, temperature and humidity (ASTM 1986). Samples are produced/obtained, presented and assessed in such a way as to eliminate irrelevant and unnecessary variability and bias. Samples may be prepared according to experimental designs, depending on the objective, for example to vary ingredients and physico-chemical properties in a systematic manner. Experimental designs for sample presentation are employed to eliminate bias, which may range from a simple balanced, complete block design to a complex nested, incomplete block design, depending upon the number of samples and experimental variables. Traditionally, samples are presented in a sequential monadic fashion and all attributes are assessed for each product. Descriptive testing has also been carried out on an attribute-by-attribute basis, often using ranking or rank-rating (Kim & O'Mahony 1998) (later termed positional relative rating by Cordinnier and Delwiche (2008)), in which all products are assessed on each attribute in turn. In comparisons between serial monadic and attribute-by-attribute protocols with untrained assessors (Ishii et al. 2007) and trained assessors (Ishii et al. 2008), untrained assessors performed better using attribute-by-attribute evaluation, while the reverse was true for trained assessors. For newer methods, samples presentation may be simultaneous.

Data generated can be purely qualitative although, most often, they are quantitative and can be generated using a variety of measurement techniques, such as ranking, category scales, line scales and magnitude estimation, all of which have advantages and disadvantages. Replication of typically between two and four replicates is used to provide reliability, that is, to demonstrate that the data are reproducible under the same experimental conditions. Data are compared using statistical analysis, such as ANOVA, to determine significant differences in sensory characteristics between samples. Multidimensional statistics are used to produce descriptive maps of sample sets, and are the most appropriate method of analysis for some methods, such as FCP and sorting. Typically, data from traditional methods can be combined across studies with the use of suitable experimental elements, such as common controls, references and samples, and across an extended period for data mining, whereas this is more difficult for some of the newer methods, such as rank rating, sorting and projective mapping.

Descriptive analysis is used to give a precise description of the sensory properties of products and comprehensively describe the nature of the differences between them. It may be used to assess sensory characteristics from all sensory modalities and traditional methodology can be used to provide a full sensory description. Some methods, such as flavor and texture profiling, focus on restricted modalities. It is also possible to focus only on selected modalities and sensory attributes, such as those that are important to consumers. Traditional methods measure attributes individually, whereas newer methods, such as sorting, projective mapping and PSP, compare many attributes simultaneously to assess overall sensory similarities and differences holistically, that is, without the need to be trained to identify individual attributes.

Conventional profiling-type panels are intensively trained on and work with a technically based, well-defined attribute list. QDA panels are trained on and work with less technical language. The language tends to become more predefined for studies subsequent to the initial study in which attributes are generated. FCP assessors are able to choose individual attributes without training that are in effect consumer terms, such as creamy, refreshing. Other rapid techniques may allow for description of products or product groups before or after measurement. Technical terms (e.g. vanillin) are more informative to product development as they can be related directly to ingredients and process variables, but may need to be linked to consumer data for directional guidance. Consumer terms reflect the language of the target population better than technical terms from the experts and more traditional techniques, and are of more interest to marketing teams, but may be difficult to interpret and action for product development.

The sensory characteristics of products change over time. The time period may be as short as a single bite, for example the change of a frozen dessert in the mouth from a hard, cold solid to a warm, liquid releasing increased flavour volatiles, to a much longer time period, of perhaps many weeks, for example an air freshener gel gradually releasing less fragrance. These changes in perceived product sensory characteristics over time are partly due to changes in consumers' sensory systems that these products induce, such as short- and long-term adaptation, as well as changes in the products themselves. Descriptive analysis can be used to measure sensory changes in time, either by simply applying typical descriptive methods at specific time points or by using specially adapted descriptive methodology. It is beyond the scope of this book to cover such time-intensity methodology, which is given comprehensive coverage in another book in this series (Hort et al. 2017).

Descriptive analysis techniques are flexible and most methods have been adapted to suit the needs of particular industries, products, projects or applications. As discussed at the end of section 1.2.2, many companies develop proprietary, in-house, customized methodology. These can take elements from other descriptive methods, and often include generic steps that are common across several descriptive techniques, which allow the most suitable philosophies of various methods to be modified, combined and used. Different descriptive techniques may be used in combination to provide more comprehensive information or to improve efficiency. For example, a rapid method may be used to obtain an initial 'look see' before employing a longer, more in-depth and more expensive technique.

A major benefit of descriptive analysis is its ability to give additional, comprehensive information above other sensory methods and it is often used in conjunction with them. Discrimination testing is more sensitive than descriptive analysis at detecting differences and, as it is typically quicker and cheaper, it is often used as a first step to determine whether there are significant sensory differences between products. If it is unclear whether a sensory difference exists between products, it is sensible to carry out discrimination testing first to confirm this before committing additional resources to descriptive analysis. Discrimination testing cannot, however, give the level of information about the nature of sensory differences that descriptive analysis can.

Without objective, detailed descriptive data provided by descriptive analysis, understanding consumers' sensory needs would be much more difficult. Consumers are good at articulating their sensory likes and dislikes, but are not good at telling us why. Consumer product testing provides information on sensory hedonics and preference, but it is difficult to interpret in an actionable manner. Data from descriptive analysis can be linked to consumer liking and preference data using techniques such as preference mapping, in order to identify sensory drivers of liking and groups of consumers with similar sensory preferences.

Descriptive analysis produces data that are objective, precise and repeatable. In other words, it produces data that are on a par with data produced from analytical instruments, but with greater variance due to inherent variation within and across individuals. This enables sensory data to be robustly linked with instrumental data, such as physico-chemical data, leading to better understanding of the perception of products. The human senses remain more sensitive than instruments, such as the electronic nose, and these methods must be calibrated against the human senses, often using descriptive analysis, before they can be used as stand-alone methods.

Data from descriptive analysis are at their most powerful when used with consumer and instrumental data for product design. This enables the key attributes for consumer liking and disliking to be identified and manipulated to give optimum benefits.

Data from descriptive analysis has become more important in marketing to understand how consumer factors, such as language, perceptions, liking, expectation, emotion, values and behaviour, are related to a product's sensory properties for use in branding, communication and advertising. Further details are given in section 1.4.3.

1.3.2 Advantage and Disadvantages

The main advantage of descriptive analysis is the type of information it can provide: a comprehensive sensory description of a product that enables the comparison of multiple sensory characteristics within and across products. It can be, however, more time-consuming and expensive than other sensory methods. Traditional descriptive methodology takes more time to design, set up and carry out. In general, compared with other sensory methods, data analysis is more complex, particularly when linking descriptive data to other data and displaying results. Although newer descriptive techniques are faster and cheaper than traditional methods, as they can be completed with little or no training and hence less lead time for data generation, the trade-off is that data analysis is longer and more complex, and the information gained less detailed.

The use of a panel of trained assessors has advantages and disadvantages. The human sensory system is more sensitive than any instrument and so more subtle and complex sensory nuances can be picked up. A disadvantage is that humans are not as consistent as instruments in that they are subject to individual variation and bias. Assessors are all individuals who perceive things differently, for example due to physiological differences; process sensory information differently, for example using words and scales differently; and have different abilities to carry out the work, such as focus, maintaining motivation, personality, etc. Some argue that humans may have limited capacity to distinguish components of mixtures and question whether the theory of psychophysical modelling upon which traditional descriptive analysis is based is meaningful for complex attribute description (Laing 1991). Rapid methods make holistic measurements, and are therefore said to be closer to 'true' perception. However, they are more idiosyncratic and so less reliable than traditional methods.

Bias can be overcome with the use of training, references and controlled experimental conditions, such as procedure, design and environment, although this increases time and costs. However, once a trained panel is set up, in general, it can operate efficiently and continuously with few breaks as a workhorse for data generation, although it requires some maintenance. There is a risk that previous experience can influence current studies. For example, once attribute labels, references, scale types or scale ranges have been learned and used in a particular way, it can be very difficult for panellists to change usage. This is also true of the conceptual structure derived from sorting (Ishii et al. 1997). Some newer methods do not use trained panels, which can reduce risk associated with prior knowledge, as well as time and costs, although this may introduce other disadvantages, such as the need to recruit assessors for each study and the ability to compare across sessions and sample sets.

Descriptive studies can use large varieties and quantities of samples that need to be ready for testing simultaneously. In-house sample production can be time-consuming and expensive, taking up resources and potentially pressuring the department producing the samples. If produced via the production line, it may be difficult to balance requirements with those of immediate commercial needs. If samples need to be purchased, such as competitive products, this can be more expensive, and also resource intensive in order to ensure products are from the same batch, age, etc. Storage can be problematic, as all samples must be stored under stable, similar conditions (unless investigating effect of storage conditions). Control and reference samples must typically be available consistently, and be stored stably, over a longer period.

Descriptive analysis is typically carried out in a laboratory under controlled conditions and this may not represent what occurs under real-life conditions. However, as discussed in section 1.2.1, descriptive analysis is increasingly being carried out in more realistic contexts, aided by new technology.

The success of descriptive analysis boils down to stringent panel screening and training (where appropriate), and proper sensory execution and management, which are rarely inexpensive and easy. It requires a long-term commitment from the company or research centre. However, the benefits of having this important and sensitive analytical descriptive analysis usually outweigh the disadvantages. For this reason, descriptive analysis continues to be an important tool.

1.4 Application of Descriptive Analysis

The evolution of descriptive analysis techniques described in the previous section has been driven by the desire to broaden the scope of its application from a stand-alone method for comparing the sensory properties of products, which was most useful for QA/QC, to a key element in sophisticated consumer-driven product design and marketing. It is now used across many industries, including food, beverages, personal care, household, cosmetics, fragrances, pharmaceuticals, automotive and many others.

1.4.1 Product Development and Design

Descriptive analysis is used throughout the product design and development process to provide information about product differences, sensory drivers of liking and competitor analysis as a function of ingredient, packaging and process variables.

1.4.1.1 Market Overview and Opportunity Definition

One of the prime uses of descriptive analysis in product development is to compare the perceived sensory characteristics of products on the market. This might be to compare current products and their competitors for benchmarking and monitoring, to carry out a category or market review or to look for opportunities, such as sensory characteristics and combinations that are not delivered by the market or the current product range. Originally, this was carried out using pure sensory description, but linking sensory to consumer data, such as hedonics, enables sensory-based consumer segments to be identified and targeted.

1.4.1.2 New Product Design and Development

Once opportunities have been identified, a precise sensory target can be defined using descriptive analysis. The product development process can become more time- and cost-effective when prototypes matching the target have been developed using descriptive analyses coupled with experimental design and modelling to link physico-chemical properties, obtained via instrumental measures, to sensory and consumer data, rather than using trial and error. Such models can help product developers focus on attributes that are important to consumer segments, rather than just noticeably different, at different stages of consumer/product interaction, such as before, during and after product use. In addition, descriptive analysis helps to guide product developers to develop innovative products, through pilot plant scale-up, production benchmarking, package design, storage trials, cost reduction, product/processing change and product optimization.

1.4.1.3 Product Optimization

Product optimization may be carried out, for example, to improve liking, reduce costs, create value or substitute ingredients. Descriptive analysis can be used for all of these applications, using modelling techniques linking sensory, instrumental and consumer data as described above. However, descriptive analysis may not be the most appropriate technique when the objective is to match the current product or the sensory differences are expected to be very small. In these cases, discrimination techniques may be more appropriate.

1.4.1.4 Protection of Competitive Advantage

Once a product, ingredient or process has been developed, the competitive advantage gained needs to be protected. Sensory properties and/or the means by which they are produced form part of the competitive advantage and descriptive analysis plays a role in their protection.

Data from descriptive analysis have been used to support patents. It is recommended that a study for the sole purpose of defending the patent is carried out. A disadvantage is that the full experimental protocol must be disclosed in a patent, which could mean putting proprietary descriptive methodology in the public domain. The protocol used to support the patent needs to be chosen with care and could be a modification of proprietary descriptive methodology to prevent full disclosure or a non-proprietary descriptive method already in the public domain.

Some signature sensory characteristics, such as a precise colour, flavour or fragrance, that form part of a branding mix, or a perception that is an integral part of the product, such as the smell of a fine fragrance, can be protected by copyright or trademark. In this case, a precise specification of the sensory characteristic is needed, which is often obtained using descriptive analysis. Such data have also been used in court to support cases brought for breach of copyright.

1.4.2 Quality Assurance and Quality Control (QA/QC)

Initially, descriptive analysis had its roots in 'golden tongue' experts used to assess product quality. Trained panels were introduced to move away from the idiosyncracies of individual experts and produce structured, comparable and validated data that could provide documentation and comparison of perceived sensory characteristics of the current products for quality control and assurance. Formal quality control sensory programmes using trained panels were initiated in the 1950s; however, awareness of their importance arose in the early 1960s, resulting in the establishment of QC sensory programmes in industry. There was increased publication of quality control sensory techniques in the 1990s, including two important texts (ASTM 1992; Munoz et al. 1992). QA/QC has become increasingly consumer driven, in line with technique development as described above, determining and focusing on the sensory attributes that are important to consumers, and ensuring that production remains within ranges giving the optimal perceived intensity of such sensory characteristics.

Trained descriptive QA/QC panels are used to:

- assess the sensory quality and consistency of ingredients
- assess consistency of production runs and batches
- assess consistency of products from different manufacturing sites
- detect sensory defects before products go to market
- describe how sensory characteristics of the product change over time for shelf-life testing
- assess the quality and consistency of the product through the supply chain
- investigate consumer complaints
- investigate taints.

The most important criteria of a QA/QC sensory programme are the training of assessors, the type of established sensory specification and the use of controlled test conditions (Munoz 2002). Kilcast (2010) gives a practical guide to using descriptive analysis for QA/QC. It typically involves a shortened version of the full sensory descriptive profile, focusing on key sensory attributes of importance to the consumer or the manufacturing process. It may be more convenient to use instrumental rather than sensory measures for routine QA/QC, in which case instrumental measurements must be validated against sensory measures through sensory-instrumental data relationship studies to be useful and reliable (Munoz 2002).

1.4.3 Marketing

Sensory and consumer research now routinely focuses on the relationship between sensory properties and overall liking. Modelling, such as preference mapping, enables marketing teams to identify and target consumer segments or regional markets differentiated by sensory preferences. However, in these days of very mature and competitive markets, potential interactions of sensory properties and other factors have become interesting and important. The role of descriptive analysis in market research and consumer science has evolved from hedonic measures to more explicit behavioural measures, such as purchase intentions, and to implicit behavioural measures such as emotional benefits (e.g. chocolate makes me happy and will calm me), expectation, functional attributes (e.g. healthy but less flavourful food is good for me). There is a need to understand how intrinsic factors (food stimulus dependent) and extrinsic product characteristics such as environment, advertising, packaging or production variables influence consumer perception, expectation, emotion, values (Thomson et al. 2010) and hence hedonic responses. The need to understand such complex relationships between product characteristics and consumer behaviour has led sensory scientists to adopt methods from other scientific disciplines, such as the repertory grid method (Gonzáles-Tomás & Costell 2006; Piggott & Watson 1992), mean-ends chain (Brunsø & Grunert 2007; Costa et al. 2007) and conjoint analysis (Enneking et al. 2007).

The applications of sensory descriptive data in marketing and advertising include:

- defining and describing the 'sensory journey' of product usage
- identifying/optimizing the sensory characteristics of a brand
- understanding the perceptual basis of consumer sensory language for use in communication
- understanding the longer-term sensory-based relationship between the user and the product/brand
- understanding, creating and optimizing sensory cues for functional and emotional benefits
- making and substantiating sensory-based advertising claims.

1.4.4 Research

Descriptive analysis has been applied in basic research in order to understand the human sensory system, and to investigate the basis of physical and chemical stimulation of this system. It has been used in areas such as mechanisms of sensory receptors, individual differences in physiology, molecular basis of sensory functionality, flavour perception mechanisms, sensory processing in the brain, etc.

Model systems and instruments, such as the electronic nose, electronic tongue, Instron®, etc., that mimic and predict the human sensory system have not only enhanced understanding of mechanisms of sensory perception, but eventually could be a cost- and time-effective replacement for a trained descriptive panel. However, our sensory systems are complex, and research has made steady yet slow progress to date (Ross 2009). Combinations of descriptive analysis and instrumental techniques, including those providing data on chemical properties, physical properties and brain functionality, have enabled multimodal flavour perception and the perceptual interactions between the senses to be studied (Eldeghaidy et al. 2011; Hort et al. 2008).

1.5 Contributions of Descriptive Analysis

1.5.1 Contribution to Industry

Descriptive analysis has made a major contribution to industry, partly by virtue of the type of robust sensory data it generates, as outlined earlier. As a consequence, innovation and new product development (NPD) have become more precise, structured and consumer focused, creating better products that better meet consumer needs, taking into account product, pack and process variables.

The manufacturing process is delivering improved sensory quality that is more focused on consumer sensory needs. Prior to descriptive analysis, the QA process relied on individual experts with all their idiosyncrasies. Descriptive analysis has enabled consumer-relevant sensory elements of product specifications to be identified, precisely defined and measured in a way that is stable over time. Product defects can be precisely characterized.

Marketing can better analyse the sensory-based categorization of products, competitors and consumers in the market and target them. Positioning, such as branding, advertising and advertising claims, can be better linked to the product's sensory characteristics to reinforce positioning, and enable sensory elements to become integrated into the positioning.

As the needs of industry have become more demanding, descriptive analysis has evolved to meet the challenges. Descriptive analysis has contributed in many product categories, such as food, beverages, personal care, household, cosmetics, pharmaceuticals, textiles, etc., and the list continues to grow. It has proved adaptable over the years, and a selection of methods is now available for different situations, such as those requiring sensory data that are more detailed, quick to generate, cheap, technically orientated, consumer orientated, stable over time, real to life, etc.

1.5.2 Contribution to Other Scientific Disciplines

1.5.2.1 Psychology

Descriptive analysis involves using humans as an analytical 'instrument'. Unlike instruments, however, human judgements can be easily affected by psychological factors, and sensory scientists must ensure that the chosen procedure, scale and experimental design minimize psychological biases. The development of descriptive analysis methods has therefore lead to a greater understanding of sensory psychology.

In the early days of traditional descriptive analysis, there was much debate over which type of scale produced the 'best' data. Much research was put into understanding how different scales worked perceptually, (e.g. category versus continuous scaling, scale-end bias, etc.), how individuals used scales differently (e.g. conservative scaling, effect of experience, differences in perceptual sensitivity) and how to train individuals to use scales in the same way (e.g. references, controls, blind samples, etc.). Scales were developed and applied to descriptive analysis to attempt to overcome different biases (e.g. the Labelled Intensity Magnitude Scale (LIMS) is used to overcome scale-end bias). Today, it is recognized that there is no one 'best' scale, but the most appropriate scale for the particular application (see Lawless & Heymann (2010c) and Lawless (2013) for an overview of scaling).

In a similar way, much research has been carried out on understanding how individuals perceive and use descriptive attributes. Attributes are perceptual concepts, consisting of a qualitative perceptual experience named with a descriptive word (descriptor). Research has been undertaken on perceptual categorization to understand how individuals designate a perceptual sensation to one attribute or another and learn a system of perceptual categorization. Research has also been undertaken on concept alignment to understand how to train individuals to have the same understanding of an attribute (e.g. use of perceptual references), and on the application of language to perceptual concepts to understand how descriptors are applied to sensory concepts.

In order to perform descriptive analysis, panellists must learn to recognize a range of sensory attributes and remember perceptual intensities. They must use their memory and be motivated to maintain performance, and research has been undertaken on sensory learning and memory in order to devise better training techniques.

The research referred to above has contributed to other disciplines, such as psychophysics and market research.

1.5.2.2 Physiology

As descriptive analysis involves using humans as an analytical 'instrument', it is desirable that all panellists produce data in the same way. Panellists, however, are all individuals with differing perceptual sensitivities that can lead to differing use of attributes and scales. Much research has gone into understanding the factors affecting individual sensitivity, including age, health, lifestyle, such as smoking status, genetics, etc., which has been applied to develop a selection process to identify the most sensitive potential panellists. Some selection tests have been used in other disciplines; for example, in-mouth texture related tests have been used in oral-related sciences (Fillion & Kilcast 2001) and odour recognition tests have been used in the medical field. Descriptive analysis has brought individual differences into sharp focus, as few people have good sensitivity across all sensations, and this has lead to a greater understanding of sensory physiology.

1.5.2.3 Physico-chemistry

In attempting to understand the relationship between taste, odour and molecular structure using descriptive techniques, much effort has been put into identifying the chemical structure of molecules that may not otherwise have been investigated. Recently, this has been particularly true of naturally occurring compounds, particularly those with sweet taste such as stevia, the ability to block bitter taste and unique aromas.

It has been notoriously difficult to statistically link descriptive data to instrumental measures, particularly for food texture. This has lead to improvements in instruments, so that they better mimic in-mouth processes, such as mastication and flavour release.

1.5.2.4 Statistical Analysis

The need to analyse descriptive analysis data and combine them with other types of data has lead to the development of new statistical techniques, and contributed to a new field of statistics: sensometrics.

Initial descriptive analysis methods, flavor and texture profiling, did not use statistical analysis. It was first employed in QDA, when it was accepted that sensory data include inherent variability due to individuals. Classic descriptive analysis is typically analysed statistically using univariate techniques, such as ANOVA. The generation of data from scales of various types has necessitated new assumptions about the nature of the data generated, and hence adaptation of the techniques used. Sophisticated statistical analyses have also been developed to assess descriptive panel performance for discrimination, reproducibility and consensus. Considerable effort has been put into displaying descriptive data in a way that is understandable and meaningful to non-sensory scientists.

As descriptive techniques have evolved over time, so computing power has increased, and with it, the ability to combine descriptive data with other types of data to move from confirmatory to exploratory analyses. This has enabled the evolution and application of sophisticated modelling techniques and multivariate analysis, such as RSM, principal component analysis, GPA and preference mapping. More complicated studies have also necessitated the development of more sophisticated experimental design techniques. The increasing sophistication of modelling technique has, in turn, made the product design process more sophisticated. Software for sensory data collection, analysis and presentation continues to develop.

1.6 Summary

Descriptive analysis is undoubtedly one of the most sophisticated, flexible and widely used tools in the field of sensory analysis. Over the years, many techniques have been developed to meet different objectives and applications, each with their own advantages and disadvantages. Descriptive analysis plays an imperative role in industry, being utilized at all stages of a product's life cycle from new product development to postlaunch monitoring to provide insights to research, marketing and QA/QC teams to help guide development, commercial, communication and maintenance strategy. Descriptive analysis is also used as part of multidisciplinary scientific investigations involving other fields for mutual benefit, such as food science, nutrition, physio-chemistry, psychophysics, psychology, physiology, neuroscience and genomics. A key aim of this multidisciplinary approach is to improve understanding of consumer perception, hedonics and behaviour. Descriptive analysis has continued to broaden its application from its roots in food and beverages to non-food categories.

1.7 Future Developments

It is anticipated that descriptive analysis will continue to develop faster and more flexible techniques, while seeking to improve the level of detail delivered and reduce the time required for data analysis. These will be combined and customized by users to develop techniques to meet their specific situations and test objectives. Research continues on newer, rapid methodology and Ares (2015) highlights particular areas for consideration: understanding the cognitive processes involved in samples evaluation, development of tools for evaluating reliability and the identification of limitations of new methodologies. There may well be a rise in the use of rapid methods with trained panels from traditional techniques to achieve the advantages of speed combined with reliability, reproducibility and improved interpretation of results. Technology will continue to impact development, so that testing in context and during real-time usage becomes commonplace, using tablets, mobile phones, apps and social media forums, etc. In short, descriptive analysis will maintain its place as a key technique in sensory evaluation, through its ability to adapt and enhance its capabilities and, hence, its usefulness.

1.8 Overview of Book

This book gives comprehensive coverage of time-static descriptive analysis. Descriptive techniques specifically designed to measure time-dependent changes in sensory properties are beyond the scope of this book, but are covered elsewhere (Hort et al. 2017). Descriptive techniques related to consumer hedonics, such as ideal and just-about-right profiling, are also beyond the scope of this book. The book is written by sensory scientists with industrial and academic backgrounds, selected for their expertise in descriptive analysis. It is organized into four sections.

Section 1, *Introduction*, covers the general principles of descriptive analysis. Chapter 1 gives an overview of descriptive analysis, and traces its evolution, applications and contributions. Chapter 2 addresses general considerations when conducting descriptive analysis, including psychological factors relevant to qualitative and quantitative assessment. Chapters 3 and 4 provide detailed information on panel set-up and training, and panel quality management (monitoring, performance and proficiency) respectively. Chapter 5 covers analysis of descriptive data.

Section 2, *Techniques*, consists of 12 chapters each detailing a descriptive technique, including an overview of its method, history, theory, analysis, practical considerations, advantages/disadvantages, applications, case studies and future directions. Chapters are generally in order of historical development, except in a few logically dictated cases. Traditional techniques covered are consensus methods (Chapter 6), original and modified/derivative profile methods (Chapter 7), Quantitative Descriptive Analysis (Chapter 8) and the Spectrum Method (Chapter 9). The newer techniques covered are ranking and rank-rating (Chapter 12), free choice profiling (Chapter 13), flash profiling (Chapter 14), projective mapping and sorting methodologies (Chapter 15), polarized sensory positioning (Chapter 16) and check-all-that-apply (CATA) and free choice description (FCD) (Chapter 17). Many practitioners customize descriptive methods to meet their own specific requirements. Two examples are given based on traditional techniques: Quantitative Fragrance Profiling (Chapter 10) and the A^{5}daptive Profile Method (Chapter 11).

Section 3, *Applications*, consists of two chapters describing applications of descriptive analysis in food (Chapter 18) and non-food products (Chapter 19). Both chapters include general considerations, considerations related to specific products and examples.

Section 4, *Summary*, consists of a final summary (Chapter 20) that compares the descriptive methods from previous chapters and outlines future directions for descriptive analysis.

References

Adams, J., Williams, A., Lancaster, B. & Foley, M. (2007) Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks. 7th Pangborn Sensory Science Symposium, 12–16 August, Minneapolis, USA.

Amerine, M.A., Pangborn, R.M. & Roessler, E.B. (1965) *Principles of Sensory Evaluation of Food*. New York: Academic Press.

Ares, G. (2015) Methodological challenges in sensory characterization. *Current Opinion in Food Science*, **3**, 1–5.

Ares, G. & Jaeger, S.R. (2014) Check-all-that-apply (CATA) questions with consumers in practice: experimental considerations and impact on outcome. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 227–246.

Ares, G. & Varela, P. (2014) Comparison of novel methodologies for sensory characterization. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 365–389.

Ares, G., Gimenez, A., Barreiro, C. & Gambaro, A. (2010) Use of an open-ended question to identify drivers of liking of milk desserts. Comparison with preference mapping techniques. *Food Quality & Preference*, **21**, 286–294.

Ares, G., de Saldamando, L., Vidal, L., Antúnez, L., Giménez, A. & Varela, P. (2013) Polarized projective mapping: comparison with polarized sensory positioning approaches. *Food Quality & Preference*, **28**, 510–518.

Ares, G., Bruzzone, F., Vidal, L. et al. (2014) Evaluation of a rating-based variant of check-all-that-apply questions: rate-all-that-apply (RATA). *Food Quality & Preference*, **36**, 87–95.

Ares, G., Castura, J.C., Antúnez, L. et al. (2016) Comparison of two TCATA variants for dynamic sensory characterization of food products, *Food Quality & Preference*, **54**, 160–172.

ASTM Committee E-18 on Sensory Evaluation of Materials and Products (1986) *Physical Requirement Guidelines for Sensory Evaluation Laboratories*. ASTM Special Technical Publication 913. Philadelphia: American Society for Testing and Materials.

ASTM Committee E-18 on Sensory Evaluation of Materials and Products (1992) *Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia: American Society for Testing and Materials.

Ballester, J., Abdi, H., Langlois, J., Peyron, D. & Valentin, D. (2009) The odor of colors: can wine experts and novices distinguish the odors of white, red, and rose wines? *Chemosensory Perception*, **2**, 203–213.

Beal, A.D. & Mottram, D.D. (1993) An evaluation of the aroma characteristics of malted barley by free-choice profiling. *Journal of Science of Food & Agriculture*, **61**, 17–22.

Bécue-Bertaut, M. & Lê, S. (2011) Analysis of multilingual labeled sorting tasks: application to a cross–cultural study in wine industry. *Journal of Sensory Studies*, **26**, 299–310.

Beilken, S.L., Eadie, L.M., Griffiths, I., Jones, P.N. & Harris, P.V. (1991) Assessment of the sensory characteristics of meat patties. *Journal of Food Science*, **56**, 1470–1475.

Blancher, G., Chollet, S., Kesteloot, R., Nguyen, H.D., Cuvelier, G. & Sieffermann, J.M. (2007) French and Vietnamese: how do they describe texture characteristics of the same food? A case study with jellies. *Food Quality & Preference*, **18**, 560–575.

Bourne, M.C., Sandoval, A.M.R., Villalobos, M.C. & Buckle, T.S. (1975) Training a sensory texture profile panel and devlopment of standard rating scales in Colombia. *Journal of Texture Studies*, **1**, 43–52.

Brandt, M.A., Skinner, E.Z. & Coleman, J.A. (1963) Texture profile method. *Journal of Food Science*, **28**, 404–409.

Brunsø, K. & Grunert, K.G. (2007) Consumer attitude measures and food product development. In: MacFie, H. (ed.) *Consumer-Led Food Product Development*. Boca Raton: CRC Press, pp. 197–222.

Cairncross, S.E. & Sjostrum, L.B. (1950) Flavour profiles: a new appeach to flavour problems. *Food Technology*, **4**, 308–311.

Cartier, R., Rytz, A., Lecomte, A. et al.(2006) Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality & Preference*, **17**, 562–571.

Cartwright, L.C. & Kelly, P.H. (1951) Sharper flavor ratings with improved profile test. *Food Engineering*, **23**, 71–73, 215.

Castura, J.C., Antúnez, L., Giménez, A. & Ares, G. (2016) Temporal check-all-that-apply (TCATA): a novel dynamic method for characterizing products. *Food Quality & Preference*, **47**(A), 79–90.

Chollet, S. & Valentin, D. (2001) Impact of training on beer flavour perception and description: are trained and untrained panelists really different? *Journal of Sensory Studies*, **16**, 601–618.

Chollet, S., Valentin, D. & Abdi, H. (2014) Free sorting task. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 207–228.

Civille, G.V.D. & Dus, C.A. (1990) Development of terminology to describe the handfeel properties of paper and fabrics. *Journal of Sensory Studies*, **5**, 19–32.

Civille, G.V.D. & Dus, C.A. (1991) Evaluating tactile properties of skincare products: a descriptive analysis technique. *Cosmetic & Toiletries*, **106**, 83–88.

Cliff, M. & Heymann, H. (1993) Development and use of time-intensity methodology for sensory evaluation – a review. *Food Research International*, **26**, 375–385.

Coombe, C.H. (1964) *The Theory of Data*. New York: Wiley.

Cooper, H.R., Earle, M.D. & Triggs, C.M (1989) Ratio of ideals – a new twist to an old idea. In: Wu, L.S. (ed.) *Product Testing with Consumer for Research Guidance*. Philadelphia: American Society for Testing and Materials, pp. 54–63.

Cordinnier, A.M. & Delwiche, J.F. (2008) An alternative method for assessing liking: positional relative rating versus the 9-point hedonic scale. *Journal of Sensory Studies*, **23**(2), 284–292.

Costa, A.I.d.A., Schoolmeester, D., Dekker, M. & Jongen, W.M.F. (2007) To cook or not to cook: a means–end study of motives for choice of meal solutions. *Food Quality & Preference*, **18**, 77–88.

Courcoux, P., Quannari, E.M. & Faye, P. (2014) Free sorting as a profiling technique for product development. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 153–186.

Dairou, V. & Sieffermann, J.M. (2002) A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, **27**, 826–834.

Da Silva, R., Minim, V.P.R., Simiqueli, A.A., da Silva Moraes, L.E., Gomide, A.I. & Minim, L.A. (2012) Optimized Descriptive Profile: a rapid methodology for sensory description. *Food Quality & Preference*, **24**, 190–200.

Da Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Lucia, S.M.D. & Minim, L.A. (2013) Qualitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality & Preference*, **30**(2), 169–179.

Dehlholm, C. (2012), Descriptive sensory evaluations. Comparison and applicability of novel rapid methodologies. PhD thesis, Copenhagen, Denmark.

Dehlholm, C. (2014a) Projective mapping and Napping. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 229–253.

Dehlholm, C. (2014b) Free multiple sorting as a sensory profiling technique. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 187–196.

Dehlholm, C., Brockhoff, P.B., Meinert, L., Aaslyng, M.D. & Bredie, W.L.P. (2012) Rapid descriptive sensory methods – comparison of free multiple sorting, partial Napping, napping, flash profiling and conventional profiling. *Food Quality & Preference*, **26**, 267–277.

Delarue, J. (2014a) Flash profiling. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 255–270.

Delarue, J. (2014b) Flash profile, its evolution and uses in sensory and consumer science. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 27–52.

Delarue, J. & Sieffermann, J.M. (2004) Sensory mapping using flash profile. Comparison with conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality & Preference*, **15**, 383–392.

Delarue, J., Lawlor, B. & Rogeaux, M. (2014) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing.

DeRovira, D. (1996) The dynamic flavor profile method. *Food Technology*, **50**, 55–60.

Dijksterhuis, G.B. & Piggott, J.R. (2000) Dynamic methods of sensory analysis. *Trends in Food Science &Technology*, **11**, 284–290.

Duizer, L.M., Bloom, K. & Findlay, C.J. (1996) Dual-attribute time-intensity measurement of sweetness and peppermint perception of chewing-gum. *Journal of Food Science*, **61**, 636–638.

Egoroff, C. (2005) How to measure perception: a case study on automotive fabrics. ESN Seminar, 25–26 May, Madrid, Spain.

Eldeghaidy, S., Marciani, L., Pfeiffer, J.C. et al. (2011) Use of an immediate swallow protocol to assess taste and aroma integration in fMRI studies. *Chemical Perception*, **4**, 163–174.

Enneking, U., Neumann, C. & Henneberg, S. (2007) How important intrinsic and extrinsic product attributes affect purchase decision. *Food Quality & Preference*, **18**(1), 133–138.

Ennis, D.M. & Ennis, J.M. (2013) Analysis and thurstonian scaling of applicability scores. *Journal of Sensory Studies*, **28**(3), 188–193.

Faye, P., Bremaud, D., Durand Daubin, M., Courcoux, P., Giboreau, A. & Nicod, H. (2004) Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality & Preference*, **15**, 781–791.

Fillion, L. & Kilcast, D. (2001) *Towards a Measurement of Oral Tactile Sensitivity and Masticatory Performance: Development of Texture Tests*. Research Report No.781. Epsom: Leatherhead Research Association.

Gacula, M.C. (1997) *Descriptive Sensory Analysis in Practice*. Oxford: Wiley-Blackwell.

Gains, N. & Thomson, D.M.H. (1990) Contextual evaluation of canned lagers using repertory grid method. *International Journal of Food Science & Technology*, **25**, 699–705.

Giboreau, A., Navarro, S., Faye, P. & Dumortier, J. (2001) Sensory evaluation of automotive fabrics: the contribution of categorization tasks and non-verbal information to set-up descriptive method of tactile properties. *Food Quality & Preference*, **12**, 311–322.

Gonzáles-Tomás, L. & Costell, E. (2006) Sensory evaluation of vanilla-dairy desserts by repertory grid method and free choice profile. *Journal of Sensory Studies*, **21**, 20–33.

Gordin, H.H. (1987) Intensity variation descriptive methodology: development and application of a new sensory evaluation technique. *Journal of Sensory Studies*, **2**, 187–198.

Gower, J.C. (1975) Generalised Procrustes analysis. *Psychometrika*, **40**, 33–50.

Halpern, B.P. (1991) More than meets the tongue: temporal characteristics of taste intensity and quality. In: Lawless, H.T. & Klein, B.P. (eds) *Sensory Science Theory and Applications in Foods*. New York: Marcel Dekker, pp. 37–105.

Heymann, H., King, E.S. & Hofer, H. (2014) Classical descriptive analysis. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 9–40.

Hoggan, J. (1975) New product development, *Master Brewers Association of the Americas*, **12**(2), 81–86.

Holway, A.H. & Hurvich, L.M. (1937) Differential gustatory sensitivity to salt. *American Journal of Psychology*, **49**, 37–48.

Hort, J., Kemp, S.E. & Hollowood, T. (2017) *Time-Dependent Measures of Perception in Sensory Evaluation*. Oxford: Wiley-Blackwell.

Hort, J., Pfeiffer, J., Marciani, L. et al. (2008) Developing realistic fMRI protocols for the study of cross modal flavor perception. The impact of body position on flavor release and perception. In: Hofmann, T., Meyerhof, W. & Schieberle, P. (eds) *Recent Highlights in Flavor Chemistry & Biology*. Freising: Deutsche forschungsanstalt für lebensmittelchemie, pp. 47–52.

Hough, G., Contarini, A. & Munoz, A. (1994) Training a texture profile panel and constructing standard rating scales in Argentina. *Journal of Texture Studies*, **25**, 45–57.

Ishii, R., Kemp, S.E., Gilbert, A.N. & O'Mahony, M. (1997) Variation in sensory conceptual structure: an investigation involving the sorting of odor stimuli. *Journal of Sensory Studies*, **12**(3), 195–214.

Ishii, R., Chang, H.K. & O'Mahony, M. (2007) A comparison of serial monadic and attribute-by-attribute protocols for simple descriptive analysis with untrained judges. *Food Quality & Preference*, **18**(2), 440–449.

Ishii, R., Stampanoni, C. & O'Mahony, M. (2008) A comparison of serial monadic and attribute-by-attribute protocols for trained judges. *Food Quality & Preference*, **19**(2), 227–285.

Jack, F.R., Paterson, A. & Piggott, J.R. (1993) Relationships between rheology and composition of cheddar cheeses and texture as perceived by consumers. *International Journal of Food Science & Technology*, **28**, 293–302.

Jack, F.R., Piggott, J.R. & Paterson, A. (1994) Analysis of textural changes in hard cheese during mastication by progressive profiling. *Journal of Food Science*, **59**, 539–543.

Jaeger, S.R., Cadena, R.S., Torres-Moreno, M. et al. (2014) Comparison of check-all-that-apply and forced-choice Yes/No question formats for sensory characterisation. *Food Quality & Preference*, **35**, 32–40.

Johnsen, P.B. & Civille, G.V. (1986) A standardized lexicon of meat WOF descriptors. *Journal of Sensory Studies*, **1**, 99–104.

Johnsen, P.B. & Kelly, C.A. (1990) A technique for the quantitative sensory evaluation of farm raised catfish. *Journal of Sensory Studies*, **4**, 189–199.

Kemp, S.E. (2013) Consumers as part of food and beverage innovation. In: Martinez, M.G. (ed.) *Open Innovation in the Food and Beverage Industry*. Cambridge: Woodhead Publishing, pp. 109–138.

Kemp, S.E., Hollowood, T. & Hort, J. (2009) *Sensory Evaluation: A Practical Handbook*. Oxford: Wiley-Blackwell.

Kilcast, D. (2010) *Sensory Analysis for Food and Beverage Quality Control*. Cambridge: Woodhead Publishing.

Kim, K.O. & O'Mahony, M. (1998) A new approach to category scales of intensity I: traditional versus rank-rating. *Journal of Sensory Studies*, **13**(3), 241–149.

Kostyra, E., Baryłko-Pikielna, N., Dąbrowska, U. & Wasiak-Zys, G. (2008) Possibility of simultaneous measurement of pungency and leading flavour attributes in natural food matrices – temporal aspects. 3rd European Conference on Sensory and Consumer Research, 7–10 September, Hamburg, Germany.

Kuesten, C., Bi, J. & Feng, Y. (2013) Exploring taffy product consumption experiences using a multi-attribute time-intensity (MATI) method. *Food Quality & Preference*, **30**, 260–273.

Labbe, D., Schlich, P., Pineau, N., Gilbert, F. & Martin, N. (2009) Temporal dominance of sensations and sensory profiling: a comparative study. *Food Quality & Preference*, **20**, 216–221.

Labuza, T.P. & Schmidl, M.K. (1985) Accelerated of shelf-life testing of foods. *Food Technology*, **39**, 57–64, 134.

Laing, D.G. (1991) Characteristics of the human sense of smell when processing odour mixtures. In: Laing, D.F., Doty, R.L. & Breipohl, W. (eds) *The Human Sense of Smell*. Berlin: Springer-Verlag, pp. 241–259.

Larson-Powers, N. & Pangborn, R.M. (1978a) Descriptive analysis of the sensory properties of beverages and gelatins containing sucrose or synthetic sweeteners. *Journal of Food Science*, **43**, 47–51.

Larson-Powers, N. & Pangborn, R.M. (1978b) Paired comparison and time-intensity measurements of sensory properties of beverages and gelatins containing sucrose or synthetic sweeteners. *Journal of Food Science*, **43**, 41–46.

Lawless, H.T. (1989) Exploration of fragrance categories and ambiguous odors using multidimensional scaling and cluster analysis. *Chemical Senses*, **14**, 349–360.

Lawless, H.T. (2013) *Quantitative Sensory Analysis: Psychophysics, Models and Intelligent Design*. Oxford: Wiley-Blackwell.

Lawless, H.T. & Heymann, H. (2010a) Descriptive analysis. In: *Sensory Evaluation of Food: Principles and Practices*, 2nd edn. New York: Springer, pp. 227–258.

Lawless, H.T. & Heymann, H. (2010b) Time intensity analysis. In: *Sensory Evaluation of Food: Principles and Practices*, 2nd edn. New York: Springer, pp. 179–202.

Lawless, H.T. & Heymann, H. (2010c) Scaling. In: *Sensory Evaluation of Food: Principles and Practices*, 2nd edn. New York: Springer, pp. 149–178.

Lawless, H.T., Sheng, N. & Knoops, S.S.C.P. (1995) Multidimensional-scaling of sorting data applied to cheese perception. *Food Quality & Preference*, **6**, 91–98.

Lê, S., Lê, T.M. & Cadoret, M. (2014) Napping and sorted napping as a sensory profiling technique. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 197–214.

Lorho, G. (2005) Individual vocabulary profiling of spatial enhancement systems for stereo headphone reproduction. 119th Convention of the Audio Engineering Society, New York.

Lorho, G. (2010) Perceived quality evaluation and application to sound reproduction over headphones. PhD thesis, Aalto University, Department of Signal Processing and Acoustics, Espoo, Finland.

Meilgaard, M., Carr, B.T. & Civille, G.V. (2006) *Sensory Evaluation Techniques*, 4th edn. Boca Raton: CRC Press.

Methven L., Rahelu, K., Economoua, N. et al. (2010) The effect of consumption volume on profile and liking of oral nutritional supplements of varied sweetness: sequential profiling and boredom tests. *Food Quality & Preference*, **21**, 948–955.

Meyners, M. & Castura, J.C. (2014) Check-all-that-apply questions. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 271–306.

Morzel, M., Sheehan, E.M., Delahunty, C.M. & Arendt, E.K. (1999) Sensory evaluation of lightly preserved salmon using free choice profiling. *International Journal of Food Science & Technology*, **34**, 115–123.

Moskowitz, H.R., Stanley, D.W. & Chandler, J.W. (1977) The eclipse method: optimising product formulation through a consumer generated ideal sensory profile. *Canadian Institute of Food Science Technology Journal*, **10**, 161–168.

Munoz, A.M. (2002) Sensory evaluation in quality control: an overview, new developments and future opportunities. *Food Quality & Preference*, **13**, 329–339.

Munoz, A.M., Civille, G.V. & Carr, B.T. (1992) *Sensory Evaluation in Quality Control*. New York: Van Nostrand Reinhold.

Murray, J.M., Delahunty, C.M. & Baxter, I.A. (2001) Descriptive sensory analysis: past, present and future. *Food Research International*, **34**, 461–471.

Neilson, A.J. (1957) Time-intensity studies. *Drug and Cosmetic Industry*, **80**, 452–453, 534.

Neilson, A.J., Ferguson, V.B. & Kendall, D.A. (1988) Profile methods: flavour profile and profile attribute analysis. In: Moskowitz, H.R. (ed.) *Applied Sensory Analysis of Foods*. Boca Raton: CRC Press, pp. 21–41.

Ng, M., Lawlor, J.B., Chandra, S., Chaya, C., Hewson, L. & Hort, J. (2012) Using quantitative descriptive analysis and temporal dominance of sensations analysis as complementary methods for profiling commercial blackcurrant squashes. *Food Quality & Preference*, **25**, 121–134.

Pagès, J. (2003) Direct collection of sensory distances: application to the evaluation of ten white wines of the Loire Valley. *Sciences des Aliments*, **23**, 679–688.

Pagès, J. (2005a) Collection and analysis of perceived product inter-distances using multiple factor analysis: application to the study of 10 white wines from the Loire Valley. *Food Quality & Preference*, **16**, 642–649.

Pagès, J. (2005b) Analyse factorielle multiple et analyse procusteenne. *Revue de Statistique Appliqué*, **LIII**, 61–86.

Pagès, J., Cadoret, M. & Lê, S. (2010) The Sorted Napping: a new holistic approach in sensory evaluation. *Journal of Sensory Studies*, **25**, 637–658.

Pecore, S.D., Rathjen-Nowak, C. & Tamminen, T. (2011) Temporal order of sensations. 9th Pangborn Sensory Science Symposium, Toronto, Canada, 4–8 September.

Perrin, L., Symoneaux, R., Maitre, I., Asselin, C., Jourjon, F. & Pagès, J. (2008) Comparison of three sensory methods for use with the Napping procedure: case of ten wines from Loire valley. *Food Quality & Preference*, **19**, 1–11.

Petit, C. & Vanzeveren, E. (2014) Adoption and use of flash profiling in daily new product development: a testimonial. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 335–344.

Piggott, J.R. (2000) Dynamism in flavour science and sensory methodology. *Food Research International*, **33**, 191–197.

Piggott, J.R. & Watson, M. P. (1992) A comparison of free-choice profile and the repertory grid method in the flavor profile of cider. *Journal of Sensory Studies*, **7**, 133–145.

Pineau, N. & Schlich, P. (2014) Temporal dominance of sensation (TDS) as a sensory profiling technique. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 269–306.

Pineau, N., Cordelle, S. & Schlich, P. (2003) Temporal dominance of sensations: a new technique to record several sensory attributes simultaneously over time. 5th Pangborn Sensory Science Symposium, 20–24 July, Boston, Massachusetts, USA.

Pineau, N., Schlich, P., Cordelle, S. et al. (2009) Temporal dominance of sensations: construction of the TDS curves and comparison with time-intensity. *Food Quality & Preference*, **20**, 450–455.

Piqueras-Fiszman, B. (2014) Open-ended questions in sensory testing practices. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 247–268.

Poinot, P., Grua-Priol, J., Arvisenet, G. et al. (2007) Optimisation of HS-SPME to study representativeness of partially baked bread odorant extracts. *Food Research International*, **40**, 1170–1184.

Popper, R. (2014) Use of just-about-right scales in consumer research. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 137–156.

Popper, R. & Heymann, H. (1996) Analyzing differences among products and panelists by multidimensional scaling. In: Naes, T. & Risvik, E. (eds) *Multivariate Analysis of Data in Sensory Science*. Amsterdam: Elsevier, pp. 159–184.

Qannari, E.M., Courcoux, P., Taylor, Y., Buck, D. & Greenhoff, K. (2010) Statistical issues relating to hierarchical free sorting task. 10th Sensometric Meeting: Past, Present, Future, 25–28 July, Rotterdam, The Netherlands.

Richter, V.B., de Almeida, T.C.A, Prudencio, S. & Benassi, M.T. (2010) Proposing a ranking descriptive sensory method. *Food Quality & Preference*, **21**(6), 611–620.

Risvik, E., McEwan, J.A., Colwill, J.S., Rogers, R. & Lyon, D.H. (1994) Projective mapping: a tool for sensory analysis and consumer research. *Food Quality & Preference*, **5**, 263–269.

Rodrigue, N., Guillet, M., Fortin, J. & Martin, J.F. (2000) Comparing information obtained from ranking and descriptive tests of four sweet corn products. *Food Quality & Preference*, **11**(1), 47–54.

Ross, C.F. (2009) Sensory science at the human–machine interface. *Trends in Food Science & Technology*, **20**, 63–72.

Saint Eve, A., Paci-Kora, E. & Martin, N. (2004) Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality & Preference*, **15**, 655–668.

Seo, H.S., Lee, M., Jung, Y.J. & Hwang, I. (2009) A novel method of descriptive analysis on hot brewed coffee: time scanning descriptive analysis. *European Food Research & Technology*, **228**, 931–938.

Sieffermann, J.M. (2002) Flash profiling. a new method of sensory descriptive analysis. AIFST 35th Convention, July 21–24, Sydney, Australia.

Sjostrom, L.B. (1954) The descriptive analysis of flavor. In: *Food Acceptance Testing Methodology*. Chicago: Quartermaster Food and Container Institute, pp. 4–20.

Stampanoni, C.R. (1994) The use of standardized flavor languages and quantitative flavor profiling technique for flavored dairy products. *Journal of Sensory Studies*, **9**, 383–400.

Stoer, N. & Lawless, H.T. (1993) Comparison of single product scaling and relative-to-reference scaling in sensory evaluation of dairy products. *Journal of Sensory Studies*, **8**, 257–270.

Stone, H. (2014) Alternative methods of sensory testing: advantages and disadvantages. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 27–52.

Stone, H., Sidel, J.L., Oliver, S., Woolsey, A. & Singleton, R.C. (1974) Sensory evaluation by Quantitative Descriptive Analysis. *Food Technology*, **28**, 24–33.

Stone, H., Bleibaum, R.N. & Thomas, H.A. (2012) *Sensory Evaluation Practices*, 4th edn. San Diego: Academic Press.

Symoneaux, R. & Galmarini, M.V. (2014) Open-ended questions. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 307–332.

Szczesniak, A.S. (1963) Classification of textural characteristics. *Journal of Food Science*, **28**, 385–389.

Szczesniak, A.S., Brant, M.A. & Friedman, H.H. (1963) Development of standard rating scales for mechanical parameters of texture and correlation between the objective and the sensory methods of texture evaluation. *Journal of Food Science*, **28**, 397–403.

Szczeniak, A.S., Loew, B.J. & Skinner, E.Z. (1975) Consumer texture profile technique. *Journal of Food Science*, **40**(6), 1253–1256.

Talavera-bianchi, M., Edgar Chambers, I.V. & Chambers, D.H. (2010) Describing flavor using fewer and simpler 'HITS' (high identity traits) profiling: an example with cheese. *Journal of Sensory Studies*, **25**, 481–493.

Tarea, S., Cuvelier, G. & Siefffermann, J.M. (2007) Sensory evaluation of the texture of 49 commercial apple and pear purees. *Journal of Food Quality*, **30**, 1121–1131.

Tárrega, A. & Tarancón, P. (2014) Free-choice profiling combined with repertory grid methods. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 157–174.

Teillet, E. (2014a) Projective sensory positioning methodologies. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 255–270.

Teillet, E. (2014b) Polarized sensory positioning (PSP) as a sensory profiling technique. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 215–226.

Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010) Sensory methodologies and the taste of water. *Food Quality & Preference*, **21**, 967–976.

Thomson, D.M.H., Crocker, C. & Marketo, C.G. (2010) Linking sensory characteristics to emotions: an example using dark chocolate. *Food Quality & Preference*, **21**, 1117–1125.

Tilgner, D.J. (1962a) Anchored sensory evaluation tests. *Food Technology*, **16**, 47–50.

Tilgner, D.J. (1962b) Dilution tests for odor and flavour analysis. *Food Technology*, **16**, 26–29.

Valentin, D., Chollet, S., Lelievre, M. & Abdi, H. (2012) Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, **47**(8), 1563–1578.

Van Trijp, H.C.M., Punter, P.H., Mickartz, F. & Kruithof, L. (2007) The quest for the ideal product: comparing different methods and approaches. *Food Quality & Preference*, **18**, 729–740.

Varela, P. & Ares, G. (2012) Sensory profiling, the blurred line between sensory and consumer science. A review of novel method for product characterization. *Food Research International*, **48**, 893–908.

Varela, P. & Ares, G. (2014) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press.

Williams, A.A. & Arnold, G.M. (1985) A comparison of the aromas of six coffees characterised by conventional profiling, free-choice profiling and similarity scaling methods. *Journal of the Science of Food & Agriculture*, **36**, 204–214.

Williams, A.A. & Langron, S.P. (1984) The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food & Agriculture*, **35**, 558–568.

Worch, T. & Punter, P.H. (2014a) Ideal profiling. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 85–136.

Worch, T. & Punter, P.H. (2014b) Ideal profiling as a sensory profiling technique. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Cambridge: Woodhead Publishing, pp. 307–332.

Worch, T., Lê, S. & Punter, P. (2010) How reliable are the consumers? Comparison of sensory profiles from consumers and experts. *Food Quality & Preference*, **21**, 309–318.

Worch, T., Lê, S., Punter, P. & Pagès, J. (2012) Assessment of the consistency of ideal profiles according to non-ideal data for IPM. *Food Quality & Preference*, **24**, 99–110.

Worch, T., Lê, S., Punter, P. & Pagès, J. (2013) Ideal profile method (IPM): the ins and outs. *Food Quality & Preference*, **28**(1), 45–59.

Zimoch J.F. & Findlay, C.J. (2006) Effective discrimination of meat tenderness using dual attribute time intensity. *Journal of Food Science*, **63**, 940–944.

Zwillinger, S.A. & Halpern, B.P. (1991) Time-quality tracking of monosodium glutamate, sodium saccharin and a citric acid–saccharin mixture. *Physiology & Behavior*, **49**, 855–862.

CHAPTER 2
General Considerations

Sylvie Issanchou

2.1 General Introduction

Descriptive analysis is a sensory method providing a sensory description of products. Generally, this description is not only qualitative but also quantitative. The sensory profile is defined (AFNOR 2009) as a 'description of the sensory properties of a sample consisting of the sensory attributes, in the order of perception, and with assignment of an intensity value for each attribute' (the term 'descriptor' is also used and is synonymous with 'attribute'). This means that the assessors (also named 'panellists') have to rate the perceived intensity of each attribute. This is a method defined as 'objective' in the sense that the methodology aims to minimize the effect of personal opinions. It also means that the hedonic dimension of the perception is not supposed to influence the evaluation of the qualitative and quantitative dimensions of the different perceptions elicited by the physico-chemical properties of the product. As for other 'objective' sensory methods, descriptive analysis is carried out in well-defined conditions.

The sensory description concerns all relevant sensory modalities for the type of product which is studied. For example, appearance, odour, flavour and texture are generally evaluated for food products, odour only for perfumes, texture only for fabrics, odour and texture for care products.

The size of the panel is usually quite small (8–24), but the assessors are selected for their abilities and trained to perform this task, and the panel leader monitors their performance.

As the usual approaches require a selected and well-trained panel, this method is time- and money-consuming. Thus, several more rapid methods have been developed and are more and more frequently used. Some of these rapid methods keep the same essential features of the traditional method, in the sense that perceived intensity for different attributes is collected. In the free-choice profiling method (see Chapter 13) and the flash profiling method (see Chapter 14), the difference is that each panellist chooses her/his own attributes. This avoids

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

the quite long step required for the definition of the vocabulary and for training the panellists to use each attribute similarly.

In the classic approach where the same attributes are used by all panellists, data are analysed by univariate methods, and in particular by analysis of variance. Whatever the technique, data can be analysed by multivariate methods, but the method depends on the profiling technique. However, in all cases these multivariate approaches produce a map, most often in two or three dimensions, where all samples which have been evaluated are positioned. The sorting methodologies (see Chapter 15) are different as panellists are not asked to rate intensity of different attributes but are simply asked to sort the different samples according to their own criteria to form groups of similar samples.

2.2 Aims of Descriptive Analysis

Descriptive analysis is not relevant if the differences between samples are believed to be very small, that is, if the experimenter is not sure that these differences are perceptible. In such cases, discriminative tests must be used (Kemp et al. 2009). Descriptive analysis can be used for different purposes such as:

- comparing different varieties or different cultural practices in the vegetal production domain
- comparing different species or different rearing practices in the animal production domain
- defining a standard for manufactured products or ingredients and raw materials
- assessing the quality of ingredients and raw materials
- studying the impact of a change in the origin of raw materials or ingredients, or in the product formulation
- studying the impact of different parameters of the process
- obtaining useful information for determining the 'best before date' and to compare different packagings in shelf-life testing
- assessing competitive products.

Results from descriptive analysis can also be related to physico-chemical data, to help determine which chemical and/or physical parameters explain sensory perception. When a good prediction is obtained from a set of samples, the correlation between predicted and observed values must be checked with another set of samples before replacing the sensory measurement with physico-chemical measurements.

Results from descriptive analysis can also be related to hedonic data obtained with a panel of consumers. Such analyses determine the importance of the different attributes for liking either for the panel or for subgroups of consumers; they also enable definition of the intensity of the different attributes to obtain an optimal product for a defined target of consumers. Then, if relationships between ingredients and/or process and sensory profile have also been identified, it becomes possible to optimize the product.

2.3 Choices of Methodology

When the product to be studied is a key product for the company and when the objective is to set up a tool to follow the quality of the production over time, the recommended option is to select conventional profiling, that is, a profile where a specific vocabulary is developed for this product and used by the panel. When the panel has to describe different types of products and to work long term, the Spectrum™ Method is the most appropriate as the scales are universal and this enables the intensity of different attributes to be rated for a large range of products (see section 2.4.4).

When the objective is to obtain quick results on a given type of product, the alternative rapid methods are more appropriate. However, results obtained by free choice profiling, flash profiling or sorting can only be analysed by a multivariate method, enabling a map of the different samples to be derived. This means that these methods can only be employed when the number of samples is at least equal to five. However, as flash profiling is based on a comparison of the samples on each attribute, the number of samples cannot be very high (Varela & Ares 2012) unless only appearance or tactile characteristics are evaluated. If odour or flavour are dimensions to be evaluated, the range of samples which can be tested is 6–8. Moreover, it must be noted that the interpretation of the sensory space derived from the multivariate analysis of the results obtained by free choice or flash profiling could be difficult due to the diversity of descriptors used by the panellists (Varela & Ares 2012), especially for flavour description. With sorting methods, the sensory map can be interpreted only with external data, such as the composition or the process of the samples, as panellists are not required to describe samples. These data can be insufficient to interpret the different dimensions of the sample map. In this case, it is only possible to identify groups of similar samples. To avoid this problem, panellists can be asked, once they have sorted the samples into groups, to describe these different groups with one or two words (Lawless et al. 1995).

Thus, the rapid methods are more appropriate for applications in product development than for applications in quality control. They can, in particular, answer questions such as:

- Amongst the different variants of a product present on the market, which are similar?
- Are there variants similar to the market leader?

These rapid methods can be used as the first step in a reformulation process to identify which parameters are the most influential on the sensory characteristics. Then, results can be used to choose the parameters and the levels of these parameters for a more detailed study where a conventional profiling can be used. A flash profiling approach can also be useful during the first sessions of vocabulary development for a conventional profiling. To conclude, as noted by Delarue and Sieffermann (2004), 'It would be misleading to consider Flash profile as a substitute for conventional profiling … Furthermore, these two methods do not fulfil exactly the same objectives'.

In the following section, we will consider the case where a conventional profiling is used, that is, with all panellists using the same set of descriptors.

2.4 Generic Procedure for Descriptive Analysis

Carrying out descriptive analysis on a product requires several steps before data collection that can be used for one of the aims previously defined. The first step is to select assessors; after this several steps take place in parallel:

- generation of an attribute list
- choice of assessment protocol, which is particularly important for the evaluation of texture
- choice of references or particular samples which illustrate the attributes, in particular the flavour attributes, and
- choice of the type of scale and of the labels on the scale for each attribute.

After these steps, rating scales are defined, assessors are trained to evaluate the intensity of the different attributes, and their performance is monitored. When performance is judged to be sufficiently good, it is possible to undertake studies.

After each study, data are analysed. First, panel performance can be checked. Second, univariate and/or multivariate analyses are carried out to determine the differences between samples. Finally, a report may be written to present the results.

2.4.1 Panel Selection

This section presents key stages and some key points concerning panel selection but readers should refer to Chapter 3 for a full description. A panel can be recruited either within the company or outside the company. In both cases, motivation and availability are the first criteria to consider. The greatest problem with an internal panel is that participation in sensory tests is a supplementary task. Thus, long-term availability can be difficult with an internal panel if there is no global agreement, in particular amongst the managerial staff, to promote sensory analysis and to recognize the importance of the work of the sensory panellists. Moreover, the number of candidates is generally too low to perform a selection based on the candidates' abilities, which is essential as training cannot correct some deficiencies. The advantages of an external panel are a larger number of candidates, unless the company is located in a small city, a higher availability, and less potential bias as external subjects are not aware of the purpose of the experiments, and there are no hierarchical relationships between the panellists.

Concerning cost, an external panel is often considered to cost more than an internal one, but company employees are paid and so the average salary of these employees must be compared to the salary or fee given to the external panel.

Even with external assessors, long-term availability is not guaranteed. Thus, the long-term availability must be checked and considered as the first recruitment criterion. The absence of health problems or regular medications which could affect the different senses must be checked, as well as the absence of allergy for working with flavours, perfumes or foods. Dental health must be checked if the panel has to work on perception of textural characteristics in mouth. Candidates must also be asked if they consume and like the type of products that they will have to test. These criteria can be checked during a phone interview.

Candidates who satisfied these first criteria are then invited to participate in one or two selection sessions. Descriptive analysis is clearly the most complex task in sensory analysis and requires additional abilities compared to those required for difference tests or ranking tests. Thus, several sensory and non-sensory abilities must be evaluated during the selection stage:

- normal acuity for the different senses which will be involved in the future task
- ability to discriminate stimuli which are known to differ in quality or intensity
- ability to express oneself verbally about sensory perception
- ability to understand the tasks
- ability to concentrate on a task
- ability to follow instructions
- ability to work within a group.

For the evaluation of sensory abilities, the stimuli used must be chosen depending on the type of product that the panel will have to test. For example, a subject's ability to rank sweet aqueous solutions according to their sweetness is not related to his ability to rank food samples according to their sweetness. Thus, the procedure needs to be product specific. When simple stimuli such as tastants are used to determine sensitivity, relevant compounds which are present in the actual products to be studied should be used. For example, for a taste recognition test, lactose, lactic acid and L-leucine could be chosen for a panel dedicated to cheese evaluation to illustrate sweet, acid and bitter tastes respectively (Issanchou et al. 1995).

The most frequent approach used in the selection procedure is to eliminate candidates with the poorest results. As underlined by Sauvageot (1975), the abilities of the panel will then depend on the abilities and, especially, on the number of candidates. The ideal way is to require a qualifying score for each test of selection. However, results for different tests aiming to explore different senses and different types of abilities are not highly correlated (Issanchou et al. 1995), and with about 100 participants, it is difficult to select only the candidates with a high level of performance on each test. Thus, some weaker abilities of some panellists have to be compensated by other panellists in order to build an efficient sensory tool.

It is also necessary to decide on the panel size. In the literature, the descriptive panel size varies from eight, the recommended minimum number according to

Table 2.1 Recommended number of panellists for a descriptive panel according to the sensory dimension.

Sensory dimension	Number of panellists
Visual	6
Texture	8
Taste	8
Odour	8
Aroma	10

Source: Adapted from Mammasse (2012).

Lyon et al. (1992), to 24 (Muir & Hunter 1991-1992). More generally, it varies between eight and 12. A small panel size can be compensated by more intensive training. However, training cannot enable all 'deficiencies' to be compensated. It is clear that if the panel is too small, the results are very dependent on individual sensitivities. This is particularly true in the olfactory domain where there are specific hyposmias. Thus, if the panel must be able to detect off-flavours, it is particularly important not to have too small a panel or to include, in the selection procedure, an evaluation of the candidates' sensitivity to the compounds responsible for off-flavours. Moreover, from a statistical point of view, the smaller the size of the panel, the larger the chance of not detecting a small difference. In a meta-analysis of the SensoBase based on 405 data sets of descriptive analysis (Mammasse 2012), the observed median of the number of panellists was 12 (10 and 12 being the most common size). Resampling in these 405 data sets was applied to determine adequate panel size; this consisted of removing k subjects from the N of the original panel and then measuring the loss of information in product comparisons. The recommendations given by the author on the basis of these analyses differ according to the type of attributes, as shown in Table 2.1. When the panel is required to work long term, it is also recommended to recruit a larger number in anticipation of some panellists giving up over time.

2.4.1.1 Testing Sensory Acuity

If panellists will have to evaluate colour, a colour blindness test (e.g. the Ishihara test) must be included in the selection test set.

Concerning olfactory acuity testing, the tests generally include absolute threshold measurement, and other tasks (discrimination, detection and/or identification) for odorants presented at suprathreshold concentrations. The 'Sniffin' Sticks' (Hummel et al. 1997) are well suited for the clinical assessment of olfactory dysfunction, but cannot be self-administered. The Lyon Clinical Olfactory Test (LCOT) (Rouby et al. 2011) is also designed to assess olfactory deficits but has the advantage that it is a self-administered test. The LCOT includes threshold measurement for two compounds, one stimulating the trigeminal nerve, and a combination of a detection and an identification task. In this LCOT test, supraliminal detection (i.e. detection at a concentration above the threshold level) in particular emerges as a simple tool to classify subjects as anosmic, hyposmic or normosmic. It must be noted that the identification task is quite easy and can be considered as only a first step in the selection procedure of candidates for a descriptive panel. The UPSIT test is another self-admininistered test for smell ability. It is based on an identification task and enables definition of the smell function in five levels (see Chapter 3).

Concerning taste acuity measurement, the tests generally use suprathreshold concentrations and are based on taste identification. However, it is important to note that naive subjects often confuse different tastes, and in particular bitterness and acidity. Thus, a first series of samples labelled with their respective taste can be presented as a training set before asking candidates to recognize the taste of a series of coded samples (see, for example, Issanchou et al. 1995).

2.4.1.2 Ability to Discriminate Stimuli

To evaluate the candidates' ability to detect differences between samples, using variants of the same type of product as the one they will have to evaluate is highly recommended. As it is important to limit the error of classifying a 'non-sensitive candidate' as 'sensitive', it is necessary to use a test with a quite low probability of a correct answer by chance, for example the two-out-of-five test (probability of right answer by chance = 1/10).

A ranking test can also be used to evaluate the candidates' ability to perceive differences in intensity for a given attribute by choosing or preparing samples varying, for example, in one ingredient known to be related to a sensory attribute which is quite easy to identify by naive panellists. For example, samples with different sugar contents, or with different fat contents, or with different concentrations of one flavour can be presented. However, this test is not the most important as this ability is one which can be improved by training.

2.4.1.3 Ability to Express Oneself Verbally About Sensory Perception

Finally, if a list of attributes is not yet established and the first task of the panel will be to generate attributes, it is important to check the candidates' ability to describe their sensory perceptions. This can be achieved by presenting samples to the candidates simultaneously and asking them to compare the samples for specified dimensions (odour, texture, etc.); they should then be asked to describe the samples without using hedonic terms. This ability is particularly important for free choice or flash profiling as each panellist will not be helped by the other members of the panel to generate the descriptors.

To test candidates' ability to express themselves verbally about their odour perception, volatile compounds or odorous preparations giving aroma notes likely to be found in the product to be studied should be chosen. When a list of descriptors for the product to be studied has been published, some of the references can be used for this test; see, for example, the wine aroma wheel (Noble et al. 1987) and the beer flavour wheel (Meilgaard et al. 1982). To test candidates' ability to express themselves verbally about their texture perception, products which are particularly characteristic of the attributes should be chosen (see AFNOR 1995b for examples of foods).

2.4.1.4 Ability to Memorize Odours

If the olfactory component is to be included in the descriptive analyses that the panel will have to perform, it is important to check candidates' abilities to memorize odours. This ability seems particularly important to test as it does not appear to be improved by training (Lesschaeve & Issanchou 1996). These authors used a recognition test. Sixteen odorants (target set) were presented successively during the first session to subjects for familiarity rating and a written description. The recognition session took place 7 days later with 32 odorants (including the 16 of the target set and 16 'new' odorants). Subjects were asked, for each odorant, to indicate if it was presented during the first session (yes/no). This task has the advantage that it combines the evaluation of the ability to describe and to memorize odours. Moreover, it evaluates the memory ability close to the one that normally occurs when experiencing products in everyday life and this could explain the absence of a training effect.

2.4.1.5 Ability to Concentrate on a Task

The Bourdon test T.IB. (Swets & Zeitlinger, Lisse, NL) can be used for this purpose. This test was originally developed to evaluate the concentration ability of airline pilots. The test sheet (A3 format) has 50 lines. Each line contains 25 groups of three, four or five black dots. Participants have 8 sec per line, and for each line they must mark by a vertical line all groups of four points. The score corresponds to the average number of groups of four dots correctly ticked over all lines.

2.4.2 Initial Training

Once the panellists have been selected, it is important to give them some basic training; this should include the functioning of the different senses that they will have to use for their descriptive task. It is particularly important that they are aware of the interindividual differences, except for free choice and flash profiling. It is also important to give them the list of instructions they will have to follow such as:

- to indicate if they use any medication
- to be on time at each session
- to keep quiet during the session
- to read attentively all instructions given at each session, in particular those concerning break or rinsing procedures between the evaluation of different samples

- not to use perfumes or other perfumed body products if they have to work on odour or flavour
- not to smoke or consume strong flavour products such as gums, sweets, coffee at least 1 hour before the session if they have to work on odour or flavour.

Panellists should also be told to inform the panel leader of any event that could affect their sensory aibilities such as a cold, medical treatment and pregnancy.

2.4.3 Vocabulary Development

In the conventional profile, the list of descriptors is identical for all the subjects, whereas with the free choice procedure proposed by Williams and Langron (1984), each subject uses her/his own list of descriptors. In this chapter, the case of the development of a common list of descriptors is considered.

Once panellists are selected, vocabulary development is the first step to conduct when there is no previously established list of descriptors on the product to be studied. Even if a list already exists, it is important not to impose this pre-established list on the panel. This is particularly true if the list has been established by product technologists who use a technical vocabulary that is difficult for non-technologists to understand, even after sensory training. It is also quite common that the range of samples used to establish the list is different from the range of samples that the panel will have to evaluate. For example, Vangtal and Hammond (1986) working on Swiss-type cheeses proposed a list of eight descriptors established in a previous study, but this list was expanded to 16 descriptors during the 10 1-hour training sessions as the cheeses used in the experiment contained f1avour notes not present in the established list. When using a pre-established list, it is advisable to present panellists with a range of samples and ask them to identify sample(s) which is(are) particularly representative of each attribute and sample(s) where each attribute is not present or at a low intensity.

Typically, generation of descriptors is carried out over several sessions as follows: assessors are presented with the largest possible range of samples compared to what they are supposed to encounter later, at a rate of a few samples per session. During each session, each assessor works independently and is asked to write down the largest possible number of descriptors characterizing each sample on the sensory dimensions defined by the panel leader. After this individual work, terms are gathered, discussed and compared. As panellists are required not to use hedonic terms, the panel leader asks them to eliminate hedonic terms such as good, bad but also tasty. Such discussions at the end of each session also enable the discarding of imprecise terms, elimination of obvious synonyms and definition of a protocol to evaluate some characteristics, in particular those refering to texture. This is also an opportunity to start work on definitions which are particularly useful for descriptors related to appearance and texture. When the panel leader notices that there is some disagreement between panellists on one descriptor, s/he must encourage discussion and, for descriptors related to flavour, look for a standard to be presented at the next session.

According to different authors (Civille & Lawless 1986; Meilgaard et al. 2007; Rainey 1986; Stampanoni 1993), qualitative standards are necessary to achieve sensory concept alignment. A standard is defined as 'any chemical, spice, ingredient, or product' (Rainey 1986). It must be easily prepared and reproducible. An improvement of alignment when using multiple standards for a given attribute was also observed (Ishii & O'Mahony 1991). With the Spectrum Method, several standards are used for a given attribute and these standards also represent different levels of intensity along the scale. For taste attributes, one compound at different concentrations is used, but different products are used for texture and aroma attributes (as described in Meilgaard et al. 2007).

It appears that it could be difficult to generalize sensory concepts to products when external standards such as chemicals, spices or ingredients are used and not added in the product type to be evaluated. Noble et al. (1987) suggested presenting standards in neutral white or red wines. It was also found that when describing the odour of orange juices, a panel who had learned descriptors from different samples of orange juice performed better than panels who were presented with one or three external standards for each (Sulmont et al. 1999).

It is generally recommended to ask panellists to decompose their perceptions into dimensions as monodimensional and independent as possible. However, as pointed out previously (Giboreau et al. 2007), this rule could be questioned. In fact, as explained by these authors:

> This requirement actually relies on an implicit theory of cognition that assumes that an object can be identified through the description of a sum of properties, independent one from the other (and mainly corresponding to physical or chemical dimensions).

However, this assumption is not relevant as many recent works have demonstrated the perceptual interactions between different sensory modalities, and in particular between taste and smell (see Delwiche 2004 for a review). There are also neuroimaging data revealing taste and smell integration in the brain (Small et al. 2004). Thus, considering multidimensional attributes could be useful and even essential and, as mentioned by Delwiche (2004), this is particularly important when descriptive analysis is conducted with the aim of relating descriptive data to consumers' preference data.

The ISO standard (AFNOR 1995a) discourages the use of terms which describe the product by itself, such as carrot aroma for a carrot soup or purée. It is clear that panellists must be encouraged to go further in their description and that the carrot aroma attribute will not be very informative to describe different carrots. However, it could be sufficient to describe the flavour of vegetable soups or purées made with a mixture of vegetables. Moreover, the panel leader must also consider the human limits in olfactory information processing (Lawless 1999). Thus, it could be relevant to keep a global term, otherwise the description of the products may be incomplete. The ISO standard also recommends that non-relevant terms, such as 'acid' for describing odour, are eliminated. It is clear that this does not correspond to a 'pure' taste perception. However, it is known that the information from our different senses merges and is integrated into the brain. Thus, we can have learned to associate a given odour to an acid taste, leading us to recognize a given odour as acid. As we have no specific vocabulary to describe odours, it could be useful to use a term which normally refers to a gustatory perception, and eliminating this term can lead to missing important information.

Nevertheless, before deciding to include a term which seems *a priori* irrelevant, the panel leader must encourage discussion and ask panellists to describe their perception more precisely, and should help them if necessary. In the example mentioned above, the panel leader could invite panellists to describe this acid odour, and if they have difficulties, to ask them, for example, if it is a pungent odour or if they can quote a product which is typical of this acid odour.

When writing a definition, it is important to avoid circularity. For example, 'burning' should be defined as 'describes a product causing a sensation of heat in the buccal cavity (as produced by chilli and pepper)' (AFNOR 2009) and not as 'a product which burns'.

> It should rather be recommended to favor description with paraphrases, analogies, synonyms, antonyms, as these processes provide explicit information … In addition, reference to products could be used as a complement to the definition … (Giboreau et al. 2007)

Some of the recommendations given by a lexicographer, Landau (1989), and presented by Giboreau et al. (2007), are worth noting:

- to begin the definition with the most important elements of meaning
- to avoid difficult wording and try to write simple phrasing
- to prefer brief definitions
- not to use ambiguous words.

2.4.4 Response Scale

The panel leader must choose the type of response scale to be used for asking panellists to record their perceived intensity of each descriptor. Two main options are possible: a discrete category scale or a continuous linear scale (also called graphic linear scales or visual analogue scales).

2.4.4.1 Mathematical Properties of Data Obtained with Different Types of Scales

Category scales may yield data which have only ordinal properties. Thus, a score equal to 5 is higher than a score of 4 but, for example, the difference in intensity between 4 and 5 is not equal to the difference in intensity between 3 and 4. However, in most cases category scales, such as line scales, are supposed to generate data with interval properties. This means that equal numerical differences on the response scale correspond to equal intensity differences.

Consequently, mean values can be calculated. In practice, verification of this property is not carried out. Deviation from this assumption would, however, be bigger in the case of a discrete scale, in particular if a label is used to describe the intensity of each level of the response scale.

Data obtained with category and line scales are not ratio-scale data. This means, for example, that a sample with a sweetness intensity equal to 8 cannot be considered twice as sweet as a sample with a score equal to 4. Magnitude estimation scaling is a procedure which is supposed to generate ratio-scale data. With this procedure, panellists are required to assign a numerical value to each sample in such a way that these numbers reflect relative proportions of perceived intensities for the attribute in the samples. In practice, panellists receive a first sample and are asked to choose a number for the intensity of the attribute to be assessed (the number for this first sample can also be assigned by the experimenter; it is referred to as the modulus). Panellists are then asked to assign numbers to all subsequent samples in proportion to the number assigned to the first sample.

With category and line scales, individuals tend to discard extreme ends of the response scale. This could be more problematic with category scales, in particular if they have a low number of levels. Thus, line scales are preferred but require that panellists enter their responses on a computer. When opting for a discrete scale, it is recommended to choose a 7- or 9-point scale. One advantage of magnitude estimation is that this procedure solves the problem of the so-called floor and ceiling effects (i.e. the avoidance of using the extreme ends of the scale).

A category scale is easier to use than a continuous scale and could be appropriate at the beginning of training. Magnitude estimation also requires some training. This is often done by asking panellists to score the length of lines or the area of different surfaces.

In practice, magnitude estimation is rarely used when panellists are required to perform a descriptive analysis for a large number of attributes. Magnitude estimation has been mainly used for studying simple systems, such as one tastant or one volatile compound at different concentrations in a given solvent. In such cases, it is possible to establish the psychophysical function corresponding to the relationship between the concentration of the stimulus (C) and the perceived intensity (I) and to determine the key parameter of this function, n, the exponent of the psychophysical function ($I = k\ C^n$), that is, the slope of the linear relationship: $\log I = n \log C + \log k$, k being a constant.

A hybrid technique, called the labelled magnitude scale (LMS), has been proposed (Green et al. 1993). The response scale is a vertical or horizontal line with verbal labels (barely detectable, weak, moderate, strong, very strong, and strongest imaginable) placed along this line according to a quasi-log spacing and with or without ticks and numbers along the line scale. In Green et al.'s work (1993), panellists were asked to express the perceived intensity relative to oral sensations of all kinds that they have experienced in daily life, including

> 'sensations as mild and strong tastes, the hotness, coldness and texture of foods and beverages, and even sensations like burning your mouth or biting your tongue'.

Bartoshuk et al. (2003) proposed a generalized version of this scale (gLMS) by labelling the maximum of the scale with 'strongest imaginable sensation of any kind'. To emphasize the general nature of the top of the scale, participants were asked

> 'to think of intense experiences from different modalities (e.g. staring at the sun, hearing a nearby jet plane take off, experiencing intense pain)'.

The gLMS is not only supposed to generate ratio-scale data and to avoid ceiling effects but also to give absolute values of perceived intensities within and across modalities and thus enables comparison of results between different attributes but also between individuals. Consequently, as pointed out by Hayes et al. (2013a),

> if 'two individuals rate the sweetness of a sucrose sample as 20 and 30 on the gLMS, but both rate the brightness of the sun near 60, we are more confident the two individuals perceive the sucrose differently'.

However, it is not sure that these properties are verified and, as underlined by Schifferstein (2012), the cognitive process used by participants when using the LMS or gLMS is not known. Moreover, Hayes et al. (2013a) found that participants clustered their responses near the verbal labels and that providing explicit written instructions to rate between the adjectives was not successful in reducing this bias. It also appears that using the gLMS is not so obvious as when asked to rate the intensity of different remembered sensations; 11% of participants failed to rate the intensity of a whisper, a conversation and the loudest sound experiences in this expected order (Bartoshuk et al. 2003).

In practice, the gLMS has been extensively used for studying perception of different taste and trigeminal coupounds and in particular to compare PROP and non-PROP tasters (see section 2.5.1.1), but not so much in descriptive analysis.

2.4.4.2 Quantitative Frame of Reference

The question of the quantitative frame of reference is a key point when using a scale. As explained by Muñoz and Civille (1998), it represents 'boundaries or limits' that panellists use when rating the perceived intensity of attributes. The quantitative frame of reference is set by defining the highest intensity reference. As mentioned above, with the gLMS scale, panellists are asked to consider all kinds of sensations. If panellists follow this instruction, comparisons between attributes, between product categories and between groups of panellists are possible. In some variants of the LMS, panellists are asked to restrict their frame of reference to the attribute to be evaluated, for example, the strongest imaginable sweetness, saltiness or bitterness they have ever imagined experiencing (Green et al. 1996), the wettest or the driest imaginable (Guest et al. 2007). In these last cases, comparisons between product categories and between groups of panellists are possible, but not comparisons across attributes.

When category and line scales are used, panellists are generally asked to rate the perceived intensity of each attribute considering their experience within the product category they have to evaluate. One objective of the training is to agree on products which are typical examples of different levels of intensity, and in particular of the highest intensity. The purpose is that panellists use all the scale and in particular its upper end. This is particularly important for odour and aroma descriptors. In fact, panellists have difficulties in using the whole scale for these descriptors (Muir & Hunter 1991-1992). This could be explained by the fact that odour and aroma characteristics are generally described using the product from which the odour originates (for example, rose odour) and the intensity of the descriptor in any product is generally quite low compared to the source (rose odour or aroma in a wine is lower than in a typical rose). If the frame of reference is the product category to be studied, one can suppose that a greater sensitivity of the scale is achieved but comparisons between product categories are not possible. With the Spectrum Method, intensities are rated on an absolute and universal basis, that is, the 'highest' intensity point on the scale is defined for all types of product categories. Additionally, these scales were designed to be equi-intense across attributes so that, for example, a 5 on sweetness scale and a 5 on salty scale should reflect the same intensity (Murray et al. 2001). Lawless and Heymann (1998) mentioned that

> 'this goal has been achieved for fragrance, aroma and flavor scales but not for texture scales'.

However, to the best of our knowledge, no published data support this feature, even for the different tastes.

Nevertheless, for a given attribute, comparisons between product categories as well as between different panels are possible. This method is based on the extensive use of reference points along their range. A reference sample or a sapid reference solution is assigned to each reference point. This method requires extensive training, from 35 to 75 hours after the vocabulary development stage (Murray et al. 2001). If the panel will have to work on different types of products, the Spectrum Method is the most appropriate choice. The primary disadvantage of this method could be that the ratings within a product category may only occupy narrow ranges of the whole scale (Muñoz & Civille 1998). In fact, it was found that 77–99% (depending on the studied tastes) of the intensity scores collected on 590 common French foods were located in the first third of the scales (Martin et al. 2014). However, according to Muñoz and Civille (1998), a well-trained panel can show differences of 0.5 points or less that are real, meaningful and possibly significant.

2.4.5 Training

Once a list of descriptors is established with a definition or references for each descriptor, panellists are trained to quantify the perceived intensity of each descriptor in a range of samples. This training is a key element in the conventional profile to achieve a panel which will be a reliable instrument.

It is important to give feedback on the results and performances at each session. This allows problems to be detected as soon as possible and thus reduces training time; this is also a good way to motivate panellists. With a computerized data acquisition system, this feedback can be given at the end of each session. Otherwise, it is given at the beginning of the next session. More immediate feedback displayed to panellists following the evaluation of several attributes appears to be an effective way to reduce training time (Findlay et al. 2006). With this procedure, when a panellist has scored the intensity of the attributes presented on a screen (up to 5), the panellist marks given on each linear scale remain on the screen and the feedback is given by displaying, for each attribute, the 90% confidence interval around the mean obtained by the panel during a previous training session.

Depending on the results obtained at one session, the panel leader decides the content of the next training session. For example, if large inter-individual differences are observed for some attributes, it is important to focus the next session on these problematic areas. Thus, at the next session, the definition of these attributes must be restated and appropriate samples to illustrate different levels of intensity of these problematic attributes must be presented.

2.4.6 Panel Performance and Monitoring

Permanent feedback does not exclude the need to monitor the performance of a conventional profiling panel before starting actual measurements. Then, performance monitoring should also occur at regular periods when the same panel is working for a long time on the same type of products or when new panellists are recruited to compensate for those who drop out.

Before the beginning of actual measurements, data analyses enable identification of any redundant descriptors; when redundant descriptors are identified by the panel leader, these results must be presented to the panel and discussed in order to decide if it is possible to eliminate descriptors and to choose which ones to keep in the final list. Criteria used to evaluate the panel performance (see later in this chapter) should also be considered for selecting one descriptor amongst several redundant ones. Of course, the descriptor with the highest repeatability and the highest levels of agreement between panellists and discrimination between samples should be preferred.

Compared to an instrumental measurement, evaluation of the reliability of the sensory instrument differs on two points. First, a sensory instrument is a group of individuals, so performances must be checked at both levels – individually and as a group. Second, the accuracy of the measurement is no longer applicable for a sensory instrument. In fact, a true response exists only in rare cases, such as for the ranking of solutions of sucrose in water according to their sweetness; in such a case, the experimenter can expect a ranking according to the concentration in sucrose.

Thus, the criteria that are important to evaluate are repeatability, consistency (i.e. the degree of agreement between panellists) and discrimination.

To evaluate discrimination ability at the assessor level, it is necessary to have replicates of several samples per assessor. As discrimination ability depends not only on the panellists but also on product variability, it is clear that the panellists' discrimination ability must be based on the same set of samples and, if possible, on the same batch. For an interlaboratory test, it is also necessary that each laboratory proposes the same set of samples (same products, same batches, same preparation if any) for comparing the performances of the different panels.

As mentioned in Chapter 4, it is really useful for the panel leader to have a rapid way to identify problems, such as panellists and descriptors for which performance is poor. Different softwares provide reports on panellist/panel performances. Visual representations can also be very useful and are easier for reporting results to panellists during the training than tables. In Chapter 4, an example based on a quality control chart approach is presented (Gatchalian et al. 1991).

An approach was proposed by Pineau et al. (2007) for following up the evolution of a panel working on the same type of product for many years. The approach used by these authors is based on a mixed model that takes into account the evolution of both panel and panellists in the same model. The performance index followed in this paper is the individual repeatability measured by standard deviation over replicates. However, other indices can be followed using the same approach. The method enables detection of whether some panellists perform better than others, and whether this difference remains the same or evolves over time. A control chart method is a very useful graphical tool for assessing time points for which a panel's or panellist's performance differs from the overall evolution over time.

2.4.7 Measurements

2.4.7.1 Samples

As in all sensory tests, samples are presented blind, that is, without any information about their origin, process, etc. They are generally coded with random three-digit numbers to avoid any expectation error (see section 2.5.2.2). Ideally, a double-blind presentation should be used, such that the panel leader does not know during the session the correspondence between the codes and the nature of the samples. However, this practice, recommended by Depledt (2009), is not commonly applied.

The method of sample preparation is defined after preliminary testing and could be finalized during the training. However, there are two key points to consider concerning sample preparation:

- to present all samples in the same way, that is, the same cooking if any, same vessel, same serving temperature, same quantity and same shape for solid products such as cheese
- to define conditions which are as close as possible to the normal consumption conditions.

This last rule is particularly important when the aim is to relate descriptive data and hedonic data. However, in some cases, panellists are able to discriminate better when the samples are prepared differently from normal (for example, strong alcohols are more discriminated if diluted in water).

These rules, which seem obvious, are not always simple to apply. For example, when working on coffee, samples are generally presented as black (i.e. without sugar or milk/cream). However, for a hedonic test, it seems relevant to let consumers taste the samples as they are used to. In such a case, trying to relate descriptive data and hedonic data could be problematic.

It is generally recommended to mask colour differences for flavour evaluation. This practice avoids panellists using colour cues to identify the sample or making inferences about its nature (origin, composition, process, ripeness, etc.) and consequently evaluating the product not according to their perception but according to their knowledge and expectations. In fact, it is important to define conditions avoiding expectation, stimulus and logical errors (see sections 2.5.2.3–2.5.2.4). However, it must be pointed out that flavour perception results from integration of the stimulation of smell, taste and trigeminal systems (Delwiche 2004). Thus, the rule concerning masking colour presents some merit and could be relevant if the aim is to relate the descriptive data to instrumental data (Delwiche 2004), but this rule could be questioned if the aim is to relate descriptive data to hedonic data.

2.4.7.2 Design

When designing an experiment, several parameters must be considered.

2.4.7.2.1 Simultaneous or Sequential Presentation

The first parameter to fix is the presentation of the samples, that is, whether the panel will evaluate the samples comparatively (samples presented simultaneously) or not (samples presented successively one by one, i.e. monadic sequential presentation). The advantage of the comparative approach, when used with non-universal scales (i.e. scales other than Spectrum, LMS or gLMS), is that it is more sensitive in revealing small differences, but its use depends on the set of samples tested within a given session, making it difficult to aggregate results collected at different sessions. Thus, a comparative evaluation should be limited to the first steps of training, or for flash profiling and sorting tasks since the comparative approach is inherent to these methods. It must be noted that the comparative approach creates a difficult task when there are many descriptors to evaluate and when the number of samples is higher than six.

2.4.7.2.2 Number of Samples Per Session

The maximum number of samples clearly depends on the sensory modalities to be included and the number of descriptors to be evaluated. If only a visual evaluation is performed, 20–30 samples can be assessed (Meilgaard et al. 2007). The ISO standard (AFNOR 2003), which concerns the food domain, indicates as a general rule to have a maximum of six samples per session for full descriptive analysis (i.e. when all sensory modalities are considered), or a maximum of 10 samples for partial profiling with less than 10 attributes. It is also advisable to have a smaller number when the type of product to be tasted has a strong or persistent flavour. If the product has a strong or persistent flavour but the panel leader wants to limit the number of sessions (this could be important with an external panel), the session could be organized into two blocks with a quite long break (e.g. 15 minutes) between the two blocks.

The maximum number of samples per session can be determined during the last steps of the training stage. The maximum number could be determined using a specific test where, for example, one or two samples are repeated at the end of the session; it is then possible to examine if the replicates are evaluated differently (different means and/or different standard deviations). As noted by Larmond (1977)

> 'Panelists often lose their desire to discriminate before they lose their capability'.

Thus, motivation is an important factor to take into account for deciding on the number of samples to be tested within a session. If the number of samples in a given experiment is too large, an incomplete block design must be used (see later in this chapter).

2.4.7.2.3 Order of Presentation

The third point of the design concerns the order of presentation. Sample evaluation may be affected by the position and the previous sample (see section 2.5.2.5). The carry-over effect corresponds to the fact that one sample may impact the evaluation of the next samples. This effect is named carry-over in reference to experiments testing the impact of a given diet or treatment on a given outcome because it corresponds to the persistence of the effect of a diet/treatment on the next periods. One speaks about a first-order carry-over effect when considering the effect on the next immediate sample.

As the size of the panel is limited, it is better to use a balanced design rather than orders chosen randomly. The use of a Williams Latin square (MacFie et al. 1989) control of the position effect and first-order carry-over effect. For a given square, each sample is presented once at each position and each sample is preceded once by each other sample. If the number of samples p is even, the design is balanced if the number of panellists is a multiple of p. If the number of samples p is uneven, the design is balanced if the number of panellists is a multiple of $2 \times p$. When the number of panellists is not a multiple of p (or $2 \times p$), the rule is to use the full squares as much as possible, and then to draw supplementary lines in another square for the other panellists. For example, if the number of samples is four and the number of panellists is 10, two full squares will be used, then two lines of a third square will be randomly selected; if the number of samples is five and the number of panellists is 12, two squares will be used to balance for the position effect and for the first-order carry-over effect, then two lines of a third square will be randomly selected.

2.4.7.2.4 Incomplete Block Designs

When the total number of products is too large to be tested within one session, only a subset of products is tasted by each assessor in one session. Specific designs must be used, ideally balanced incomplete block (BIB) designs in which all products are evaluated by the same number of panellists and all possible pairs of products occur together equally often within the subsets of products presented to the panellists. A BIB design is specified by the following parameters (Gacula et al. 2009).

- t = total number of products
- k = number of products tested by each panellist, chosen by the experimenter as equal to or less than the number of samples that a panellist can test within a session without fatigue
- r = number of presentations of each product over the panel
- b = number of panellists
- λ = number of panellists testing each pair of sample.

A BIB design can be built if:

$$r\,t = b\,k = N, \text{ the total number of observations in the experiment}$$
$$\text{and } \lambda\,(t-1) = r(k-1).$$

Thus, the constraints are important, in particular for the usual size of a trained panel. Moreover, as noted by Wakeling and Buck (2001), such a design specifies only the allocation of a subset of products to each panellist but not the order of presentation of products for each subset. However, these authors indicate in their list of possible designs those for which it is possible to balance for position effect.

When possible, each panellist evaluates a subset of products at a given session but all products over several sessions. It is possible to construct experimental designs for such situations by using different arrangements of the same BIB design for the different sessions (see Wakeling and Buck 2001 for an example). With this approach all products must be prepared for each session.

However, as pointed out by Deppe et al. (2001), a kitchen constraint may arise due to available kitchen facilities or available staff for preparing the products, controlling cooking times and temperatures. If such a kitchen constraint occurs, only a given subset of products is prepared for a given session. One option is that the experimenter decides to split the whole range of products into different subsets. This option requires preliminary knowledge of the sensory differences between the samples in order to choose subsets so that, as far as possible, the size of the differences between the products is the same whatever the subset. In other words, one must avoid having similar samples in one subset presented at one session and very different samples in another subset presented at another session. Another option is to construct a design taking this constraint into account while also trying to find the optimal combination in terms of occurrences of a given pair over all subsets presented to the panellists, of position of each product over the whole design, and of first-order carry-over effect (Deppe et al. 2001).

2.4.7.2.5 Replicates

The number of replicates depends on the objective of the measurements. If the measurement series is conducted after a period of training in order to check panel performances and to finalize the list of descriptors, it is necessary to present each sample at least twice. This rule also applies if the measurement series is conducted to compare the performance of different panels (interlaboratory test). In these cases, for each sample, replications are generally performed on a given batch (see later in this chapter). For measurements on a set of products, the number of replicates is generally equal to two or three.

It is also important to consider the nature of the replicates, that is, if they are intra- or interbatches. If all panellists are presented, for each product, with a sample from a given batch (intrabatch replicates), the estimated variability corresponds to the measurement error. This option can be used during training and for panel monitoring. However, during measurement, the key objective is to determine if several products are perceived differently. Consequently, it is important to take into account the product variability (i.e. the experimental error) and thus to present independent replicates of the products (i.e. interbatch replicates). Surprisingly, the nature of the replicates seems to be rarely specified in papers reporting sensory experiments.

Running sensory profiling studies with replicates from a single batch is justified in psychophysical experiments where no variation in ingredients or preparation is expected. As pointed out by Meilgaard et al. (2007), it is more convenient to serve all panellists from a single batch (i.e. single pack/bottle); it is also less expensive. Nevertheless, this practice enables us only to conclude that the panel detect differences between the particular samples they evaluated but it is not possible to conclude about product differences. Thus, replicates from the same batches (case 1 of Figure 2.1) should be limited to specific cases where both the

Case	Replicate	Design	Type of design
a	Intra-batch	Product A, Batch X ↙ ↓ ↘ Replicate 1 2 3 Product B, Batch Y ↙ ↓ ↘ Replicate 1 2 3	Repeated measures
b	Inter-batch	Product A ↙ ↓ ↘ Batch T U V; Replicate 1 2 3 Product B ↙ ↓ ↘ Batch X Y Z; Replicate 1 2 3	Randomized split-plot
c	Inter-batch	Raw materials/ingredients Batch X ↙ ↓ ↘ Product A B C ↓ ↓ ↓ Replicate 1 1 1 Batch Y ↙ ↓ ↘ Product A B C ↓ ↓ ↓ Replicate 2 2 2 Batch Z ↙ ↓ ↘ Product A B C ↓ ↓ ↓ Replicate 3 3 3	Randomized block split-plot

Figure 2.1 Examples of experimental designs with intra- or interbatch replicates.

raw materials/ingredients and the preparation are known to be very constant. In other cases, replicates should correspond to different batches. Two types of designs are presented in Figure 2.1. Case 2 corresponds, for example, to the situation where products A and B are produced by two companies which use different raw materials/ingredients as well as different recipes/processes. Case 3 corresponds to a situation where a producer wants to compare different recipes/processes but taking into account possible variation due to raw materials. It is important to note that different models of analysis of variance must be used depending on the design.

2.5 Factors Affecting Results in Descriptive Analysis

Between the stimulation of a sensory receptor and the overt response of an individual, there is a sequence of processes which can be affected by several factors. These can be classified into two categories: physiological factors and psychological factors; some effects inducing errors in sensory measurements could be due to both physiological and psychological factors. Here, we will present the main factors that could affect the judgement of intensity of one sensory characteristic and how a panel leader can minimize the risk of these factors occurring and thus how a human measurement instrument can be optimized.

2.5.1 Physiological Factors

2.5.1.1 Individual Differences in Sensitivities

It is clear that individuals differ in their sensory equipment and that, consequently, each individual has their own pattern of sensitivities to different stimuli. Selection of panellists is the way to avoid recruiting individuals with a low sensitivity for a number of stimuli. Some diseases as well as many medicines can affect sensitivities, in particular chemosensory sensitivities. Thus, it is important during the selection process to ask volunteers if they have chronic disease(s) and/or a regular medical treatment. It is also important for panellists to be aware of the factors that could decrease their sensory ability, even temporarily. The panel leader must record these events and, if possible, discard panellists who have illness and/or medications that could impact their sensory ability, or at least carefully check their performance.

Besides a low sensitivity for a given sensory modality, there are numerous cases of specific low sensitivity to some volatile compounds (specific hyposmia). If some of these compounds are key components of the flavour of products that panellists will have to taste, it could be important to specifically test the candidates' ability to perceive these compounds. However, it must be noted that after repeated exposure to a given compound, it seems that sensitivity to this compound could increase. Thus, ability to perceive androstenone was induced in 10 out of 20 individuals initially insensitive to this compound (Wysocki et al. 1989). Another study found a similar effect with different odorants but only for women

(Diamond et al. 2005). Repeated short-term exposures to odorous wine key compounds reduced the olfactory detection thresholds in wine experts (Tempere et al. 2012). These latter two studies revealed that the effect of training is specific to the odorant used during the training and does not generalize to other odorants.

Regarding taste function, large interindividual differences, related to genetic differences, have been observed for bitterness of two synthetic compounds: phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP) (Drewnowski 2003). Though earlier research considered only two groups of individuals, the tasters and the non-tasters, more recent studies indicate that perception of PTC (and PROP) is on a continuum (Mennella et al. 2011).

It has been suggested that sensitivity to PTC/PROP would be related to sensitivity to other bitter compounds, in particular those with a similar structure, that is, containing a N-C=S group. However, there are divergent results which could be explained by the fact that humans are equipped with different bitter taste receptors. One study found that mutations in TAS2R38, responsible for differences in sensitivity to PTC/PROP, play a role in perception of goitrin, a bitter compound found in crucifers and containing a N-C=S group but that other factors must also be involved (Wooding et al. 2010).

If bitterness is an important characteristic of the studied products, it would appear relevant to eliminate non-tasters in order to have a panel as sensitive to bitterness as possible. However, considering our present knowledge on bitter taste genes, it seems not relevant to screen panellists on their sensitivity to PROP and, as pointed out by Hayes et al. (2013b), selection must be done by using the food of interest or the key ingredients present in this food (e.g. hop for beer).

2.5.1.2 Ageing

Ageing is accompanied by a decline in chemosensory abilities, such as the ability to perceive an odour or a taste (Doty et al. 1984; Mojet et al. 2001; Murphy 1986; Schiffman 1993). This decline can be due to physiological mechanisms that are inherent in the ageing process *per se*, but also to other factors related to lifetime events such as diseases or chronic exposure to toxic compounds. However, major declines generally do not occur before the age of 60 years and again, the selection must enable identification of individuals with low chemosensory sensibilities.

2.5.1.3 Gender and Hormonal Influence

It is known that women generally outperform men in odour tests, in particular in odour identification. However, odour performance is affected by menstrual cycle phase and duration of oral contraception intake and thus appears to be modulated by hormonal changes (Derntl et al. 2013).

Menopause is thus a stage where modifications of chemosensory perceptions could occur. It is well known that many women experience a reduction in salivary flow during this period. This could induce alterations in taste perception but also in mouthfeel perception. Women often report oral discomfort (Wardrop et al. 1989); however, the real impact of this decrease in salivary flow on mouthfeel perceptions does not seem to have been investigated very much.

There are numerous anecdotal reports of a higher smell sensitivity during pregnancy. However, studies are rather limited and results do not always confirm the hypothesis of a higher sensitivity. Nevertheless, in a review of this question, Cameron (2014) concludes that there are changes in odour perception during pregnancy. It seems that, in comparison to non-pregnant women, pregnant women give higher intensity ratings to some odorants and demonstrate an improved ability to identify some odorants, but also do worse for others.

There are only a few studies concerning gustatory function of pregnant women, suggesting differences from non-pregnant women and during the course of pregnancy. However, the effect of pregnancy seems to differ according to tastes and results also differ between studies. Results of tests aimed at evaluating the ability to discriminate among different concentrations of salt and sucrose solutions revealed that pregnant women were significantly less able to correctly identify concentration differences for salt solutions than were non-pregnant women (Brown & Toma 1986). Duffy et al. (1998) reported higher intensity ratings to bitter solutions of quinine hydrochloride in the first trimester of pregnancy. In contrast, other authors did not find any differences in intensity ratings for different tastants, including quinine, but observed that pregnant women had a lower ability to identify bitterness of low concentrations of caffeine and tended to have a lower ability to identify saltiness of low concentrations of NaCl (Kölble et al. 2001).

In conclusion, pregnancy could explain changes in performance in descriptive analysis and the panel leader should be informed about panellists' pregnancy to be able to explain changes in performances.

2.5.1.4 Hunger Influence

There are a limited number of studies examining the impact of hunger on olfactory function. Most of them are not recent and results are inconsistent; several found no effect of the hunger state, some showed higher thresholds after a meal than before, and surprisingly others found the opposite (see Albrecht et al. 2009 for a review). These inconsistencies could be due to several factors, one being the nature of the odorant. Albrecht et al. (2009) conducted a study with young non-smokers, healthy and normal-weight women using two odorants: a food-related odour (isoamyl acetate) and a non-food-related odour (n-butanol). A significant change in detection threshold was found only for the food-related odour, the sensitivity being higher in the satiety state. A similar result for a food odour (herb-based odorant) was found in a study conducted by Stafford and Welbeck (2011) using a between-subject design. However, this effect was found only for individuals with a BMI higher than the median. In this paper, the authors reported a higher olfactory sensitivity in the non-satiated group versus the satiated one for the non-food-related odour (n-butanol).

There are a very limited number of studies examining the impact of hunger on gustatory function, in particular in healthy non-obese individuals, and their conclusions are not in agreement. Pangborn (1959) did not find any effect of hunger on taste thresholds. Kaplan and Powell (1969) did not report any effect on threshold for 6-n-propylthiouracil but observed an increase in sensitivity for quinine after eating lunch for a majority of subjects. More recently, with a sample of non-smoking, non-drinking, non-obese, healthy young male subjects, significantly lower recognition thresholds for sucrose and salt were found after 14–16 hours of food deprivation compared to 1 hour after a meal while the values of recognition thresholds for bitter substances did not differ significantly (Zverev 2004). These results were not confirmed in a more recent study in which taste recognition thresholds were determined for six compounds (sucrose, fructose, sodium chloride, quinine sulfate, PROP and liquorice) for young, non-smokers, normal-weight subjects (Pasquet et al. 2006).

It is noticeable that in most studies the olfactory and gustatory functions were evaluated on the basis of detection or recognition thresholds. Despite the fact that there is no report of the effect of hunger on performances of a descriptive panel, it is advisable to always conduct measurements at the same time of day and to ask panellists to refrain from eating 2 hours before the session. Some authors also recommend scheduling the evaluation of a product type for the time of day when that product is normally consumed (Meilgaard et al. 2007). The same authors also indicate that it is not recommended to schedule the evaluation of highly flavoured or alcoholic products in the early morning. In general, end of the morning and mid-afternoon are recommended periods for sensory testing.

2.5.1.5 Adaptation and Build-Up Effects

Continued exposure to a stimulus results in a decrease in sensitivity to that stimulus (adaptation) and to other stimuli which activate the same receptors (cross-adaptation). Adaptation is particularly important for olfactory and taste modalities. The higher the intensity for a particular characteristic, the higher the adaptation. Consequently, the intensity of the characteristic for the following sample(s) will be perceived as lower than it should do.

The build-up effect corresponds to the cumulative effect of one sensation over successive ingestions. This could particularly occur for astringency and several authors have reported that intensity of astringency increased over successive samples.

Both effects can be minimized by asking panellists to have a break between samples and/or to rinse their mouth between samples. As noted by Johnson and Vickers (2004), common palate cleansers include water, sparkling water, crackers, carrots and apples. In the case of red wine, it was found that amongst four palate cleansers (water, pectin, carboxymethylcellulose and cracker), cracker was the most effective at reducing perceived astringency while water was found to be the least effective (Ross et al. 2007). With tannin solutions, none of the palate cleansers which were tested (water, carboxymethylcellulose, crackers, milk, chewing wax or nothing) prevent astringency build-up from occurring but panellists discriminate better between samples on astringency when nothing or water was used as palate cleanser (Lee & Vickers 2010). For the evaluation of bitterness of cream cheese samples varying in caffeine content, no differences were found between water, carrots, crackers, plain cream cheese or no rinsing for their ability to control adaptation and build-up, or for their ability to increase panellists' discrimination among samples (Johnson & Vickers 2004). This study showed that rinsing with sparkling water depressed the perceived bitterness at all levels of caffeine and revealed that residual cleansers can produce changes in the flavours of subsequently tasted products.

Moreover, Vickers et al. (2008) caution against using panellists' opinions of the effectiveness of a palate cleanser because people may not be aware of perceivable levels of residual tastants. Thus, the choice of palate cleanser must be done by the panel leader after trials conducted during the training period. This is particularly important for products containing irritants such as capsaicin, piperine or menthol which stimulate the trigeminal nerve and induce burning, warming, tingling and cooling sensations that may persist for several minutes after the exposure (see, for example, Allison et al. 1999 and Suwonsichon et al. 2009).

The effect of adaptation and build-up effects on results may be minimized by using a design of presentation (see section 2.4.7.2.3) balanced for order or position effect and carry-over effects.

2.5.2 Psychological Factors

2.5.2.1 Suggestion Error

It is well known that a panellist's answers can be influenced by the other panellists. Therefore, panellists are separated in individual booths in order to avoid possible discussions and in such a way that each panellist cannot see the facial expressions of the other panel members.

2.5.2.2 Expectation Error

Information about the objective of the experiment or about the nature of the products can induce expectation about the sensory characteristics of the products based on panellists' previous knowledge or beliefs. These expectations can influence panellists' evaluation. To avoid such expectation bias, the experimenter must avoid disclosing information about the products and the aim of the experiment, unless it is necessary for ethical reasons (Kemp et al. 2009). Expectation bias is more likely to occur with an internal panel, so the experimenter must be particularly careful not to include panellists who have too much knowledge about the product. Coding samples with random three-digit numbers (as noted in section 2.4.7.2) avoids any expectation error due to the use of codes such as letters or one-digit numbers.

2.5.2.3 Stimulus Error

A stimulus error occurs if when evaluating one characteristic, panellists use other characteristics of the product. In fact, since it is extremely difficult to evaluate samples presented in a blind condition, panellists tend to look for cues that could help them in their evaluation in order to give a right answer and please the panel leader. Therefore, uniform presentation of the samples is required to avoid such bias.

2.5.2.4 Logical Error, Cross-Modal Interactions and Dumping Effect

A logical error occurs when panellists are unconsciously influenced in their rating of one sensory characteristic by the presence of one or several other characteristics. A classic example is the impact of colour on evaluation of flavour intensity. A logical error depends on previous knowledge and consequently on the culture and/or expertise of the panellists and thus could be classified as an expectation error. Consequently, a logical error could be more frequent with experts in the product than with trained panellists.

As noted by Spence et al. (2010), a key question is whether a logical error corresponds to a response bias and/or to an actual effect on perception. The response bias is problematic for the reliability of the data. Until recently, sensory textbooks have considered only response bias and thus, it was generally recommended to mask differences (in particular colour differences) that could impact flavour evaluation. However, in recent years it has been more frequently admitted that some sensory characteristics could actually contribute to the perception of other characteristics. Unfortunately, it is not easy to disentangle response bias from perception effect. In order to minimize bias, the panel leader needs to teach the panellists that one characteristic is not necessarily predictive of the presence/intensity of another one. This could be done by presenting adequate samples or 'modified' products during the training period. However, the panel leader must be conscious that flavour perception is a multimodal perception due to the central integration of signals from the gustatory, olfactory and trigeminal systems (Delwiche 2004).

The role of experience in odour/taste integration has been demonstrated: a given odour enhances the perceived intensity of a given taste if this odour and this taste have been experienced simultaneously (Small & Prescott 2005). This odour and this taste are thus considered as congruent. Since congruency results from experience, it is not surprising that a given odour-taste pair could be considered more or less congruent, depending on the culture. For example, the association between strawberry and sweetness is strong for American students but only slight for French students (Sauvageot et al. 2000). It is noticeable that only one coexposure of an unknown odour with a sweet taste can render this odour sweeter smelling (Prescott et al. 2004). Moreover, it has also been found that this effect does not diminish over time, even if the odour is repeatedly experienced without the taste (Stevenson et al. 2000).

It is important that the panel leader instructs panellists to focus their attention on each single stimulus attribute and ignore all others when giving their responses. However, as pointed out by Frank (2003), it is important to be aware that panellists do not necessarily respond to instructions in the way that the experimenter anticipates and that instructions are not necessarily effective. Training is certainly more effective to improve panellists' analytical capabilities and consequently the selectivity of their answers.

It has also been shown that strawberry odour enhances sweetness but this is observed when only sweetness is rated but not if sweetness and fruitiness are rated. Thus, the effect of one characteristic on another could be due to the fact that a relevant attribute is missing in the ballot. This is the so-called 'dumping effect'. If panellists perceived a characteristic for which there is no descriptor in the ballot, they can report their perception on one or several of the available descriptors (see, for example, Teerling 1994). It has been shown that it is not the number of attributes which is the key factor but the relevance of the attributes. Consequently, it is really important during the vocabulary development stage to make panellists test a range of samples that covers the range which will be encountered during the measurement phase to be sure that all relevant descriptors are included in the ballot.

If in general, a panel leader must choose conditions promoting a more analytic approach, there could be situations where it is also interesting to collect results under conditions promoting a synthetic approach. For example, if a company looks for a way to reduce salt content, it could be important to know how the addition of a given flavour can increase saltiness perception in order to maintain product liking. In such a case, working with a trained panel, or only with a trained panel, is not necessarily the best option. As mentioned previously, colour may also impact flavour perception but again, should we consider such cross-modal interaction as a response bias and/or as an actual effect on perception? For the practitioner, it is important to consider whether masking colour is or is not relevant depending on the purpose of the study (see section 2.4.7.1).

2.5.2.5 Errors Related to the Presentation Order of the Samples

2.5.2.5.1 Positional Bias

This bias has been mainly described in difference testing as well as in hedonic measurements. However, it is common practice in descriptive analysis to use a presentation design balanced for position, that is, to present a given sample an equal number of times at each position. Analysis of variance enables the testing of position effect. However, this is rarely done in practice since designs balanced for position are used but this could be important if the number of panellists and the number of products do not enable to fully respect the balance. The few cases reporting the analysis of the order effect in descriptive analysis indicate that the order effect depends on the attribute. Schlich (1993) found that the sample evaluated in the first position seemed to have lower scores; such a result could be explained by panellists' timidity at the beginning of the session. In a interlaboratory study on coffee (ESN 1996), it was found that the occurrence of the order effect depends on the descriptor, and the direction also depends on the descriptor; ratings for acid, astringent and floral flavour are lower when a coffee sample is evaluated in the first position while ratings for bitter and burnt flavour are higher when a coffee sample is evaluated in the first position.

2.5.2.5.2 Effects Related to the Sequence of Presentation

The contrast effect corresponds to the situation where two quite different samples are presented successively and when panellists exaggerate their actual differences when giving their ratings.

The opposite is the convergence effect, which may occur when two similar samples are presented within a set of very different samples; the difference in ratings is lower than it should be actually.

These two effects are minimized by using a balanced order of presentation (see section 2.4.7.2). Moreover, when the number of samples to be evaluated is large and requires several sessions, it is important to try to have the same range of variation between samples at each session.

2.5.2.6 Range-Frequency Effect

This is also called the 'context effect'. It corresponds to the fact that intensity ratings of different samples are affected by the frequency of each sample within one testing session. Thus, it was found that the same drinks were rated sweeter when lower concentrations were presented more frequently and less sweet when higher concentrations were presented more frequently (Riskey et al. 1979). These authors also found that drinks preceded by a drink with a higher sucrose concentration were judged significantly less sweet than the same beverages preceded by a lower concentration, but noted that this effect was small compared to the effect of stimulus frequency and concluded that the effect of stimulus frequency is not mainly due to adaptation. In fact, this error, as well as errors related to the presentation order of the samples, could be due to the fact that individuals' judgements are in nature relative, not absolute. This type of error can be minimized by training. Using the Spectrum scale, LMS scales or magnitude estimation should minimize this effect. However, a context effect was observed with magnitude estimation and a LMS scale (Diamond & Lawless 2001).

2.5.2.7 Central Error

This error corresponds to the fact that panellists tend to use the middle of the scale and avoid the extremes. This results in a low standard deviation of their ratings and less discrimination between samples. Training is the way to minimize this error. Again, using the Spectrum scale, LMS scales or magnitude estimation should minimize this effect.

2.5.2.8 Idiosyncratic Scale Use

Some panellists tend to give low ratings and others high ratings. This results in low (high) average scores. Moreover, if they tend to use one extreme of the scale, this results in a low standard deviation of their ratings.

If, for a given panellist, these ways to rate samples occur for some descriptors, this could be due to a difference in sensitivity. If this is more systematic, this corresponds to an idiosyncratic way of using the scale which should be corrected by training and feedback.

2.6 Data Analysis

2.6.1 Evaluation of Panel Performance

As noted above, performance must be checked for a conventional profiling panel but is generally not checked for rapid methods.

Several parameters have been proposed to examine individual or panel performance. However, as noted previously, the most important criteria to evaluate are:

- each assessor's discrimination
- the panel's discrimination
- the agreement between assessors and the contribution of each assessor to any disagreement.

It is important that the panel leader has a convenient tool which gives him/her an easy-to-read report of the performance. One such tool, named CAP (for Control of Assessor Performances), was developed by Schlich (1997); this provides two tables summarizing results from univariate and multivariate analyses of variance. For each descriptor, an analysis of variance (ANOVA) is performed per assessor and for the whole panel. At the multivariate level, a MANOVA is conducted. Other approaches are possible and have been included in software dedicated to sensory evaluation (see Chapter 4).

2.6.2 Analyses of Product Differences

Readers should refer to Chapter 5 for a complete description but the sections below give some key points about data analyses.

2.6.2.1 Univariate Analyses

As noted before, such analyses are possible only when all panellists used the same descriptors. So, they cannot be applied for analysing data collected with free choice or flash profiling.

For each descriptor, an analysis of variance on the panel data is performed according to the following models:

1. attribute = sample + panellist + sample × panellist + error, for a design with intra-batch replicates (see section 2.4.7.2), and
2. attribute = sample + panellist + replicate + sample × panellist + sample × replicate + panellist × replicate + error, for a design with interbatch replicates (see section 2.4.7.2).

Many discussions have taken place on whether the assessor effect should be considered as a fixed or a random effect. Actually, trained panellists are not chosen randomly and thus, some consider the assessor to be a fixed effect. However, as the experimenter wants to provide a conclusion which is valid not only for its panel but could have been the same with another panel of similarly selected and trained individuals, it appears logical to consider the assessor effect as random. If from one replicate to another there is some random variation, it is also logical to consider the replicate effect as a random effect.

When the sample effect is significant, a multiple comparison of means, with the same error term as for the global test (for example, $MS_{sample \times panellist}$ for model 1), is performed to determine which samples are different. Another option is to calculate a contrast test to compare each sample to the global mean and thus identify samples which have a higher or lower intensity than the grand mean. This option is very useful when the number of samples is high (above five or six) to obtain easy-to-read results. When one sample is a reference, for example the product usually produced, a Dunnett test can be performed to compare each sample to this reference.

It is useful for the experimenter to have a convenient tool for reporting results in an easy-to-read form, particularly when there are a large number of samples and attributes. The Flash table, proposed by Schlich (1998), summarizes results over all the attributes in a easy-to-read presentation with attributes in lines and samples in columns. Attributes are grouped into clusters. The cluster at the top of the table is the one which accounts for the most variance; then, clusters are sorted according to the correlation between a cluster and its preceding cluster. Within each cluster, the atttributes are sorted from the best to the worst representative of the cluster. The clustering technique ensures that the amount of variance recovered by replacing the attributes by the means over attributes of each cluster would be at least equal to 70%. The samples (columns) are sorted by increasing mean of the first attribute. Sample means significantly lower or significantly greater than the grand mean are underlined and respectively preceded with a minus or a positive sign as in Table 2.2.

Other possible ways to visualize sample differences are presented in Chapter 5.

2.6.2.2 Multivariate Analyses for Product Positioning in a Sensory Map

Since a high number of dimensions is necessary to fully represent the information contained in the data set, multivariate methods are useful to represent visually the samples in a low-dimensional space. The most common method used for multivariate analysis of profiling data is the principal component analysis (PCA) which is performed on a table of mean scores crossing products and descriptors. As the same scale is used for all descriptors, the logical option is to conduct PCA on the covariance matrix of this table (Borgognone et al. 2001). However, since the intensity scale is rarely used entirely for odor and aroma descriptors, these descriptors have less weight in the axes; thus, some experimenters perform the PCA on the correlation matrix. However, with this option,

Table 2.2 Example of a Flash table summarizing results obtained with 14 panellists who evaluated eight products and rated 14 attributes on a continuous linear scale (panellists' responses were converted into intensity scores from 0 to 10). Values in bold and preceded by + are significantly higher than the grand mean (gmean), values in italics and preceded by (–) are significantly lower than the grand mean.

Descriptor	F[a]	p[b]	gmean	P1	P2	P3	P4	P5	P6	P7	P8
Pepper	**18.92**	0	1.35	*(–)0.35*	*(–)0.4*	*(–)0.88*	*(–)0.94*	1.16	1.25	**+1.99**	**+3.83**
Pungent	**9.52**	0	1.20	*(–)0.58*	*(–)0.27*	*(–)0.71*	0.96	1.48	1.28	**+1.84**	**+2.45**
Salty	**18.8**	0	3.23	*(–)2.24*	*(–)1.05*	3.01	*(–)2.33*	**+4.66**	3.76	**+4.53**	**+4.28**
Viscous	**29.48**	0	4.95	*(–)4.31*	*(–)3.63*	*(–)4.26*	*(–)4.42*	**+6.80**	4.59	**+7.41**	*(–)4.19*
Sticky	**7.25**	0	4.55	4.67	*(–)3.08*	4.55	*(–)4.04*	4.72	4.47	**+6.27**	4.65
Acid	**10.92**	0	1.29	1.35	*(–)0.62*	*(–)0.77*	*(–)0.60*	**+1.74**	1.00	**+2.77**	1.45
Onion	**4.65**	0	1.26	*(–)0.32*	1.00	1.08	1.47	1.36	1.36	**+1.77**	**+1.68**
Umami	**2.7**	0.014	1.19	1.05	*(–)0.58*	1.14	0.92	1.51	1.25	**+1.88**	1.18
Beef	**54.11**	0	4.71	*(–)0.13*	*(–)3.51*	4.58	4.84	**+7.71**	5.10	**+6.86**	4.93
Granular	**57.76**	0	1.98	*(–)0.44*	2.31	*(–)0.94*	**+2.88**	**+6.61**	*(–)1.01*	*(–)0.70*	*(–)0.91*
Sweet	**13.77**	0	1.53	**+3.37**	1.53	1.58	1.54	*(–)0.58*	1.30	*(–)0.65*	1.69
Carrot	**74.69**	0	2.80	**+8.32**	2.50	2.64	2.58	*(–)0.60*	*(–)2.37*	*(–)1.56*	(–)1.88
Unctuous	**15.41**	0	4.72	**+5.58**	5.21	**+5.87**	4.93	*(–)1.57*	**+5.48**	*(–)3.52*	**+5.60**
Fatty	**2.18**	0.043	3.79	3.95	3.43	4.40	4.11	*(–)2.89*	4.39	*(–)3.01*	4.12

[a]F value of the product effect
[b]probability

all descriptors have the same variance and this could lead to giving too much importance to some odour or aroma descriptors.

Another option is to conduct a PCA per sensory modality. When there are different steps in the sensory evaluation, such as a visual examination, an odour examination and an 'in-mouth' evaluation, it could also be relevant to perform a PCA for attributes rated for each of these steps. Since the aim of performing a PCA is to obtain a map illustrating similarities and dissimilarities between products, the best option is to work on the table of panel means for each product and each descriptor. To avoid non-discriminant attributes influencing first axes, it is recommended to delete these attributes from the table to be analysed (generally those with a p-value higher than 0.10). The first axis obtained by a PCA accounts for the maximum possible variance between the products. The second one corresponds to the maximum possible remaining variance between the products and is not correlated with the previous axis, and so on. These axes are linear combinations of the original variables (attributes).

A key question for the experimenter is to select the appropriate number of axes to represent the products. This is a compromise between two criteria:

- a simple representation, that is, the lowest possible number of axes
- a sufficient percentage of initial variance.

The problem with a PCA performed on the mean scores over assessors is that there is no way to know if two products are significantly different. Consequently, it has been proposed to build confidence ellipses around each product using a resampling technique (Husson et al. 2005).

Canonical variate analysis (CVA) is another multivariate approach overcoming the PCA limit. This analysis is equivalent to a multivariate analysis of variance. CVA enables differences between preidentified groups (e.g. the different products) to be explained by the variables and thus, is more powerful for explaining product differences. Confidence ellipses can be drawn around each product on the basis of the individual assessments which can be projected as supplementary points on the map.

2.7 Reporting Results

For any study, the report should mention all information enabling the study to be identified, including its title, who commissioned the study and who conducted it. The publication date of the report should also be mentioned.

After an executive summary, the report starts with an introduction describing the context and objectives of the study. Then, experimental conditions must be described, including:

- a description of the panel (selection, previous experience, specific training duration, size)
- a description of the products (characteristics, sampling, date of manufacture, shelf-life, conditions of storage and preparations, serving temperature, vessels used to serve the samples, quantity served)

- a description of the testing sessions (place, dates, time of day), the testing conditions (room, temperature, vessels, lighting), the experimental design (number of sessions, number of samples per session, mode and order of presentation of the samples, number and nature of replicates)
- statistical methods and software used
- presentation and interpretration of the results
- conclusions.

The report must also include appendices with answer forms, a complete list of descriptors with their definition and/or standards and all information not reported in the main part of the text (e.g. details about panel training and about control of performance). Finally, the report must indicate where the raw data are stored and how to access them.

2.8 Summary

This chapter describes the general procedure for obtaining a sensory profile, focusing in particular on the conventional approach where all panellists rate the intensity of the same set of descriptors. This chapter presents some key points about the measurement of intensity, such as the advantages and limits of different types of scales and the factors (bias) that could affect the evaluation of intensity. It indicates how to minimize the risk of these biases but also highlights that the procedures applied to avoid bias or reduce variability in the answers could lead to experimental conditions that might not be very relevant for some specific objectives.

2.9 Future Developments

As mentioned by Lawless and Heymann (2010), conventional sensory profiling methods are one of the most powerful sensory tools. The online survey conducted with participants of the Pangborn Sensory Science Symposium in 2011 revealed that they are considered as the most reliable methods to track product changes (Delarue 2015). Not surprisingly, respondents to this survey underlined that their main weaknesses are that they are time-consuming and expensive. This is clearly related to the time and resources devoted to train the panel. Thus, these conventional approaches are not suitable in all situations, and in particular when product developers have asked for a rapid answer concerning products for which a trained panel has not yet been set up. The different rapid methods described in this book may be useful in these situations. However, the main weaknesses of these rapid methods, highlighted in the above-mentioned survey, were the complexity of data processing and the lack of dedicated software. It is likely that these limitations will be addressed soon thanks to training of practitioners and students and the development of software for data acquisition and processing. Another problem with these rapid methods is that as replications are usually not performed, there is no standardized tool to evaluate the reliability of the results (Ares 2015). The development of such tools is another important point for expanded use of these rapid methods.

One issue which is more and more raised, probably partly in relation to the development of rapid methods, is to run experiments aimed at obtaining a sensory map with subjects other than trained panellists. Besides the potential advantage of saving time and money, another potential interest would be to obtain a sensory map of variants of a product as perceived by the target population of this product. So besides and above the question of training, the question of the type of panellists would certainly be an important issue to consider in the future development of sensory evaluation.

As pointed out by Maitre et al. (2015), one can wonder if it is relevant to collect a sensory map with a trained panel composed of healthy adults (generally from 20 to 50 years of age) to try to explain preferences of a different target, such as children, elderly people or patients with a given disease. It would be actually problematic if the target consumers are more sensitive to one or several sensory characteristics than the trained panel (Issanchou 2015). Even if such a case does not occur frequently, it could eventually happen. It has, for example, been found that a panel of elderly people was more sensitive to mouth drying than a panel of younger adults (Withers et al. 2013). It is still an open question whether a panel of children would be relevant to provide sensory descriptions. However, considering the age-related differences in bitterness perception (Mennella et al. 2014), one can suppose that data obatained with an adult panel could be irrelevant to explain children's preferences for some specific products. Nevertheless, it is quite challenging to apply descriptive methods with children below the age of 8 (Nicklaus 2015). However, rapid methods, and in particular sorting methods, could have potential that would need to be explored (Nicklaus 2015), even with children as young as 5–6 years of age.

The method of preparing and presenting the samples must be standardized and chosen to avoid any possible bias. However, this rule could lead to asking panellists to evaluate products in conditions very different from the way in which they are generally consumed (e.g. when colour is masked); even if the defined experimental conditions represent one way to prepare and consume the product, it is only one option amongst different possible options (e.g. black coffee without sugar only). This could be problematic for some objectives, in particular to relate sensory characteristics to consumers' liking, or to obtain a sensory description of products as they are consumed at home. From a practical point of view, it could be extremely difficult (or even not possible) in a sensory laboratory to prepare a given range of products in different ways taking into account the variability of transformations made at home (recipes, cooking methods and seasoning). Thus, in order to get information on the sensory characteristics of products as they are usually consumed, it could be relevant to train a panel and then to ask this panel to evaluate the products at home as they usually prepared and consumed, as done recently by Martin et al. (2014). This approach could also take into account the fact that a given product is generally not consumed alone. Actually, it could be very useful to know how the sensory profile of a product could be modified when tasting after another product. However, there are currently only a limited number of studies reporting how one product modifies the sensory perception of another and most of them concern wine and cheese pairings; they have looked at the impact of consuming cheese before tasting wines (Madrigal-Galan & Heymann 2006; Nygren et al. 2002) or conversely at the impact of consuming wine before tasting cheese (Nygren et al. 2003a). One can expect that in the future, there will be growing interest in studying such interactions that could occur with sequential or mixed tastings of food combinations (Nygren et al. 2003b).

Thanks to an increase in knowledge about conventional descriptive methods and the particular characteristics of the human 'instrument', use of conventional descriptive methods has increased considerably in recent years. New approaches have been developed to give more rapid answers than with the conventional approaches and to answer different questions, such as the evolution of the sensory perceptions during tasting (Pineau & Schlich 2015). So, the challenge for sensory experimenters is to develop expertise in choosing the best option, depending on the problem, and to communicate with developers and marketing. This does not mean forgetting the 'conventional' approaches and current good sensory practices but building other possible tools to better meet current needs.

References

AFNOR (1995a) *NF ISO 11035. Sensory analysis – Identification and Selection of Descriptors for Extablishing a Sensory Profile by a Multidimensional Approach*. Paris: AFNOR.

AFNOR (1995b) *NF ISO 11036. Sensory Analysis – Methodology – Texture Profile*. Paris: AFNOR.

AFNOR (2003) *NF ISO 13299. Sensory Analysis – Methodology – General Guidance for Establishing a Sensory Profile*. Paris: AFNOR.

AFNOR (2009) *NF ISO 5492. Sensory Analysis. Vocabulary*. Paris:AFNOR.

Albrecht, J., Schreder, T., Kleemann, A.M. et al. (2009) Olfactory detection thresholds and pleasantness of a food-related and a non-food odour in hunger and satiety. *Rhinology*, **47**, 160–165.

Allison, A.M.A., Chambers, I.V.E., Milliken, G.A. & Chambers, D.H. (1999) Effects of interstimulus rinsing and time on measurements of capsaicin heat in tomato salsa. *Journal of Sensory Studies*, **14**, 401–414.

Ares, G. (2015) Methodological challenges in sensory characterization. *Current Opinion in Food Science*, **3**, 1–5.

Bartoshuk, L.M., Duffy, V.B., Fast, K., Green, B.G., Prutkin, J. & Snyder, D.J. (2003) Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: what we have learned from genetic variation in taste. *Food Quality & Preference*, **14**, 125–138.

Borgognone, M.G., Bussi, J. & Hough, G. (2001) Principal component analysis in sensory analysis: covariance or correlation matrix? *Food Quality & Preference*, **12**, 323–326.

Brown, J.E. & Toma, R.B. (1986) Taste changes during pregnancy. *American Journal of Clinical Nutrition*, **43**, 414–418.

Cameron, E.L. (2014) Pregnancy and olfaction: a review. *Frontiers in Psychology*, **5**, 67.

Civille, G.V. & Lawless, H.T. (1986) The importance of language in describing perceptions. *Journal of Sensory Studies*, **1**, 203–215.

Delarue. J.(2015) The use of rapid sensory methods in R&D and research: an introduction. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques*. Cambridge: Woodhead Publishing, pp. 3–25.

Delarue, J. & Sieffermann, J.M. (2004) Sensory mapping using Flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality & Preference*, **15**, 383–392.

Delwiche, J. (2004) The impact of perceptual interactions on perceived flavor. *Food Quality & Preference*, **15**, 137–146.

Depledt, F. (2009) *Evaluation Sensorielle. Manuel Méthodologique*. Paris: Lavoisier.

Deppe, C., Carpenter, R. & Jones, B. (2001) Nested incomplete block designs in sensory testing: construction strategies. *Food Quality & Preference*, **12**, 281–290.

Derntl, B., Schöpf, V., Kollndorfer, K. & Lanzenberger, R. (2013) Menstrual cycle phase and duration of oral contraception intake affect olfactory perception. *Chemical Senses*, **38**, 67–75.

Diamond, J. & Lawless, H.T. (2001) Context effects and reference standards with magnitude estimation and the labeled magnitude scale. *Journal of Sensory Studies*, **16**, 1–10.

Diamond, J., Dalton, P., Doolittle, N. & Breslin, P.A.S. (2005) Gender-specific olfactory sensitization: hormonal and cognitive influences. *Chemical Senses*, **30**, I224–i225.

Doty, R.L., Shaman, P., Appelbaum, S.L., Giberson, R., Siksorski, L. & Rosenberg, L. (1984) Smell identification: changes with age. *Science*, **226**, 1441–1443.

Drewnowski, A. (2003) Genetics of human taste perception. In: Doty, R.L. (ed.) *Handbook of Olfaction and Gustation*, 2nd edn. New York: Marcel Dekker, pp. 847–860.

Duffy, V.B., Bartoshuk, L.M., Strigel-Moore, R. & Rodin, J. (1998) Taste changes across pregnancy. *Annals of the New York Academy of Sciences*, **855**, 805–809.

ESN (1996) *A European Sensory and Consumer Study. A Case Study on Coffee*. European Sensory Network. Available from Campden & Chorleywood Food Research Association, Chipping Campden, Gloucestershire, GL55 6LD UK.

Findlay, C.J., Castura, J.C., Schlich, P. & Lesschaeve, I. (2006) Use of feedback calibration to reduce the training time for wine panels. *Food Quality & Preference*, **17**, 266–276.

Frank, R.A. (2003) Response context affects judgments of flavor components in foods and beverages. *Food Quality & Preference*, **14**, 139–145.

Gacula, M.J., Singh, J., Bi, J. & Altan, S. (2009) *Statistical Methods in Food and Consumer Research*. Burlington: Academic Press.

Gatchalian, M.M., de Leon, S.Y. & Yano, T. (1991) Control chart technique: a feasible approach to measurement of panelist performance in product profile development. *Journal of Sensory Studies*, **6**, 239–254.

Giboreau, A., Dacremont, C., Egoroff, C. et al. (2007) Defining sensory descriptors: towards writing guidelines based on terminology. *Food Quality & Preference*, **18**, 265–274.

Green, B.G., Shaffer, G.S. & Gilmore, M.M. (1993) Derivation and evaluation of a semantic scale of oral sensation magnitude with apparent ratio properties. *Chemical Senses*, **18**, 683–702.

Green, B.J.., Dalton, P., Cowart, B., Shaffer, G., Rankin, K. & Higgins, J. (1996) Evaluating the 'labeled magnitude scale' for measuring sensations of taste and smell. *Chemical Senses*, **21**, 323–334.

Guest, S., Essick, G., Patel, A., Prajapati, R. & McGlone, F. (2007) Labeled magnitude scales for oral sensations of wetness, dryness, pleasantness and unpleasantness. *Food Quality & Preference*, **18**, 342–352.

Hayes, J.E., Allen, A.L. & Bennett, S.M. (2013a) Direct comparison of the generalized visual analog scale (gVAS) and general labeled magnitude scale (gLMS). *Food Quality & Preference*, **28**, 36–44.

Hayes, J.E., Feeney, E.L. & Allen, A.L. (2013b) Do polymorphisms in chemosensory genes matter for human ingestive behavior? *Food Quality & Preference*, **30**, 202–216.

Hummel, T., Sekinger, B., Wolf, S.R., Pauli, E. & Kobal, G. (1997) 'Sniffin' sticks': olfactory performance assessed by the combined testing of odor identification, odor discrimination and olfactory threshold. *Chemical Senses*, **22**, 39–52.

Husson, F., Le, S. & Pages, J. (2005) Confidence ellipse for the sensory profiles obtained by principal component analysis. *Food Quality & Preference*, **16**, 245–250.

Ishii, R. & O'Mahony, M. (1991) Use of multiple standards to define sensory characteristics for descriptive analysis – aspects of concept-formation. *Journal of Food Science*, **56**, 838–842.

Issanchou, S. (2015) Sensory & consumer studies with special populations: children and elderly. *Current Opinion in Food Science*, **3**, 53–58.

Issanchou, S., Lesschaeve, I. & Köster, E.P. (1995) Screening individual ability to perform descriptive analysis of food products: basic statements and application to a camembert cheese descriptive panel. *Journal of Sensory Studies*, **10**, 349–368.

Johnson, E.A. & Vickers, Z. (2004) The effectiveness of palate cleansing strategies for evaluating the bitterness of caffeine in cream cheese. *Food Quality & Preference*, **15**, 311–316.

Kaplan, A.R. & Powell, W. (1969) Taste acuity in relation to hunger. *Nature*, **221**, 367–368.

Kemp, S.E., Hollowood, T. & Hort, J. (2009) *Sensory Evaluation: A Practical Handbook*. Chichester: Wiley-Blackwell.

Kölble, N., Hummel, T., von Mering, R., Huch, A. & Huch, R. (2001) Gustatory and olfactory function in the first trimester of pregnancy. *European Journal of Obstetrics, Gynecology & Reproductive Biology*, **99**, 179–183.

Landau, S.I. (1989) *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.

Larmond, E. (1977) *Laboratory Methods for Sensory Evaluation of Food*. Ottawa: Canada Department of Agriculture.

Lawless, H.T. (1999) Descriptive analysis of complex odors: reality, model or illusion? *Food Quality & Preference*, **10**, 325–332.

Lawless, H.T. & Heymann, H. (1998) *Sensory Evaluation of Food*. New York: Chapman & Hall.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evaluation of Food. Principles and Practices*. New York: Springer.

Lawless, H.T., Sheng, N. & Knoops, S.S.C.P. (1995) Multidimensional scaling of sorting data applied to cheese perception. *Food Quality & Preference*, **6**, 91–98.

Lee, C.A. & Vickers, Z.M. (2010) Discrimination among astringent samples is affected by choice of palate cleanser. *Food Quality & Preference*, **21**, 93–99.

Lesschaeve, I. & Issanchou, S. (1996) Could selection tests detect the future performance of descriptive panellists? *Food Quality & Preference*, **7**, 177–183.

Lyon, D.H., Francombe, M.A., Hasdell, T.A. & Lawson, K. (1992) *Guidelines for Sensory Analysis in Food Product Development and Quality Control*. London: Chapman & Hall.

MacFie, H.J., Bratchell, N., Greenhoff, K. & Vallis, L.V. (1989) Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, **4**, 129–148.

Madrigal-Galan, B. & Heymann, H. (2006) Sensory effects of consuming cheese prior to evaluating red wine flavor. *American Journal of Enology and Viticulture*, **57**, 12–22.

Maitre, I., Symoneaux, R. & Sulmont-Rossé, C. (2015) Conducting sensory analysis tests among the elderly. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques*. Cambridge: Woodhead Publishing, pp. 485–508.

Mammasse, N. (2012) Le nombre de sujets dans les panels d'analyse sensorielle: une approche base de données. PhD Thesis. Université de Bourgogne, Dijon (France), p. 89.

Martin, C., Visalli, M., Lange, C., Schlich, P. & Issanchou, S. (2014) Creation of a food taste database using an in-home 'taste' profile method. *Food Quality & Preference*, **36**, 70–80.

Meilgaard, M., Reid, D.S. & Wyborski, K.A. (1982) Reference standards for beer flavor terminology system. *Journal of the American Society of Brewing Chemists*, **40**, 119–128.

Meilgaard, M., Civille, G.V. & Carr, T. (2007) *Sensory Evaluation Techniques*. Boca Raton: CRC Press.

Mennella, J.A., Pepino, M.Y., Duke, F.F. & Reed, D.R. (2011) Psychophysical dissection of genotype effects on human bitter perception. *Chemical Senses*, **36**, 161–167.

Mennella, J.A., Reed, D.R., Roberts, K.M., Mathew, P.S. & Mansfield, C.J. (2014) Age-related differences in bitter taste and efficacy of bitter blockers. *PLoS One*, **9**, e103107.

Mojet, J., Christ-Hazelhof, E. & Heidema, J. (2001) Taste perception with age: generic or specific losses in threshold sensitivity to the five basic tastes? *Chemical Senses*, **26**, 845–860.

Muir, D.D. & Hunter, E.A. (1991–1992) Sensory evaluation of Cheddar cheese: order of tasting and carryover effects. *Food Quality & Preference*, **3**, 141–145.

Muñoz, A.M. & Civille, G.V. (1998) Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, **13**, 57–75.

Murphy, C. (1986) Taste and smell in the elderly. In: Meiselman, M.L. & Rivlin, R.S. (eds) *Clinical Measurement of Taste and Smell*. New York: Macmillan, pp. 343–369.

Murray, J.M., Delahunty, C.M. & Baxter, I.A. (2001) Descriptive sensory analysis: past, present and future. *Food Research International*, **34**, 461–471.

Nicklaus, S. (2015) Sensory testing in new product development: working with children. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques*. Cambridge: Woodhead Publishing, pp. 473–484.

Noble, A.C., Arnold, R.A., Buechsenstein, J., Leach, E.J., Schmidt, J.O. & Stern, P.M. (1987) Modification of a standardized system of wine aroma terminology. *American Journal of Enology and Viticulture*, **38**, 143–145.

Nygren, I.T., Gustafsson, I.B. & Johansson, L. (2002) Perceived flavour changes in white wine after tasting blue mould cheese. *Food Service Technology*, **2**, 163–171.

Nygren, I.T., Gustafsson, I.B. & Johansson, L. (2003a) Perceived flavour changes in blue mould cheese after tasting white wine. *Food Service Technology*, **3**, 143–150.

Nygren, I.T., Gustafsson, I.B. & Johansson, L. (2003b) Effects of tasting technique – sequential tasting vs. mixed tasting – on perception of dry white wine and blue mould cheese. *Food Service Technology*, **3**, 61–69.

Pangborn, R.M. (1959) Influence of hunger on sweetness preferences and taste thresholds. *American Journal of Clinical Nutrition*, **7**, 280–287.

Pasquet, P., Monneuse, M.O., Simmen, B., Marez, A. & Hladik, C.M. (2006) Relationship between taste thresholds and hunger under debate. *Appetite*, **46**, 63–66.

Pineau, N. & Schlich, P. (2015) Temporal dominance of sensations (TDS) as a sensory profiling technique. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques*. Cambridge: Woodhead Publishing, pp. 269–306.

Pineau, N., Chabanet, C. & Schlich, P. (2007) Modeling the evolution of the performance of a sensory panel: a mixed-model and control chart approach. *Journal of Sensory Studies*, **22**, 212–241.

Prescott, J., Johnstone, V. & Francis, J. (2004) Odor-taste interactions: effects of attentional strategies during exposure. *Chemical Senses*, **29**, 331–340.

Rainey, B.A. (1986) Importance of reference standards in training panelists. *Journal of Sensory Studies*, **1**, 149–154.

Riskey, D.R., Parducci, A. & Beauchamp, G.K. (1979) Effects of context in judgments of sweetness and pleasantness. *Perception & Psychophysics*, **26**, 171–176.

Ross, C.F., Hinken, C. & Weller, K. (2007) Efficacy of palate cleansers for reduction of astringency carryover during repeated ingestions of red wine. *Journal of Sensory Studies*, **22**, 293–312.

Rouby, C., Thomas-Danguin, T., Vigouroux, M. et al. (2011) The Lyon Clinical Olfactory Test: validation and measurement of hyposmia and anosmia in healthy and diseased populations. *International Journal of Otolaryngology*, **2011**, 203805.

Sauvageot, F. (1975) L'analyse sensorielle: outil scientifique? *Cahiers de Nutrition et de Diététique*, **10**, 26–44.

Sauvageot, F., Hoang Nguyen, D. & Valentin, D. (2000) Les mots évoquent-ils des saveurs? Une comparaison entre étudiants de France, du Vietnam et des USA. *Sciences des Aliments*, **20**, 491–522.

Schifferstein, H.N.J. (2012) Labeled magnitude scales: a critical review. *Food Quality & Preference*, **26**, 151–158.

Schiffman, S.S. (1993) Perception of taste and smell in elderly persons. *Critical Reviews in Food Science and Nutrition*, **33**, 17–26.

Schlich, P. (1993) Uses of change over designs and repeated measurements in sensory and consumer studies. *Food Quality & Preference*, **4**, 223–235.

Schlich, P. (1997) CAP: une méthode et un outil de contrôle rapide et synthétique des performances des sujets en évaluation sensorielle descriptive. 5èmes Journées Agro-industries et Méthodes Statistiques, December 3–5 december, Versailles, pp. 8.1–8.10.

Schlich, P. (1998) What are the sensory differences among coffees? Multi-panel analysis of variance and FLASH analysis. *Food Quality & Preference*, **9**, 103–106.

Small, D.M. & Prescott, J. (2005) Odor/taste integration and the perception of flavor. *Experimental Brain Research*, **166**, 345–357.

Small, D.M., Voss, J., Mak, Y.E., Simmons, K.B., Parrish, T.R. & Gitelman, D.R. (2004) Experience-dependent neural integration of taste and smell in the human brain. *Journal of Neurophysiology*, **92**, 1892–1903.

Spence, C., Levitan, C., Shankar, M. & Zampini, M. (2010) Does food color influence taste and flavor perception in humans? *Chemosensory Perception*, **3**, 68–84.

Stafford, L.D. & Welbeck, K. (2011) High hunger state increases olfactory sensitivity to neutral but not food odors. *Chemical Senses*, **36**, 189–198.

Stampanoni, C.R. (1993) The 'quantitative flavor profiling' technique. *Perfumer & Flavorist*, **18**, 19–24.

Stevenson, R.J., Boakes, R.A. & Wilson, J.P. (2000) Resistance to extinction of conditioned odor perceptions: evaluative conditioning is not unique. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26**, 423–440.

Sulmont, C., Lesschaeve, I., Sauvageot, F. & Issanchou, S. (1999) Comparative training procedures to learn odor descriptors: effects on profiling performance. *Journal of Sensory Studies*, **14**, 467–490.

Suwonsichon, S., Chambers, E., Chambers, D.H. & Milliken, G.A. (2009) Effects of oral rinsing on the perception of residual cooling and burn in highly mentholated toothpaste. *Journal of Sensory Studies*, **24**, 290–300.

Teerling, T. (1994) The missing link in product development. *Perfumer & Flavorist*, **19**, 17–22.

Tempere, S., Cuzange, E., Bougeant, J.C., de Revel, G. & Sicard, G. (2012) Explicit sensory training improves the olfactory sensitivity of wine experts. *Chemosensory Perception*, **5**, 205–213.

Vangtal, A. & Hammond, E.G. (1986) Correlation of the flavor characteristics of Swiss-type cheeses with chemical parameters. *Journal of Dairy Science*, **69**, 2892–2993.

Varela, P. & Ares, G. (2012) Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International*, **48**, 893–908.

Vickers, Z., Morris, E.E. & Savaria, M. (2008) Effectiveness of palate cleansers for evaluating sourness. *Journal of Sensory Studies*, **23**, 526–532.

Wakeling, I.N. & Buck, D. (2001) Balanced incomplete block designs useful for consumer experimentation. *Food Quality & Preference*, **12**, 265–268.

Wardrop, R.W., Hailes, J., Burger, H. & Reade, P.C. (1989) Oral discomfort at menopause. *Oral Surgery, Oral Medicine, Oral Pathology*, **67**, 535–540.

Williams, A.A. & Langron, S.P. (1984) The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food & Agriculture*, **35**, 558–568.

Withers, C., Gosney, M.A. & Methven, L. (2013) Perception of thickness, mouth coating and mouth drying of dairy beverages by younger and older volunteers. *Journal of Sensory Studies*, **28**, 230–237.

Wooding, S., Gunn, H., Ramos, P., Thalmann, S., Xing, C. & Meyerhof, W. (2010) Genetics and bitter taste responses to goitrin, a plant toxin found in vegetables. *Chemical Senses*, **35**, 685–692.

Wysocki, C.J., Dorries, K.M. & Beauchamp, G.K. (1989) Ability to perceive androstenone can be acquired by ostensibly anosmic people. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7976–7978.

Zverev, Y.P. (2004) Effects of caloric deprivation and satiety on sensitivity of the gustatory system. *BMC Neuroscience*, **5**, 5.

CHAPTER 3

Setting Up and Training a Descriptive Analysis Panel

Margaret A. Everitt

3.1 Introduction: Descriptive Analysis

A wide variety of descriptive analysis techniques are available, ranging from the long-established methods such as the flavor profile and the texture profile methods, Spectrum™ Method, quantitative descriptive analysis (QDA), those that enable a faster output, for example free choice profiling (FCP) and flash profiling, to those that are designed to measure dynamic changes in attribute intensity, such as time intensity (TI), progressive profiling (PP) and temporal dominance of sensations (TDS). In addition, hybrid forms of descriptive analysis consisting of combinations of elements from, in particular, QDA and the Spectrum Method are commonly used. Irrespective of method, the aim is to generate data that describe and quantify the differences and similarities among a set of products. Although the exact approach differs depending on method (see subsequent chapters), there are generic stages that are relevant, to a greater or lesser extent, to all (Figure 3.1).

In the broadest sense, a sensory panel can be defined as a group of recruited and selected individuals from whom product evaluation data are obtained. This definition can refer to either trained or untrained panels. Trained panels are required for most descriptive analysis methods, with exceptions being FCP and flash profiling which more commonly use consumers and a training phase is not included. This chapter focuses on trained panels and the generic stages involved in both screening and training people for the role.

3.2 Types of Panel Resource

Once a business has made the commitment to invest in a descriptive analysis panel, it can be set up either by utilizing internal staff, if there is sufficient resource available, or by recruiting external people specifically for the role. Various factors need to be considered when deciding which option to take.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

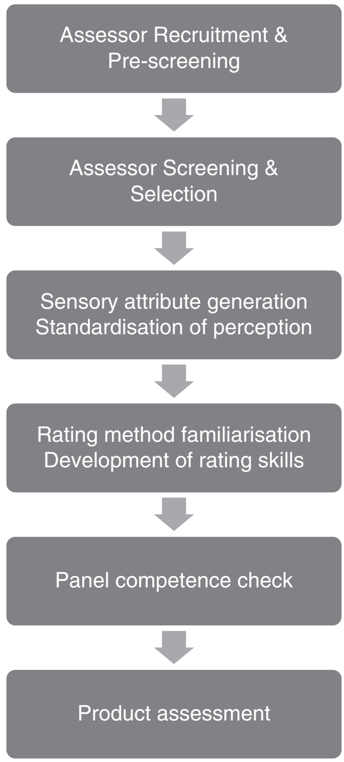


Figure 3.1 Flow chart showing the generic stages of panel selection and training for descriptive analysis.

- Frequency of testing, e.g. routinely every week, less frequently or on an irregular basis
- Range of products to be assessed
- Application of results: level of detail and accuracy required
- Implications of current and future business needs on resource
- Explicit and implicit cost

If testing will be and is likely to remain intermittent, the product range is limited and/or a high level of discrimination ability is not required then use of internal resource may be the most appropriate option. However, experience has shown that as the value of descriptive analysis data becomes recognized within a business, the volume and complexity of testing typically increase and the move to a dedicated panel often becomes a necessity.

3.2.1 Employees as Assessors: Benefits and Drawbacks

A panel comprising employees is often referred to as an ‘internal’ panel. By way of their product experience, employees should have heightened awareness of the relevant sensory characteristics and variability which potentially should help expedite panel training. In theory, they should be readily available but in practice their primary responsibilities can cause participation in tests to be sporadic which in turn impacts on the efficiency of administering the tests and the quality of the resulting data. There is a heightened risk of biased evaluations, in particular from expectation bias due to assessors having prior project or product development knowledge. Hierarchy within an organization may also give rise to problems. In particular, junior staff may become more reticent during discussion in the presence of their superiors. Costs are hidden so running expenses are difficult to measure accurately and can appear deceptively low.

3.2.2 Dedicated Assessors: Benefits and Drawbacks

The term 'external' panel is often used for one comprising dedicated assessors, whether the assessors are directly on the business payroll or that of an employment/outsourcing company. They are solely employed for sensory evaluation which allows test time to be managed with greater efficiency. Training and validation can be more easily established as an integral part of the job which in turn can aid the optimization of assessors' sensory proficiency. A dedicated panel requires management time to organize, run and maintain it so a dedicated panel leader/manager will also be required. Costs are explicit, which enables accurate budgeting, and on the face of it can make this option appear more expensive than the former. In fact, this is generally a misconception when hidden costs, such as non-panellist employee time, are taken into account.

3.2.3 Agency Panels

Some companies may choose to leave the training, running and maintaining of their sensory panel/s to a sensory research agency. The agency would oversee and provide guidance about the operation of the panel/s but would not be involved in the day-to-day activities. Companies that take this option are more typically multinationals for whom sensory analysis is an integral part of the product development process and who consequently have a constant, high demand for sensory analysis information. Other companies that may adopt this route are those needing a 'one-off' study or those who need descriptive analysis on an infrequent basis, neither of which warrants the investment of setting up their own descriptive panel. Although costly, this option may prove a better way overall for a business to utilize time and resource within its product development team.

3.3 Panel Leader

A panel leader is essential to the process of setting up and running a descriptive analysis panel. The panel leader is the person who works directly with a panel on a day-to-day basis. Their key responsibility will be to ensure that the panel is trained and performs to the best of its ability and hence operates efficiently.

Panel leader and sensory analyst/manager roles may be performed by the same person, especially if the sensory department is small and the workload fairly light. Typically, though, the person will be dedicated to the specific role and report to the sensory analyst/manager whose responsibility will be the overall functioning of the sensory unit/department.

3.3.1 Required Knowledge and Skills

To be competent, a panel leader requires good interpersonal, communication and IT skills along with a sound knowledge of scientific practice as scientific rigour is a very important aspect. The person is usually a postgraduate from a science background with an interest at least, if not experience, in sensory science.

Whether the person recruited will need to have previous panel leader experience will largely depend on how much time can be allowed for them to acquire essential basic knowledge versus how quickly they are needed to competently perform. Part 2 of ISO 13300-2 (ISO 2006) provides a good reference for general guidance on the recruitment and training of panel leaders. Training courses specifically tailored for panel leaders are available.

The training style to be used by a panel leader will depend on that required for a specific descriptive analysis method, that is, to facilitate or be directive. The QDA and Spectrum methods illustrate the distinct differences between these two approaches; in the former, the panel leader takes the role of facilitator and is not an active participant, while in the latter they are highly involved, specifically directing the training. In the case of hybridized profiling methods which use a combination of elements from specified descriptive analysis techniques, the panel leader may choose to flex between the two distinct styles according to which they think is most effective at given points during training. Irrespective of approach, strong interpersonal qualities are essential such as active listening, patience, diplomacy, etc., along with verbal communication skills and a good amount of enthusiasm.

3.3.2 Defining the Panel Leader Role

Key responsibilities of the role are as follows.

- Train the panel, and validate and monitor their performance.
  - Define and maintain a catalogue of training aids, e.g. raw ingredients, competitor products, modified samples such as spiked or aged samples, with appropriate instructions about how to obtain, store, prepare and serve.
  - Obtain a sufficient range of product samples to build a standard frame of sensory reference for each product category as relevant.
  - Record and develop sensory attribute descriptions and lexicons as appropriate.
  - Define methods for validating initial performance and also monitoring it both short and long term (Chapter 4 addresses this area in detail).

- Ensure the safety and well-being of the panel.
  - Maintain food hygiene standards.
  - Ensure health and safety procedures are understood and followed. This should include a procedure for dealing with an adverse reaction, which in the first instance typically requires the assistance of a trained first aider to initially decide the best treatment. A list of names and contact numbers should be highly visible for everyone to refer to; some companies also have a push-button paging system set up for emergencies. Applicants who stated that they had allergies or intolerances should have been screened out at the initial recruitment stage, but the possibility of a reaction occurring in the future in an 'approved' assessor still needs to be addressed.
  - Ensure adequate information is available if novel ingredients/foods are to be tested. This would require a safety risk assessment so that assessors could be informed about the given level of risk involved as they *must* give informed consent before commencing with any sensory assessments. Depending on a company's ethical policy, the risk assessment data may require ethical panel approval before sensory evaluation can be requested. Guidelines for Ethical and Professional Practices for the Sensory Analysis of Foods published by the Institute of Food Science and Technology (updated 2015), give very helpful advice and provide references to other important sources of information.
  - Record and track assessors' health, in particular any changes such as a new medical condition or change to existing medication, and review with the sensory analyst if any action is required.
- Set up and run sensory assessment sessions using computerized data capture software.#
  - Ensure the correct products and sufficient quantities are available to complete the task.
  - Ensure methods, attributes and instructions are appropriate and current.
  - Ensure the correct assessment procedures and protocols are maintained.
- Analyse and report results.
  - Maintain standard procedures and formats.
  - Ensure data are filed and stored according to company policy. The panel leader may only be required to prepare the results or conduct the initial stage of data analysis ready for the sensory analyst who will possibly apply more complex data analysis before interpreting the results and completing a report. In other situations, the panel leader may be given the full responsibility; to a large extent, it will depend on their level of sensory knowledge and experience together with the level of importance attached to the particular sensory project.
- General management of assessors.
  - Track attendance.
  - Provide feedback about performance, informal. e.g. after training exercises, and formal. e.g. performance reviews.
  - Deal with poor performance.
  - Deal with unacceptable behaviour.

Performance issues are discussed further in section 3.8.

3.4 Recruitment and Screening Programme

Whether using employees or dedicated assessors to establish the panel, a clearly defined recruitment and screening programme is essential. Thorough planning of the programme ensures that all of the important assessment criteria are included, and that the complexity of the logistics is fully appreciated, an aspect that is often liable to be underestimated. A carefully planned screening programme should help ensure that the most suitable candidates are selected to progress on to be trained, which in turn will aid the effectiveness of the training stage.

It is also important to ensure that human resource personnel are involved in the recruitment and screening programme from the very start. They will be able to provide valuable guidance about various elements of the process such as wording of items such as recruitment adverts, acceptance/reject letters, job specification, etc., selection plus analysis of questions and exercises to assess personality traits and communication of feedback.

3.4.1 Recruitment Numbers Versus Size of Panel

A descriptive analysis panel typically involves 10–16 assessors but numbers can range from as little as six to 20 or more, depending on product and objectives. In general, panels involved in the evaluation of non-food products such as hair care, oral care, shaving products, etc. tend to be larger than those used for food and drink. This is to increase the statistical robustness (a) in order to help counteract the increased potential for noise from additional factors, such as skin type, hair growth, hair type, (b) when it is impossible or impractical for each panellist to evaluate more than one sample in a given test session, such as shave panels, hair care panels.

It is advisable to aim for a panel with a sufficient number of assessors to ensure that there will always be the preferred minimum number available, for example a panel of 14 to ensure a minimum of 10 assessors for any given project.

The number of people initially recruited needs to be well in excess of the final number required as a high percentage of applicants are typically eliminated at the initial screening stage. Although there is no absolute rule, experience has shown that the target number needs to be multiplied by a factor of at least 10, for example recruit 140–200 people initially to have 30–40 at the main screening stage if the aim is to finally have a panel of 14.

Rather than employing a fixed number of assessors at a time for a defined number of hours per week, some companies choose to have a pool of assessors and enlist the required number on a project-by-project basis. A major risk with this approach is that not all assessors will be equally used over time and differences in proficiency and experience will become exaggerated. Plus the level of commitment is likely to be lower and hence panel attrition higher.

3.4.2 Advertising for Assessors

Whether recruiting for a panel internally within a business or externally, an advert that defines the job role, the location and working hours plus application details is required.

Popular means for advertising externally are via local newspapers, local trade magazines, parish notice boards, word-of-mouth, etc. Some companies also use job centres, social media and/or local radio to promote the posts. This general approach means that all types of people are likely to apply. If the product category is highly specialized or there are ethical or legal restrictions related to the product use, for example, baby foods, sensitive skin products, alcohol, tobacco, dietetic foods, etc., recruitment will need to be more focused plus restrictions may need to be applied, such as who can be targeted and hence where the advert can/cannot be placed.

Advertising for internal recruitment is commonly done via staff notice boards plus face to face or via short presentations to departments. Collaboration from senior managers and heads of departments is usually required to allow access across business sections. Having support across an organization from the start can be a huge advantage in getting sensory analysis accepted and established as an integral part of business activity.

3.4.3 Application Form

The application form is used as part of the initial screening process and should include the key criteria that will enable the least suitable applicants to be quickly identified. In addition to sections that capture standard information about the applicant, such as name, address, contact details, education and qualifications, employment history, points that need to be addressed upfront are:

- general state of health, in particular any allergies or intolerances
- attitude towards use of foods/products, in particular food neophobia, strong dislikes or likes
- availability/conflicting commitments
- any previous work experience relating to the product area and/or sensory analysis.

Plus a question asking for a brief description about their reason for applying could provide some initial insight about their descriptive ability and level of interest.

3.5 Initial Screening

An initial screening stage should be designed to aid the efficiency of the main screening sessions via elimination of the least suitable applicants. Ideally, it consists of two parts: (a) marking of the responses on the application form followed by (b) a short interview by phone or face to face, with those who pass part (a). Any queries about the responses on the application form can be probed and clarified plus information can be gathered about additional factors such as smoking habits, dental health, regular use of medication. Care needs to be taken to ask the questions tactfully. If they are preceded by a short explanation about why they are required then applicants are usually more willing to answer. The initial screening can be done by personnel either from within a business or an agency specifically employed to assist with the recruitment.

3.5.1 Assessment of Application Form Responses

- *Allergies*: as with the recruitment of people for consumer tests, most companies take a blanket approach to a question about allergies. Such is the risk that anyone recording an allergy, whether it is directly related to the product category to be tested or not, is usually excluded.
- *Intolerances*: may be handled with more leniency as they are usually less critical to health. If a positive response is recorded but it does not appear to relate to the product category then the decision might be made to approve the applicant for this criterion pending further probing during the telephone/face-to-face interview.
- *Attitude towards the use of foods/products*: applicants with strong aversion to specific foods/products, in particular the products to be assessed, should be eliminated. Any cultural, lifestyle or other reasons that prevent certain types of foods/products from being assessed also need to be established, such as non-meat eaters.
- *Availability*: the screening process together with the time commitment required for the role should be clearly explained. Applicants who work rotating shifts, have to travel frequently or have a demanding primary job may not be best suited to the role of sensory assessor.
- *Previous related working experience*: if an applicant has some previous experience of analytical sensory assessment, it should aid their understanding of the basic concepts of the work. However, the likelihood of their experience biasing their judgements needs to be considered; this is potentially more likely with internal recruited candidates. Any risk of conflict of commercial interest also needs to be addressed.

3.5.2 Assessment of the Initial Interview

The main objectives of this part of the initial screening process are to:

- clarify any outstanding queries about intolerances, attitudes to food or products, and availability
- probe factors related to dental health, smoking, use of medication, use of perfume/deodorant to determine if any pose a potential issue
- gain some initial insight about the applicant's verbal communication and descriptive skills. For example, applicants could be asked to talk about a favourite food and to describe which of the product's sensory characteristics they particularly like
- enable the applicant to ask questions and learn more about the job.

After completion of the second part of the initial screening, applicants can be placed into one of three possible categories.

- Those who have clearly passed all criteria
- Those who have been eliminated based on one or more of the key criteria
- Those for whom some degree of doubt remains but at a less critical level

How many applicants are picked from the third group to go through to the main screening stage will depend on how many are in the first group. Their suitability can be prioritized based on how important the reason/s for doubt is considered to be.

3.6 Main Screening and Selection of Assessors

The main screening stage is usually the most intensive part of the panellist selection process. It involves taking applicants through a series of exercises that are designed to determine:

- sensory impairment
- sensory acuity
- ability to describe sensory observations
- behavioural traits, e.g. ability to listen and concentrate, interest and motivation, willingness to co-operate and work as part of a team, etc.

3.6.1 Screening Tests

Many sensory analysis textbooks provide information about screening tests, basic taste identification, odour recognition, flavour and texture description plus discrimination and or matching exercises being the most common examples cited. Specific tests together with the necessary materials and preparation protocols that can be used to assess sensory impairment, acuity and descriptive ability across all sensory modalities are detailed under the panel screening and training sections in BS ISO 8586 (ISO 2012a). This standard is recommended as a useful reference when planning this stage. The American Society for Testing and Materials (ASTM) special technical publication (STP 758) (ASTM 1981) is another good reference for general guidance.

3.6.2 Sensory Impairment

An important aspect of the main screening process is to determine if any of the candidates have any sensory impairment that will affect their ability to become fully trained assessors. Tests that have been adopted for screening sensory in/sensitivity are outlined below.

3.6.2.1 Sight

The Ishihara Colour Vision Test invented by Dr Shinobu Ishihara in 1917 is widely used to detect colour vision deficiency. The full test consists of 38 specifically designed, multi-coloured plates; a smaller 24-plate version is also available.

The task involves either reading a number or tracing a line on each plate. Either version together with instructions can be obtained from a range of suppliers or downloaded free from the internet. If the latter option is used, a high-quality colour printer is needed to ensure that the resulting plates are accurate.

The Farnsworth–Munsell 100 hue test (Farnsworth 1957) is another popular colour vision test. It is a hue discrimination exercise, also called an arrangement test, in which the aim is to place a series of colour plates of a similar hue in the correct order. Any misplacement can indicate some degree of colour vision deficiency. Whereas the Ishihara test only detects red/green deficiencies, this test covers a wider spectrum and is a useful screening tool if the ability to assess subtle differences in a product's colour will be required.

3.6.2.2 Smell

Odour recognition and odour matching exercises (such as described in BS ISO 8586 (ISO 2012a), ASTM STP758 (ASTM 1981) and many of the general sensory science textbooks) are commonly used and usually wholly suitable for detecting weaknesses in smell function.

If a precise measure of smell function/dysfunction is required then application of the University of Pennsylvania Smell Identification Test (UPSIT) developed by Professor Richard L. Doty in the early 1980s could prove beneficial. Smell function/dysfunction can be quantitatively measured using the UPSIT which is a simple, self-administered exercise involving four booklets each containing 10 'scratch and sniff' odorants. Four descriptions are presented with each stimulus and the task is to make a choice about which is the correct one even if no odour is perceived, that is, forced choice. The results enable five levels of smell function to be defined ranging from normal function (normosmia) through to total impairment (anosmia). As normal smell acuity can vary by age (Doty et al. 1984), this is factored into the method for evaluation of the test results (Doty 1995). The test is available in American, Arabic, British, Chinese (Simplified), Chinese (Traditional), Dutch, French, German, Italian, Japanese, Portuguese and Spanish. The UPSIT may be an appropriate screening tool if sniffing will be a major part of the product sensory assessment process, for example in the assessment of fragrances, alcoholic beverages, deodorants, etc. Personal experience has also shown the test to be of benefit when recruiting sensory consumer panels.

3.6.2.3 Taste

Taste recognition and threshold detection tests as described in BS ISO 3972:2011 are commonly used to assess taste acuity. The standard describes how to prepare a series of solutions for the five basic tastes, by dissolving specified chemical compounds in water. Assessors' taste is standardized against the plain water used to make the solutions at the start of each test. Sensitivity can be determined for both detection and recognition thresholds. The lower the concentration at which a difference in taste compared to plain water is detected, the greater the sensitivity; the same applies when determining the recognition threshold. Detection and recognition thresholds may be at the same concentration point for some people. Complete ageusia, that is, loss of all taste sensations, is a rare neurological condition but partial loss/insensitivity (hypogeusia) is more common.

It is worth considering during the planning process whether some taste sensations should be weighted as more important than others, ready for when the results are assessed; for example, sensitivity for bitter, acid and sweet might be deemed more important than for salt and umami when screening for a soft drinks panel.

Screening specifically for bitter blindness, which is associated with certain genetic conditions, may be relevant in some instances such as when bitterness is a key characteristic of the product category, for example lagers, dark chocolate, coffee, etc., or control of bitter taste/aftertaste is an important consideration, such as potable water, oral care products, etc. The compounds metallic phenylthiocarbamide (PTC) and 6-n-propylthiouracil (PROP), a related bitter chemical, are used to assess bitter blindness. Edible paper taste strips dosed with either PTC or PROP together with blank control strips are commonly used for the exercise. The control strip, which helps eliminate any false positives, is first placed on the tongue for a few seconds and any taste sensation noted. The control strip is removed and the test strip is then assessed in a similar way. Those who are insensitive to the bitterness of these compounds are called non-tasters. The use of PROP has become more popular over PTC as it is less toxic, plus it enables the correlation between the number of taste buds and degree of sensitivity to be identified. Those who detect bitterness most intensely, often finding it intolerable, are called 'supertasters' and have a greater number of taste buds than average (Bartoshuk et al. 1994). Those who detect bitterness but find it tolerable are called 'medium tasters' and possess a typical number of taste buds whereas non-tasters tend to have fewer taste buds than average.

Phenylthiocarbamide and PROP taste strip test kits are both commercially available. Full health and safety data should be provided by a supplier along with the test kit. It is important that assessors are made aware of the level of risk to health associated with the test and allowed to abstain if they choose; that is, informed consent is obtained.

3.6.2.4 Touch

Ranking tests based on the method described in standard ISO 8587 (2006) are commonly used to evaluate texture discrimination ability by mouth or by hand. Four to five samples are selected that have subtle differences in the intensity of a specific texture attribute, such as firmness, chewiness, smoothness, etc., as appropriate for the particular product category. The samples are presented in a random order under code to potential assessors who are required to rank them in order of increasing intensity for the specified attribute. It may be practical to assess discrimination ability of a few texture attributes using the same sample set, each of the specified attributes being ranked individually. It is advisable to qualify an attribute name with a standard definition to help standardize applicants' comprehension of it. For the purpose of screening, it may be most effective to focus on a product's most distinctive attributes.

Whilst a wealth of published research exists about the oral processing of foods and measurement of sensory texture perception, the methods used tend not to have been adopted as regular screening tools. They are either impractical for routine use or are too complex to enable the mechanism for an observed behaviour or correlation with sensory texture perception to be established (Stokes et al. 2013). The measurement of saliva flow rate is one approach sometimes used for screening and could be particularly relevant where assessment of astringency will be a key factor, such as in wines. People can be classified into high flow or low flow; those having low saliva flow demonstrate a longer and more intense response to astringency stimuli (Dinnella et al. 2009). Saliva flow can be measured unstimulated but use of a stimulus facilitates measurement, citric acid being a general choice due to its ease of use and the high level of saliva response (Engelen et al. 2003). Saliva is collected from each applicant at regular intervals over a specified period, for example every 10 seconds over 60 seconds. The net saliva weight from each applicant is statistically analysed versus time to determine flow rate.

Three texture assessment tests, which arose from a project designed to study the effect of ageing on tactile sensitivity and masticatory performance, that lend themselves to use as screening tools (Fillion & Kilcast 2001) are oral shape recognition, particle size discrimination and chewing efficiency. The principle of the oral shape recognition test is to identify five sugar letters (Figure 3.2), A, P, O, S, H, by oral touch only. A warm-up letter (U) is given first to familiarize assessors with the protocol. The assessor's task is to recognize each letter when placed in the mouth, with the help of show cards, 10 letters on each card, of which only one is correct. The scoring system is based on the level of difficulty of recognition assigned to each of the five letters, with O being easiest and S most difficult, and points are only awarded for certain letters other than the correct ones as shown in Figure 3.3.The particle size discrimination test comprises three pairs of powdered sugar of different grades, from coarse to very fine. The task is to recognize, within a pair of samples, which one has the finest grain (smallest size of particle) by oral touch only. Points are awarded for each correct identification, the maximum score being 6, that is, all three pairs correctly identified.

The chewing efficiency test uses two-colour (pink and white) chewing gum. Assessors chew the gum for a defined number of chew strokes; the length of time taken to complete the required number of chews is also recorded. The chewed gum is then evaluated and scored against an eight-point photo scale to identify which is the closest match: 1 showing the two colours barely mixed, 8

Figure 3.2 Alphabet sugar letters.

LETTER	ANSWER	POINTS
A	A	3
	V	2
	Y	1
P	P	2
	Y	1
O	O	1
S	S	5
	J	4
	F	1
H	H	4
	N	3
	M	1

Figure 3.3 Scoring scheme for the oral shape recognition test. No points are allocated to any other letter selected. Points allocated to each of the five letters are added to give the total score (minimum score = 0, maximum score = 15).

showing the two colours well mixed. Half points are allowed if the colour of the mix is judged to be in between two of the photos. Full details of each of these three tests are given in Filion & Kilcast (2001).

Tactile sensitivity of the fingers can be assessed using the Touch-Test® Two-Point Discriminator or the Semmes–Weinstein five-piece hand aesthesiometer. The tests are straightforward to administer and interpret, and full instructions are provided.

3.6.3 Behavioural Traits

Certain personality, traits especially interest, motivation and co-operation, are just as important as an assessor's sensory acuity. In their paper about the selection of sensory assessors, Piggot and Hunter (1999) proposed that concentration and personality tests are better predictors than intricate sensory screening procedures of the future ability of an individual to perform well as a panellist. A calm yet confident disposition is also desirable as assessors need to be able to articulate their perceptions clearly and ask questions without dominating the group.

It is highly advisable to include a group assessment exercise in addition to the sensory exercises in the main screening programme to help assess character traits such as team working, communication skills, diplomacy and tact. These exercises typically involve applicants being given some information relating to a business scenario before the group exercise begins, with a short amount of time allowed to read it and make notes. The group is then instructed to discuss the information provided and reach a conclusion over a specified time after which they are asked to explain the reasoning behind their decision. The behaviour of each applicant is observed and noted during the whole process. From the author's experience, inclusion of this type of exercise not only proves beneficial in aiding the selection of the most appropriate individuals to progress onto the training stage but subsequently aids the effectiveness of the training itself.

The human resource department or sourcing agency should be able to help with the selection, application and assessment of a suitable exercise but generic ones, of which many examples are available on the internet, can be adapted.

3.6.4 Screening Programme and Timetable

Points to consider when planning the programme and timetable include:

- number of screening sessions required
- types of test exercises to be included
- time allocated for each test and length of overall session
- whether groups need to be subdivided in order to manage differing durations of tests
- meeting and greeting
- closing and departing
- venue, range of facilities and support required
- implication of other aspects within the business, such as parking, security, health and safety, etc.

3.6.5 Number and Length of Overall Screening Sessions

The number of main screening sessions required will depend on the number of candidates going forward from the initial screening stage divided by the maximum number that can be processed at any one time, for example 38 candidates in total divided by 10 maximum per occasion = 4 screening sessions.

The maximum number of candidates that can be handled on each occasion will largely depend on the size of the available facilities and the number of staff that can be dedicated to the task. Tests such as basic taste identification, odour recognition and discrimination exercises are typically conducted in sensory booths or partitioned areas, as each candidate needs to complete them discretely, whereas descriptive exercises and personality tests, whether conducted as one-to-ones or as groups, can be managed in standard meeting rooms. As a key tip, full consideration is advised when planning the number and type of staff required for the preparation, running and overseeing of the tests. Experience has shown that the support required especially for the latter point, can be easily underestimated. If time allows, it is best to run a pilot session beforehand using internal personnel to ensure all goes well on the day.

From experience, the total time required for a screening session involving 10–12 candidates, especially if a behaviour assessment exercise/s is to be included, is typically around 3 hours.

3.6.6 Selection Criteria and Mark Scheme

An objective mark scheme should be set up that enables each candidate's results to be formally recorded and easily interpreted. Such a scheme can be easily set up using Excel (Figure 3.4). The specific choice of test materials, which type of discrimination test/s are considered most appropriate and whether certain sensory modalities are given more focus than others will be decided at the discretion of the panel leader/sensory manager and depend on the type of products to be routinely assessed. For example, it may be decided to include an odour description test and an odour matching test; the first could show a general range of common odour compounds (recommend six maximum per test to minimize sensory adaptation and fatigue), whilst the latter could focus on ones that will be routinely experienced. If products have strong flavours or mouthfeel characteristics that linger and are likely to cause carry-over issues, it could influence the choice of discrimination test/s.

When marking results, tests that have definitive answers, for example basic tastes, odour matching, discrimination tests, etc., can be awarded 1 point for a correct answer. A graded marking approach can be used for descriptive exercises such as odour identification and sensory modality description, where 3 = most accurate/good and 1 = least accurate/poor. If several people are involved in the marking, they should all be instructed to interpret the grading in a similar manner so that they apply it fairly and consistently. A simple grading approach can also be used to record the outcome of personality exercise/s. A written comment such as

Candidate Name	Basic Taste Tests: 1 = correct						Total Basic Tastes	Discrimination Tests: 1 = correct				Total Discrim' Tests	Odour Description: 3 = exact, 2 = close, 1 = attempt						Odour Matching: 1 = correct						Total Odour Tests	Descrip tive Tests	Total Descrip tive Tests	Ranking Test: 1 = correct					Total Ranking Tests	Total Score (Max possible =)

Figure 3.4 A basic format for creating a screening mark scheme in Excel.

'suitable', 'suitable with some reservation' or 'unsuitable' can be recorded alongside each candidate's total score or colour coding, for example green, amber, red, can be used to convey the same. As having the preferred personality traits is as important as having the required sensory acuity, any candidates who are judged to be 'unsuitable' should be eliminated before the results of the sensory tests are assessed.

As Figure 3.5 illustrates, totals can be recorded for each exercise as well as the overall total, enabling potential weak versus strong abilities to be readily identified for each person; 70% is commonly set as the desired pass level with 60% being the minimum. In the illustrated example, candidates would need to achieve a total of 34 points to meet the desired pass level. Candidate B clearly passes with a total of 39 and candidate H just meets it with 34. Candidate E passes the minimum level with 32 points (65%) and would therefore be kept in reserve. The most suitable candidates to progress from the 'reserved' list can be prioritized by identifying those who have performed best in the tests that are considered most critical for selection. The need for selection from the reserve list will depend on how many candidates pass at the desired level.

3.6.7 Formal Offer of Employment

Once the selection process is complete, it is advisable, prior to sending out the formal written offer of employment, to contact each of the chosen candidates by phone to ensure that they still wish or are able to take up the role.

Under the guidance of human resources, an employment pack should be compiled. This will include items such as:

- job offer letter with rate of pay, and start date and time
- employment pack, including standard terms and conditions of contract, holiday entitlement, benefits, parking and security information, confirmation of acceptance form for signing and returning.

It is advisable to plan to take a panel through the company induction process when they first start. For efficiency, this is usually completed with the whole panel over their first few sessions of attendance. The following areas are typically addressed.

- Information about the company as appropriate, such as mission statement, product range, how sensory analysis information is used by the business, growth and performance, etc.
- Health and safety protocols, both general and those specific to the job role.
- Security aspects, such as sign-in/sign-out protocol, importance of confidentiality, use of computers, etc.
- Medical records and checks, importance of keeping information current, if ongoing health checks will be necessary.
- Reiteration of points from the company's standard terms and conditions of contract as appropriate.

An induction period can provide a valuable settling-in phase for assessors before they commence with the sensory training intensive.

	Discrimination Tests: 1 = correct					Odour Description: 3 = exact, 2 = close, 1 = attempt				Odour Matching: 1 = correct					Descriptive Test 2			SCORE TOTAL
Candidate Name	Test 1	Test 2	Test 3	Test 4	Total Discrim' Tests	Odour 1	Odour 2	Odour 3	Odour 4	Odour 1	Odour 2	Odour 3	Odour 4	Total Odour Tests	3 = good, 2 = fair, 1 = poor	Total Descriptive Tests		Possible Max Score = 49
A	1	0	1	0	**2**	1	3	2	2	1	1	1	1	**12**	2	**3**		**26**
B	1	0	1	1	**3**	2	2	1	2	1	1	1	1	**11**	2	**5**		**39**
C	0	0	0	0	**0**	2	2	2	2	1	1	0	1	**11**	1	**3**		**24**
D	0	1	1	1	**3**	1	2	2	2	1	1	1	1	**11**	2	**4**		**28**
E	1	0	1	1	**3**	1	2	2	2	0	1	1	1	**10**	2	**5**		**32**
F	1	1	1	1	**4**	1	2	2	1	1	1	0	0	**8**	1	**2**		**20**
G	1	0	1	0	**2**	2	1	2	3	0	1	0	1	**10**	2	**4**		**27**
H	1	1	0	1	**3**	1	2	2	2	1	1	1	1	**11**	3	**5**		**34**

Figure 3.5 An extract from a screening mark scheme in Excel showing total scores for individual tests as well as the grand total score achieved by each candidate.

3.7 Panel Training

Three core stages are involved in training a panel to reach a desired level of competency in descriptive analysis although the specific approach and time allocated to each stage will vary according to the particular descriptive method.

1. *Descriptive stage (qualitative).* Standardize assessors' understanding and perception of the sensory attributes relevant to the test set of products.
2. *Scale and rating stage (quantitative).* Develop assessors' ability to rate attribute intensities consistently.
3. *Panel performance validation stage*. Verify assessors' ability to meet a required level of performance prior to the formal evaluation of the products.

3.7.1 Key Aims of Training

In addition to developing assessors' ability to describe and discriminate stimuli, training aims to:

- develop an objective assessment approach across the panel
- fully determine strengths and weaknesses of each panellist's discrimination ability
- develop the concept of scales and the ability to quantitatively rate sensory attribute intensity consistently
- standardize assessors' 'frame of reference' of the sensory attributes
- instil test and assessment procedures and protocols
- build assessors'confidence in their assessment abilities.

With new assessors, it is advisable to spend some time when they first start, for example 8–9 hours, familiarizing them with the basic principles of sensory evaluation and the need for assessments to be as objective and unbiased as possible. This can help them to accept certain procedures more readily which may seem unnecessary and even bizarre at first, and also help them to start to better understand how to apply their own senses.

The facets of the three stages of the panel training and validation process are discussed below and summarized in Figure 3.6.

3.7.2 Descriptive Stage

The purpose of the descriptive phase is to establish a comprehensive set of attribute terms that accurately describe all relevant sensory characteristics for the product range to be evaluated.

It is important that attribute terms should be as objective and precise as possible. Assessors usually grasp the concept of objectivity relatively quickly and understand that subjective terms such as like, good, bad, nice, pleasant, etc. are not acceptable. However, the ability to identify distinct attributes from within a complex stimulus is very difficult (Murray et al. 2001). The process of perception is complex and it is important for panel leaders to understand that

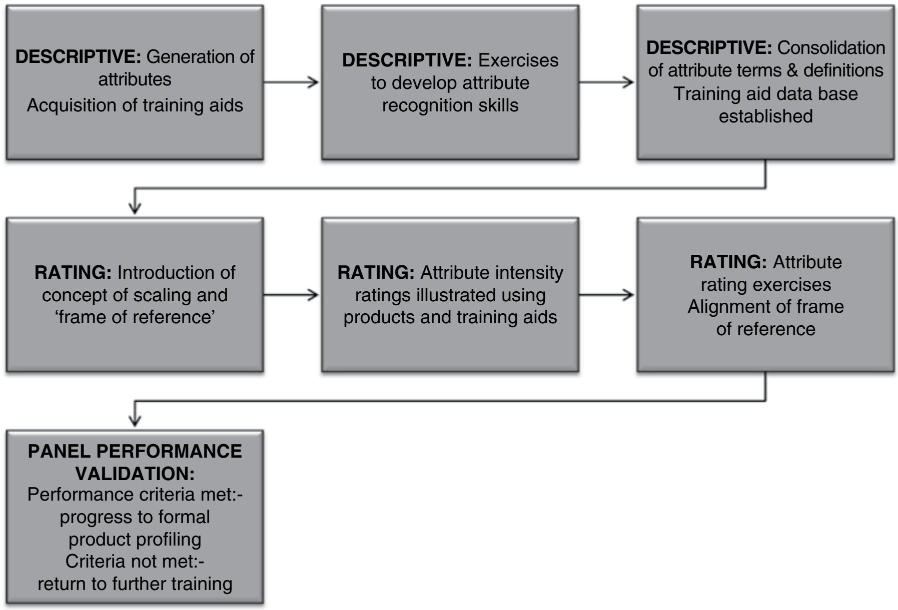


Figure 3.6 A summary of the key facets of the training process.

initially, descriptions for a given stimulus will vary between assessors due to differences in their inherent sensitivity to it, their level of familiarity with it, and their basic natural ability to clearly describe what they perceive (Meilgaard et al. 2007).

3.7.2.1 Generation of Attribute Descriptions

To aid the process, it is best to first show assessors products that demonstrate clear differences to enable them to easily recognize and describe some of the attributes which in turn will help them develop confidence in their abilities from the start. To help identify the most relevant descriptors and also avoid verbal bias, it is advisable to get assessors to independently write down their own descriptions prior to any group discussion. The most popular descriptions can then be identified by tallying the number of assessors who have used each description for each product. This can be done by the panel leader during a session or afterwards for feedback at the next session, as best suits. This also enables descriptions to be identified and discussed that potentially relate to the same stimulus which in turn helps to start the standardization of attribute terms and agreement of the most apt definitions.

In some instances, such as when new assessors are being integrated into an existing panel or to comply with certain descriptive analysis techniques, assessors will be trained to use an existing set of attribute terms and definitions from the start. In this case, the panel leader will take a directive approach, explaining and defining each attribute term whilst using a specified range of samples to illustrate each characteristic.

3.7.2.2 Association Effects

Assessors' descriptions are also likely to be affected initially by 'association effects', a psychologically based bias whereby the description relates to another product that is more familiar to the panellist, rather than the specific stimulus; for example, 'vanilla' may initially be described as ice cream, custard or cookies. Or a stimulus may spark the generation of additional descriptions not actually present in the product, for example the fruity character of plain chocolate triggering 'orange flavour' to be described as well.

3.7.2.3 Distinguishing Attributes

It is advisable to use a single term for each attribute to help assessors with the mental process of focusing onto a specific stimulus. Avoid 'bundled' terms that could be easily misinterpreted even after training. For example, 'confectionery' would be better broken down further, such as sweet, estery, candy, as appropriate. A tiered approach may prove useful in helping assessors with the mental process of attribute identification: a general 'umbrella' term is used to start, becoming more specific depending on how well the stimulus signal can be perceived, for example, fruity >>citrus, red fruit, dried fruit >>lime, raspberry, prune. Care should also be taken by the panel leader on how to use product information such as knowledge of a product's ingredient list, as this may create false expectations of what supposedly should be described. Coriander, for example, is used in a wide range of products from breads to alcoholic beverages. In many, it will be unidentifiable specifically as coriander, but will contribute at a subtle level to characteristics such as floral, citrus, sweet, musty. In other products, it will be distinctly perceivable and hence the name may then become one of the sensory attribute terms.

3.7.2.4 Multisensory Perception

Over the past decade, significant progress has been made in understanding the multisensory perception of flavour. Some odours are consistently described as 'sweet' or 'sour', terms traditionally only classed as taste sensations in sensory evaluation, and have been shown to enhance or suppress overall sweetness perception (Auvray & Spence 2008). The findings have also shown that culture, such as western versus non-western, plays a role in the perceived extent to which specific odours affect sweetness perception. Besides having important implications in helping developers produce healthier foods without lessening flavour quality (Spence 2010), the author suggests this area of psychological research will help sensory scientists to better adapt training methods to standardize perceptions and terminology by advancing understanding of psychologically based biasing factors.

3.7.2.5 Training Aids

Physical training aids help the panel leader consolidate assessors' perception and understanding of specific characteristics. A wide range of materials can be used; the following lists some popular sources.

- Raw ingredients, e.g. herbs, spices, oils, cereals, nuts, fresh and dried fruits,10
- Odour and flavour references from external suppliers and/or within the business
- Visual and colour standards from external suppliers and/or those developed within the business
- Commercial competitor products
- Modified products, e.g. aged, spiked with additional amounts of an ingredient or an extra ingredient
- Base formulation without, for example, final colour/perfume/flavouring/texture components added

The panel leader will often have to be creative and experiment with various samples to establish what works best for a given product range. A reference list of all proven training aids, detailing the product category/ies and attribute to which they each apply, should be set up and maintained by the panel leader. Additional information for each sample should include, as appropriate, supplier, reference code, preparation instructions, storage area, disposal date, disposal instructions plus the current standard definition used by the panel.

3.7.2.6 Consolidation of Attributes

During the course of attribute generation, the panel leader will:

- collate the list of attributes generated by the panel and revise as appropriate after each training session until a comprehensive, final list is agreed
- discuss, advise and agree with the panel applicable definitions for each attribute
- discuss and agree the most applicable assessment order for the attributes to appear on the formal ballot
- discuss and agree assessment protocol along with any specific modifications required
- record any particular strengths or weaknesses that have been observed in skill to perceive specific attributes. This will start to build an understanding of the panel's as well as each assessor's potential abilities.

Sensory lexicons are also useful aids to help the panel leader with attribute generation and consolidation as well as with the phrasing of definitions. Drake and Civille (2003) give a comprehensive summary of flavour lexicon development and application. Odour matching exercises (using approved odour compounds obtained either from a flavour house or from the raw materials department of a business) against a written list of standardized attribute terms can also be used as training progresses to provide a guide of assessors' recognition abilities.

3.7.3 Attribute Rating Stage

3.7.3.1 Quantitative Rating Scale

The most common form of rating scale used with trained descriptive analysis panels is a continuous line scale. The scale is usually unipolar with anchor points qualified by descriptions that relate to the extreme values of the attribute being assessed. The anchor points can be placed at the very ends of the scale or indented as shown in Figure 3.7. Positioning the anchors slightly in from the ends of the scale helps to avoid the psychological bias of central tendency (Lawless & Heymann 1999) and maintain linearity (Stone et al. 2012). The area outside the anchor points provides a mental 'comfort' zone for assessors which allows for the potential rating of future samples that may be considered more extreme in intensity than previously experienced. The line scale typically represents 100 discrete points from anchor to anchor although the values do vary according to the specific descriptive method, for examaple 15 points in the Spectrum Method. Assessors are taught to put a vertical mark at the point on the line that corresponds with the perceived intensity of an attribute, that is, rate intensity. The mark is then converted into a numerical value. Conversion of the value is usually done automatically via a computerized data capture system of which there are now several sophisticated commercial versions available.

An overview of other types of quantitative response scales can be found in BS ISO 4121 (British Standards Institution 2003).

3.7.3.2 Frame of Reference

A key aim of the quantitative rating stage of training is to standardize assessors to a common mental frame of reference against which to rate attribute intensity. If left to rate products without any training, assessors would use their own personal benchmark and the output would be highly variable due to differences in extent of product knowledge, such as experience of the variation in sensory attribute intensities common to a specific category of products.

Assessors' product knowledge will to some extent have been developing through exposure to product samples and additional training aids during the descriptive stage. This can be progressed by introducing comparative exercises in the form of paired comparisons and ranking. Experience has shown that it is advisable to work on two or three specific attributes at a time to start with,

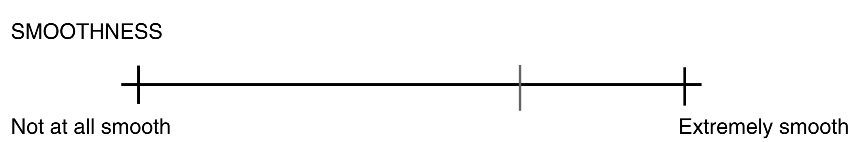


Figure 3.7 A line scale with indented anchor points and descriptive labels. The vertical line towards the top end of the scale marks the point at which the intensity of the attribute has been rated.

using three or four products that illustrate high versus low ends of the scale, progressing onto examples of more subtle differences. Assessors make their decisions independently on paper then discuss openly as a group. The panel leader collates the results during the discussion and informs assessors of the expected results so they can compare them against their own. Care needs to be taken in the choice of wording for anchor points as this can also affect assessors' comprehension of the frame of reference for an attribute and how they use the scale. For example, if it is possible for an attribute to not be present then it would be best to label the lower end of the scale as 'none' or 'not at all' as opposed to 'very weak' or 'very slight'.

Initially, to help build confidence, assessors can be allowed to use the scale simply as two halves: lower versus upper. Then as attribute recognition skills improve and the frame of reference becomes more aligned, they can be guided to use the scale in a more precise way. The panel leader can use these results as an informal check of panel performance to help them decide when the panel is ready for a formal performance validation check prior to the formal evaluation sessions.

Murray et al. (2001) suggested that a training regime that combines product-specific training and the use of training aids may improve panel performance. How and to what extent products and training aids are used will be for the panel leader to judge and he/she will be influenced by specific method and level of performance required.

When new assessors are being trained to join an existing panel, the use of established training aids and reference samples can prove invaluable in achieving efficient development and alignment of their frame of reference with that of the existing panel. The approved samples can be readily selected to demonstrate the scale anchor points for defined attributes, and also preferably illustrate some different points along the scale too.

Note: acquisition of product knowledge is an ongoing process and therefore routine assessment activities should help in refining assessors' frame of reference.

3.7.4 Performance Measurement Stage

This stage provides a statistical measure of the whole panel's and each panellist's performance for defined quality criteria. The descriptions used for the performance quality criteria may vary from text to text but primarily they refer to similar aspects. Whether assessing the panel as a whole or individual panellist performance,the following abilities need to be checked:

- discriminate between products when a true difference exists
- provide reproducible measurements that fall within accepted ranges, i.e. precision
- provide reliable data so that the panel average can be used as a representative measure, i.e. consistency of agreement in the way products are rated.

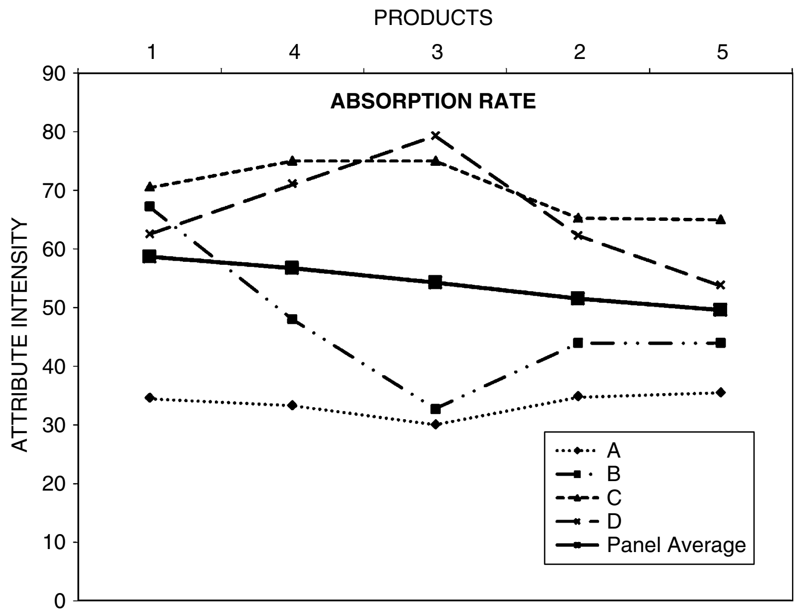


Figure 3.8 'Absorption rate' performance interaction plot showing four panellists' intensity ratings for five products.

Two factors that are regularly observed with new assessors' data are disagreement in the way that products are rated, that is, cross-over effects, and pronounced differences in the use of the scale. The former has more serious implications than the latter on how confidently the results can be used to support business decisions but both issues need to be minimized as much as possible over the training period. Figure 3.8 shows an interaction plot for four assessors who rated five hand cream products for 'absorption rate'. A distinct example of cross-over effect can be seen for product 3 plus there are pronounced differences in the use of the scale and panellist A is 'flat lining', that is, not discriminating between the products. Panel performance is weak overall and the result is poor discrimination between the products, shown by the 'panel average' line. The assessors would benefit from further training on interpretation of the attribute term and developing their frame of reference.

Figure 3.9 shows an interaction plot for visual 'shininess', again for four assessors scoring five products. Although there is some slight inconsistency in the ranking of products 2 and 3, neither the cross-over effects nor the differences in the use of the scale are as pronounced as shown in Figure 3.8. There is much better discrimination between the products overall, and although performance could still be improved, the results indicate that alignment in assessors' frame of reference is developing for this attribute.

Gaining consistency of agreement is another major challenge, especially when the signal from a stimulus is subtle and possibly close to threshold recognition

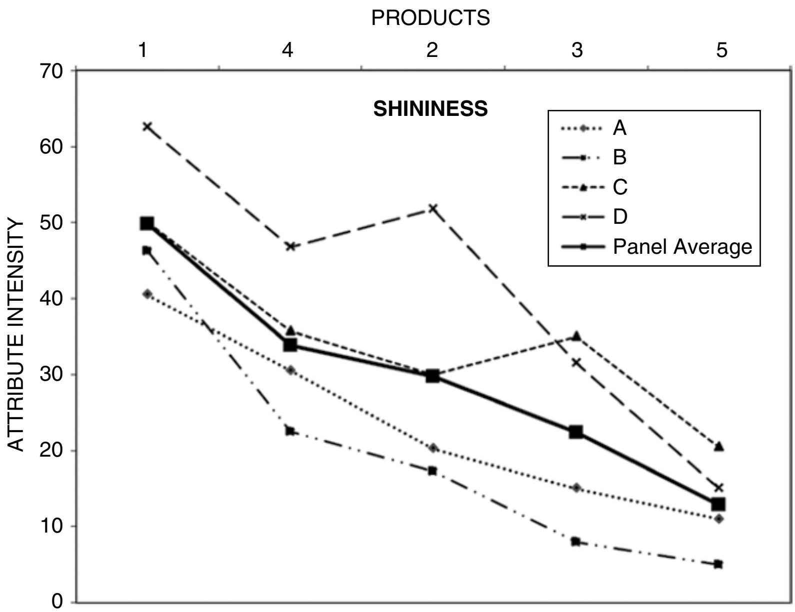


Figure 3.9 'Shininess' performance interaction plot showing four panellists' intensity ratings for five products.

point for some assessors. Perception at threshold is not a constant for a given substance (Meilgaard et al. 2007). It may constantly change from being none distinguishable to readily distinguishable and the problem may be exacerbated further if the stimulus is very volatile or there is difficulty in homogenizing the product sample. These factors need to be taken into account when defining performance criteria and interpreting results.

3.7.4.1 Selection of Products to Assess Performance

The selection of 3–4 products from those used during training is usually adequate and it is recommended that the test set should demonstrate statistically significant differences for at least a third of the attributes in the finalized list. These attributes can be used as the key performance indicators. To make a confident choice, the panel leader should refer to the informal results captured during training along with their own product knowledge; it may be useful to consult with other members of the sensory or development teams as well for guidance. Each product should be assessed at least in duplicate according to agreed protocols and standard sensory practice.

3.7.4.2 Interpreting Results

Three-factor ANOVA is most commonly used to analyse the data. This enables product, panellist and session interaction to be assessed for each attribute across the product set and identify how well both the whole panel and each panellist has performed for the key performance indicator attributes. A minimum 'pass' level should be defined, such as number of attributes required to record acceptable performance. If the required minimum is not reached, further training is usually conducted focusing on the attributes that have caused the most issues, with further priority, especially if time is limited, for those that are considered most important to the project needs. Most computerized data capture systems come with panel performance software that enables performance charts, such as interaction plots, to be easily prepared – a very useful aid when giving assessors feedback.

Figure 3.10 shows an example of a panel performance summary chart, produced using SENPAQ, licensed data analysis software supplied by Qi Statistics Ltd, UK. This type of chart is useful for providing a quick synopsis of assessor performance across the set of attributes used for a specific study. Positive results indicate good performance. Three performance criteria are used: discrimination ability, repeatability and consistency. Each assessor's performance is measured relative to the panel average. In this example, it can quickly be seen that there is wide variability in assessor ability and overall performance is poor. Assessors A, C, D, E and J show poor discrimination ability to varying degrees but panellist E

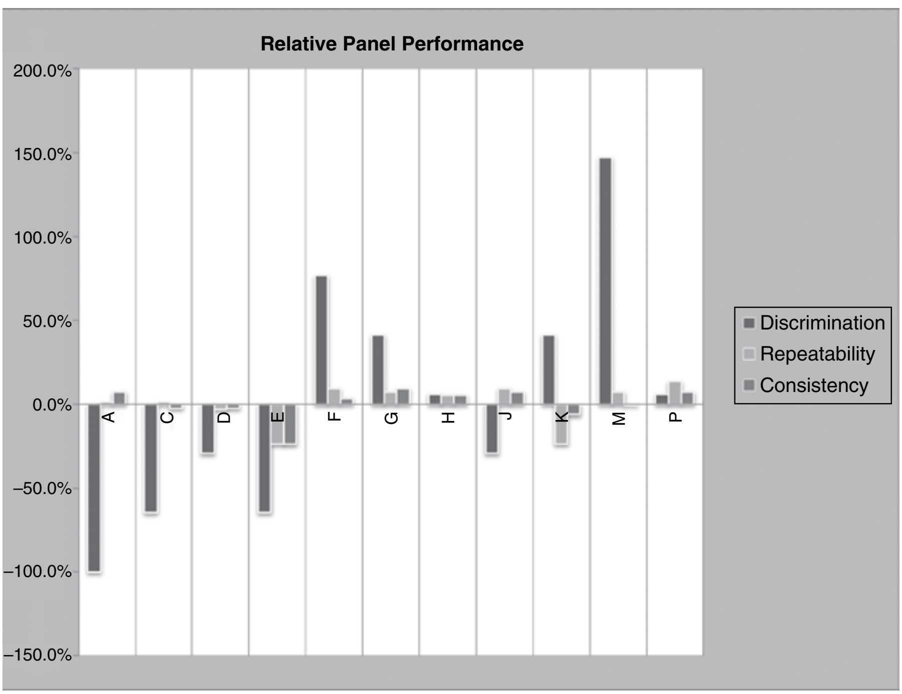


Figure 3.10 An example of a panel performance summary chart produced using SENPAQ, licensed data analysis software supplied by Qi Statistics Ltd, UK. This type of chart gives a top-level synopsis of each panellist's performance across the set of attributes used for a specific study.

shows the worst overall performance as they also have poor repeatability and are contributing the most to interaction, that is, lack of consistency. Although assessors H and P are not rated negatively for any of the criteria, the low levels for each suggest that they are not discriminating between the products very well either. To fully understand specific issues, deeper investigation using interaction plots and/or tables showing assessors' performance for each attribute would be required.

ISO 11132 (ISO 2012b) provides a useful guide for the panel leader to appraise performance. Chapter 4 of this book discusses quality measures and analysis of panel performance in further detail.

3.7.4.3 Training Period

An extensive period of time is usually required to fully train new assessors for the majority of descriptive methods and hybrid forms. From experience, around 80 hours minimum is suggested as a guide to achieve a reliable and consistent level of discrimination ability. The Spectrum Method could take even longer. Time and budget constraints may impose limitations on the amount of training that can be achieved and the compromise may be that a lower level of panel performance has to be accepted.

3.8 Panel Motivation

Recruitment of assessors who have an enthusiasm for the role and genuine interest in the product area pays dividends in helping maintain motivation over time. Selecting assessors with a good team player attitude is another valuable asset as it helps to build rapport and empathy across the panel which in turn can increase the sense of gratification in the role.

Recognition and reward schemes and team building events are often cited as motivational aids along with activities designed to improve sensory awareness, such as an informal talk from a member of the development team or visit to a production area or supplier to learn how sensory information benefits their work. Motivation has been well researched in performance psychology and education and sport science. The findings, which suggest an important role for autonomy, competence and relatedness, can be applied to sensory panels to increase intrinsic motivation (Lund 2009). Lund's research revealed that external panels were found to be more intrinsically motivated than internal panels, and that they had an increased perception of competence.

3.8.1 Feedback

Feedback plays an important role in conveying a sense of competence, an important motivating factor (Meilgaard et al. 2007). From the start, it is advisable to explain that feedback is an integral part of the training process, and indeed the continued monitoring of panellist performance. Assessors should not feel threatened by it when problems need to be addressed. In order to help them feel comfortable with all aspects of feedback, they should be encouraged to view it constructively as a means to help them develop their full potential as an assessor. This calls on the interpersonal skills of the panel leader.

Informal feedback will occur naturally during discussion sessions and again, if there is good rapport within the group, assessors should be able to openly discuss issues and share learning.

During training, assessors can be familiarized with the official style of feedback that will be used when a formal performance check has been conducted. Instruction is first given by the panel leader about how to interpret the performance charts that are routinely used. It is best if assessors are allowed to study their individual results first before the start of open discussion. This will help them be mentally prepared to positively address the weaker aspects of their performance. By the end of the training process, panellists should be fully prepared for the feedback they will routinely receive as professional assessors.

3.8.2 Poor Performance

Formal record keeping of performance needs to be set up from the start of training and be routinely recorded. This enables assessment and feedback to be kept as factual and objective as possible based on defined, measureable criteria, and is invaluable if having to deal with a case of consistent poor performance. The situation should have been detected via the routine monitoring process and the panel leader should first use one-to-one feedback and discussion to make every attempt to understand and help rectify the issue. If this fails, the human resource department may need to be consulted and the formal disciplinary procedure started.

3.9 Panel Attrition

The departure of assessors, although unavoidable, should only occur infrequently, especially if they continue to find satisfaction in the role. It is not usually cost-effective to recruit for just one new panellist so the decision of when to initiate a fresh recruitment drive can be difficult. In practice, recruitment takes place when a few new assessors are required, perhaps 3–4, but before the panel is reduced below the minimum number required for its reliable functioning, for example 8–10 for food and beverage panels, 12–16 for panels working in other product categories such as hair, personal care, etc.

3.9.1 Integration of New Assessors

A fundamental problem when integrating new panellists is that they may be overawed by the thought of the existing assessors' experience even if they have completed an initial training regime and performance check prior to being integrated. Again, the importance of ensuring via the screening process that assessors have the desired interpersonal skills can help minimize the risk of any sense of hierarchy occurring. The panel leader should, however, acknowledge the experience of existing panel members and encourage them to share their learning to help clarify points as appropriate for the new ones, a scenario that should help with both motivation and integration.

Experience has shown the value of using a classic team-building exercise in aiding integration. This would be set up and conducted by a person/people with a personnel background, away from the sensory facility. These exercises allow the group to get to know each other in a fun way in a neutral context. The author recommends that an activity that is new to all assessors is used when the new group is ready to be fully integrated. If the suggested team-building exercise is not feasible, a sensory exercise involving products from a category different from those of the business could be an alternative.

3.10 Summary

The impact that the assessor screening and selection process can have on the subsequent effectiveness and stability of the resulting sensory panel should not be underestimated. The time required to implement a successful screening programme can be more than justified. The choice of sensory screening tools, how the five senses are prioritized and the extent to which each is evaluated will largely depend on the product category/ies for which the descriptive analysis panel is required. In addition, the importance of measuring applicants' behavioural traits should not be overlooked.

Central to the whole screening and training process is the panel leader. They play a crucial role not only in training the panel to a reliable level of performance as effectively as possible but also in ensuring the panel's well-being, physically and mentally. A suitable individual therefore needs to be competent as both scientist and mentor.

During the training stage of the process, the fundamental challenge for a panel leader is to develop the panel to a defined, standard level of assessment proficiency. To achieve this, assessors have to develop specific abilities, of which key ones are: recognize and differentiate specific attributes, rate differences in attribute intensity, and align their ratings to a standard frame of reference for each attribute. A formal validation test of the panel's proficiency is conducted once a panel leader is satisfied that all assessors can provide reliable and reproducible ratings of differences between a set of products in at least the attributes that have been defined as key performance indicators. This test can also form the basis of the formal panel performance monitoring scheme.

Once the panel's performance is validated, both specific projects and ongoing routine product assessments will help to further develop and refine assessors' skills as their product knowledge increases.

3.11 The Future

Developments in statistical software packages designed to measure panel performance have gained impetus over recent years and are likely to continue as panel leaders demand quicker ways to enable them to reliably check and validate performance. Graphical output of data has been enhanced, enabling the information to be more readily interpreted and facilitating speedier feedback. Training courses that specifically address panel performance are also on the rise.

Ongoing research in areas such as those described below could lead to changes in the way that assessors are screened and trained in the future – for example, the development of more precise and practical screening tools and more efficient training processes based on a deeper understanding of the physical and psychological mechanisms affecting perception.

- *Multisensory perception*: new insights about, in particular, flavour perception and the effect of cross-modal interactions between the senses.
- *Neurological measurements*: improved knowledge of how the brain reacts to sensory stimuli.
- *Taste genetics*: deeper appreciation of the effect of genotype on the perception of taste, for example bitterness, and mouthfeel, for example creaminess, astringency.
- *Oral processing*: advances in understanding the effect of actions of the mouth during eating and swallowing on flavour and texture perception.

The added value that the application of psychometric and group assessment tests can offer to the recruitment of assessors may see them increasingly used in the future to aid the selection of people with the most suitable personality traits (see section 3.6.3). The use of these techniques should assist the development of team spirit and commitment within the panel and help achieve its successful operation, ensuring that the time and effort invested in the total screening and training process are fully recognized.

References

American Society for Testing and Materials (ASTM) (1981) *Guidelines for the Selection and Training of Sensory Panel Members*. Special Technical Publication No. 758. Philadelphia: American Society for Testing and Materials.

Auvray, M. & Spence, C. (2008). The multisensory perception of flavor. *Consciousness and Cognition*, **17**, 1016–1031.

Bartoshuk, L.M., Duffy, V.B. & Miller, I.J. (1994) PTC/PROP tasting: anatomy, psychophysics, and sex effects. *Physiology & Behavior*, **56**, 1165–1171.

British Standards Institution (BS) (2003) *Sensory Analysis: Guidelines for the Use of Quantitative Response Scales*. BS ISO 4121. London: British Standards Institution.

British Standards Institution (BS) (2011) *Sensory Analysis. Methodology. Method of Investigating Sensitivity of Taste*. BS ISO 3972. London: British Standards Institution.

Dinnella, C., Recchia, A., Fia, G. et al. (2009) Saliva characteristics and individual sensitivity to phenolic astringent stimuli. *Chemical Senses*, **34**, 295–304.

Doty, R.L. (1995) *The Smell Identification Test Administration Manual*. Haddon Heights: Sensonics.

Doty, R.L., Shaman, P., Applebaum, S.L., Giberson, R, Siksorskim L. & Rosenberg, L. (1984) Smell identification ability: changes with age. *Science*, **226**, 1441–1443.

Drake, M.A. & Civille, G.V. (2003) Flavour lexicons. *Comprehensive Reviews in Food Science & Food Safety*, **2**, 33–40

Engelen, L., de Wijka, R., Prinza, J., van der Bilt, A. & Bosman, F. (2003) The relation between saliva flow after different stimulations and the perception of flavor and texture attributes in custard desserts. *Physiology & Behavior*, **78**, 165–169.

Farnsworth, D. (1957) *The Farnsworth-Munsell 100-Hue Test for the Examination of Color Discrimination*. Maryland: Munsell Color Company.

Fillion, L. & Kilcast, D. (2001) *Towards a Measurement of Oral Tactile Sensitivity and Masticatory Performance: Development of Texture Tests*. Research Report No. 781. Epsom: Leatherhead Research Association.

IFST (2015) www.ifst.org/knowledge-centre-other-knowledge/ifst-guidelines-ethical-and-professional-practices-sensory

International Organization for Standardization (ISO) (2006) Sensory Analysis. General Guidance for the Staff of a Sensory Evaluation Laboratory. Part 2: Recruitment and Training of Panel Leaders. ISO 13300-2. Geneva: International Organization for Standardization.

International Organization for Standardization (ISO) (2010) *Sensory Analysis Methodology. General Guidance for Establishing a Sensory Profile*. ISO 13299. Geneva: International Organization for Standardization.

International Organization for Standardization (ISO) (2012a) *Sensory Analysis. General Guidelines for the Selection, Training nd Monitoring of Assessors and Expert Sensory Assessors*. ISO 8586. Geneva: International Organization for Standardization.

International Organization for Standardization (ISO) (2012b) *Sensory Analysis. Guidelines for Monitoring the Performance of a Quantitative Sensory Panel*. ISO 11132. Geneva: International Organization for Standardization.

Lawless. H.T. & Heymann. H. (1999) *Sensory Evaluation of Food: Principles and Practices*. New York: Aspen Publishers.

Lund, C.M. (2009) Investigation of New Zealand Sauvignon Blanc wine using trained sensory panels. PhD thesis, University of Auckland. Available at: https://researchspace.auckland.ac.nz/handle/2292/5168 (accessed 31 May 2017).

Meilgaard, M.C., Civille, G.V. & Carr, B.T. (2007) *Sensory Evaluations Techniques*, 4th edn. Boca Raton: CRC Press.

Murray, J.M., Delahunty, C.M. & Baxter, I.A. (2001) Descriptive sensory analysis: past, present and future. *Food Research International*, **34**, 461–471.

Piggot, J.R. & Hunter, E.A. (1999) Review: Evaluation of assessor performance in sensory evaluation. *Italian Journal of Food Science*, **4**, 289–303.

Spence, C. (2010) The multisensory perception of flavour. *Psychologist*, **23**(9), 720–723.

Stokes, J.R., Boehm, M.W. & Baier, S.K. (2013) Oral processing, texture and mouthfeel: from rheology to tribology and beyond. *Journal of Colloid & Interface Science*, **18**, 349–359.

Stone, H., Bleibaum, R. & Thomas, H.A. (2012) *Sensory Evaluation Practices*, 4th edn. San Diego: Academic Press.

CHAPTER 4

Panel Quality Management: Performance, Monitoring and Proficiency

Carol Raithatha and Lauren Rogers

4.1 Introduction to Panel Quality Management

4.1.1 Definitions and Overall Approach

Panel quality management deals with the entirety of sensory panel performance management and as such covers the three areas of performance, monitoring and proficiency as depicted in Figure 4.1. Panel quality management can only be useful when the sensory panel has been properly trained: Chapter 3 gives more details on sensory panel training.

Panel quality management includes the following subdefinitions.

- *Panel performance* (see section 4.2): the day-to-day or project-to-project assessment of profiling data to evaluate their reliability and relevance, and to determine the need for any panel interventions.
- *Panel monitoring* (see section 4.3): the ongoing tracking and monitoring programmes associated with the longer term performance and validation of each panellist's capabilities, and the panel output as a whole.
- *Panel proficiency testing* (see section 4.4): a ring-testing process in which multiple sensory profiling panels participate by assessing a common sample set. The output from each panel is compared against an expected output to validate performance.

It is important to note that any issues discovered while assessing the sensory panel data may not be associated solely with the panellists' performance: they may also be due to preparation errors, sample variability, experimental design, etc. Therefore, it is important for the panel manager as well as individual panellists to adhere to the company's standards and practices. These may include, for example, laboratory procedures for sample preparation, labelling and presentation as well as procedures for the panellists such as sample assessment protocols.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

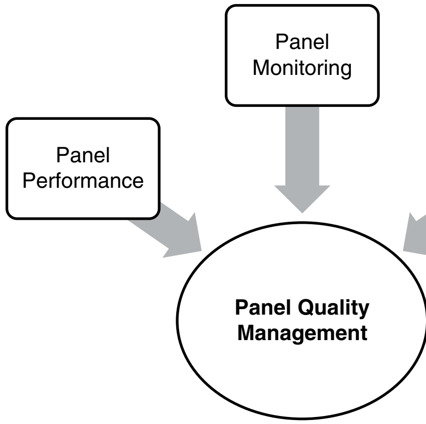


Figure 4.1 Schematic representing panel quality management.

Panel quality management is a vital part of any sensory science function but will also require other standardized approaches to be fully effective. These will be common to most sensory functions and will include aspects such as training programmes, use of sensory standards (International Organization for Standardization or ASTM International, for example), procedures, protocols and correct project design.

4.1.2 Why is Panel Quality Management Important?

In descriptive analysis, the panel is assumed to be providing analytical data much as any other piece of 'equipment' found in a quality or research and development laboratory. This implies that as well as calibration (training in the case of descriptive panels), validation and monitoring of the output of the panel must also be part of an overall quality management system.

A good panel quality management system will include control loops that measure and monitor performance of the panel and individual panellists and highlight the need for appropriate training or other action at the correct time. These feedback systems are used to ensure that the panel provides data that can be used confidently in business decision making. A panel quality management system allows the sensory manager to be aware of general data quality and helps to highlight issues that need attention. These systems may also lead to adjustments and improvements in training practices and methodologies used.

Despite providing analytical data, descriptive sensory panellists are human beings, employees and individuals with particular needs and motivations. Good panel quality management systems can help to ensure that individuals and the group perform to the best of their abilities and achieve high levels of job satisfaction.

4.1.3 Context and Panel Quality Management

Approaches to panel quality management will vary depending on context. The most important factors to consider are as follows.

- *Descriptive methodology*: because the training, procedures, data produced and use of outputs differ across descriptive methodologies, the approach to panel quality management will need to be adapted. For example, the Spectrum™ Method uses universal attribute intensity references, while quantitative descriptive analysis (QDA) works on a per sample set and relative basis. Because of this, panel quality management within the Spectrum Method context is likely to focus much more on panel precision and accuracy.
- *Business outputs required*: a related topic is the importance of business outputs in the design of a panel quality management system. For example, if the main outputs of a panel are sensory scores that will inform a quantitative process or consumer choice model, then performance measures related to precision, accuracy and reproducibility will be very important. From another perspective, if a panel is predominantly carrying out one-off benchmarking and/or concept assessments, the panel quality management scheme will focus on applicability in terms of discrimination and relevance.
- *Panel type*: panel quality management for a panel composed of those employed in the company for other purposes may, by necessity, be less comprehensive than that for a panel employed specifically for a sensory evaluation role. If an individual's main job role is as a sensory panellist, it will be appropriate for their performance assessment to include panellist quality management measures.
- *Whole panel versus individual panellist*: panel quality management measures for individual panellists are usually related to those for the whole panel, but tend to be statistically less strict.
- *Individual project versus ongoing panel work*: panel quality management within a project will normally focus on key panel and panellist measures that need to be met for data within the project to be considered suitable for use. Ongoing panel quality management takes a longer term process control monitoring type approach, looking at trends and performance over time and feeding back into longer term training and methodology development plans and cycles. These plans should be informed by past performance and highlight areas where improvements are needed or where good performance should be reinforced.
- *Individual panellist personal circumstances*: another aspect of panel quality management is the consideration of individual panellists' behaviour, motivation and personal issues which may affect performance, be it a one-off or an ongoing personal situation. A good relationship with the panel leader will help highlight any problem; in such instances, arrangements can be made to take any data issues into account. For example, this could include the removal of a panellist's data from a project if they have a head cold or closer examination of the data to ascertain if they can be used if the panellist has a personal problem that may be affecting their concentration.

Finally, the practical approach to panel quality management can vary depending on focus and resource. For example, if panel quality management is formally part of a panel leader or sensory analyst's responsibilities, time will be allotted to data collection and evaluation activities, and it will be possible to develop and maintain a more detailed system. In addition, the design and measurement of panel quality management systems can range in the level of complexity of the statistical focus. Unfortunately, panel quality management is not often given top priority and can be seen as an add-on cost rather than an inherent element of the sensory profiling process. In this chapter, the authors take a pragmatic view, describing a range of approaches but focusing examples on analysis and activities that might be possible within a typical, busy sensory testing environment, in which sensory analysts have some statistical knowledge but are not statisticians.

4.1.4 Key Approaches and Concepts in Panel Quality Management

The assessment criteria used within panel quality management are related to strategic and statistical concepts. Various authors and sensory professionals have suggested key measures and/or approaches and use a range of definitions for key terms.

Many of the early panel performance measures (Stone & Sidel 1985) were focused on individual panellists' means and subsequent rank orders of the sample set for each attribute in comparison to the rank order of the rest of the panel. Pritchett.Mangan (1992) stated that the true attribute mean is actually unknown, and hence viewed this approach with caution because it assumes that the panel mean is in some way an accuracy standard, that is, that the majority of the panel is the standard by which to measure the rest of the panel.

Naes et al. (1994) listed the main four ways in which panellists can differ:

- confusion about attributes
- individual differences in sensitivity to certain sensations
- individual differences in use of the scale
- individual differences in precision.

Wilkinson and Yuksel (1997) summarized the differences between panellists in scale use, listing them as:

- differences in what panellists perceive – their different thresholds
- panellists' ability to perceive differences in intensities of some attributes
- the use of attributes differently – scoring under a different attribute name for certain stimuli
- differences in use of the scale – location and range of the scale used.

Confusion about attributes can occur, for example during 'attribute dumping' when panellists find no suitable attribute for which to score a particular characteristic they detect or when one attribute is used to measure several different stimuli. Differences in precision are important when selecting and training panellists, and if a panel shows low precision for a particular attribute this can indicate that more training is required.

Practically, it is not sensible to force panellists to rate something they cannot detect and nor to force them to use certain areas of the scale, but the information about different sensitivities and differences in scaling is useful for the panel leader to aid with data analysis and interpretation.

Training and feedback is an important part of the overall panel quality management approach and concept. Kuesten et al. (1994) proposed the use of an automated feedback system for panellists to indicate how they had performed in various tests, including measuring basic taste intensities on an unstructured line scale. The authors' results indicated that panellists increased the amount of the line scale used and also became more precise after the automated feedback sessions. These types of automatic feedback systems are now a possibility with relevant software.

Performance monitoring requires some method of visualizing and assessing performance trends over time. Catchalian et al. (1991) used a quality control chart approach (from Statistical Process Control: SPC procedures) to measure panel performance, with the average central line on the plot the average of the means for all replicates, for all panellists for each attribute in turn. These charts were then used to show panellists their outlying scores and any attributes with wide confidence intervals. These attributes were then revisited in training sessions to improve the panellists' performance.

Panel quality management systems must account for several factors to encompass all aspects of performance. Pfeiffer et al. (2008) distinguished four technical areas in which, according to the authors, each individual of a sensory descriptive panel needs to be assessed.

- Discrimination – the ability to identify differences
- Validity – the ability to rank products in the correct order of magnitude
- Repeatability – the ability of a panellist to repeat their own results
- Agreement – the ability to be in line with the majority.

Tormod et al. (2010) described some of the main types of individual differences and these are given below.

- *Use of scale*: differences in mean and variability/range of the scores
- *Agreement/reproducibility across the panel*: disagreement in ranking of the objects
- *Reproducibility*: different level of precision – differences between independent replicates
- *Discrimination*: differences in ability to discriminate between products

Kemp et al. (2009) stated that the most important criteria for panel performance (in terms of sensory evaluation in general) are accuracy, precision and reliability. They defined accuracy as how close panellist data or panel mean data are to the *true* value (e.g. reference, spiked sample); precision as a measure of how reproducible panellist or panel data are; and reliability/validity as a measure of how close an individual panellist's score/judgement is compared to the rest of the panellists and panel mean.

Although the panel quality measures proposed are often quite different amongst authors, and in some cases the definitions of key terms are conflicting, there are underlying themes apparent. Taking this into account, as well as the authors' professional experiences, four main quality measures for panel quality management are considered important by the authors of this chapter.

- *Repeatability*: the concept that the repeated measurements **within a project** for individual panellists are within accepted ranges.
- *Reproducibility*: the concept that the repeated measurements **across projects** for both the panel and individual panellists are within accepted ranges.
- *Consistency*: the concept that there is sufficient agreement among panellists to use the panel average as a representative measure.
- *Discrimination*: the concept that a difference is detected when a real difference in samples exists.

Each of these aspects can be considered both qualitatively and quantitatively and they are discussed in full in sections 4.2–4.4.

In addition to the four measurements above, another important aspect of panel quality management is overall validity, that is, whether the data outputs of a sensory descriptive panel can be confidently applied to business action standards. Validity incorporates all the four measurements above but also relates to the extent to which attributes developed during the descriptive process are relevant to business objectives, and whether the overall conclusions taken from sensory projects relate back to the reality of the actual product set.

A final important concept linked to the human nature of sensory analysis is panel and panellist engagement. This is a measure of how much the panel and individual panellists are motivated and open to continuous improvement. Panel and panellist engagement is a mode of considering and/or measuring the likelihood that the panel shows attitudes and behaviours that will allow continued delivery of high-quality data.

4.2 Panel Performance

Panel performance measurement is the day-to-day or project-to-project assessment of profiling data to evaluate their reliability and relevance, and to determine the need for any panel interventions. These panel performance checks can be used during initial panel training, during the development phase of a profiling project prior to the final scoring sessions, and also prior to complete statistical analysis of the final data set in a particular project.

Panel performance measures are also the starting point for both panel monitoring and proficiency testing and therefore, the main elements discussed here will also be important in these later sections of this chapter.

The use of a panel performance check prior to full data interpretation for a project is critical. It would not be advisable to report results from a profiling project if the panel performance checks show severe issues. For example, panel performance checks should uncover issues such as:

- one or two panellists ranking samples in an opposite direction to that of the rest of the panel for a particular attribute, indicating a probable misunderstanding of the scale or attribute definition

- one panellist scoring in such a way as to inflate or deflate a sample mean, thus making it unrepresentative of the actual sample set
- attributes which appear to indicate the samples are not statistically significantly different but the underlying data show that in fact there may be sample differences. This can be caused by scoring issues such as bimodality, where scores are clustered at either end of the scale
- attributes which appear to indicate the samples are statistically significantly different but closer inspection reveals that only one or two panellists are contributing to this output.

Panel performance also includes the regular panel feedback element of panel quality management and as such, allows the panel manager to improve both the individual panellists' and the panel's performance as a whole. Collecting panel performance data alone will not help improve the panel's performance; regular and timely feedback is essential (Findlay et al. 2007).

As mentioned earlier, the authors have summarized the main elements of panel performance as being the measurements of repeatability, reproducibility, consistency and discrimination. Validity and panel/panellist engagement are also important considerations.

There are a multitude of approaches for measuring panel performance, and it is difficult to be prescriptive. Table 4.1 gives some examples of possible panel performance measures with respect to key parameters; however, these may well

Table 4.1 Example panellist and panel performance parameters.

Performance parameter	Panellist measure options	Panel measure options
Repeatability (assuming replicates exist)	• Replicate range for an attribute/product combination • Root mean square error (RMSE) for an attribute compared to the attribute mean or the scale range or in comparison to other panellists or the whole panel value	• RMSE values for each attribute compared to the attribute mean and/or the scale range • Significance of replication effect in three-way ANOVA
Discrimination	• Difference in sample means for an attribute • Significance of product effect in one-way ANOVA by attribute	• Significance of product effects in ANOVA • Discrimination or not of key attributes where expected • *Post hoc* comparisons between samples show expected differences or not
Consistency	• Significance of contribution to overall panel interaction for an attribute • Correlation of individual sample means to panel sample means for an attribute • Inspection of panellist by product plot (clear cross-overs may indicate possible issues with consistency)	• Significance of panellist*sample interaction effect in ANOVA

change with project context and objectives, and whether the focus is on panel or panellist performance. Many of the measurement options presented in Table 4.1 are discussed within this chapter.

It is worth noting that the areas of the panel performance elements listed in Table 4.1 are not stand-alone. For example, a panellist may produce excellent precision in their replicates mainly because they use the same part of the scale for all replicates and all samples, therefore not showing any sample discrimination. Or a panellist may show excellent sample discrimination but order the samples in a different way to the rest of the panel.

Checking the output from analysis of variance (ANOVA) can be useful prior to detailed panel performance checks as it will give an idea of the number of attributes showing significant differences between the samples and also the residual error for each attribute after panellist and sample variation have been accounted for. Performing the ANOVA to determine if expected differences have been found will also give the analyst an idea of data validity and relevance to project action standards. In addition, simple raw data visualizations (such as distribution graphs) will highlight attributes with issues such as bimodal scores, those that have very low scores or zeros for all samples, or those where panellists are using large ranges.

Once these initial evaluations have been carried out, more detailed panel performance checks can be performed. These are explained in the sections below.

4.2.1 Repeatability

Repeatability is the concept that the repeated measurements *within a project* for individual panellists are within accepted ranges.

Repeatability generally refers to the closeness of measurements by a panellist for each of the replicates within a profiling study. These replicates may be performed on the same day, in the same week or in some extreme cases, for example in personal care products, they may be weeks apart.

For a well-trained panel which works on a number of different product types, a general rule of thumb for replication is that less than 20 points difference between the replicates (measures of the same sample for the same attribute) on a 100-point scale can be considered good replication, and less than 10 is excellent replication. However, this rule of thumb will vary from panel to panel depending upon the level of panel training, the profiling methodology used and the objectives of the specific test.

4.2.2 Reproducibility

Reproducibility is the concept that the repeated measurements *across projects* for both the panel and individual panellists are within accepted ranges.

Reproducibility can be regarded as the ability of the panel to reach similar conclusions to previous projects they have carried out on similar product types. Checks on reproducibility may be carried out by performing profiles on the same set of samples, or a sample set with spiked samples, on a regular basis. Reproducibility is a criterion related to panel performance monitoring and is, therefore, discussed in more detail later in this chapter.

4.2.3 Consistency

This is the concept that there is sufficient agreement among panellists to use the panel average as a representative measure.

Consistency is concerned with the panellists' ability to agree with each other and how an individual panellist agrees with the panel as a whole. Some authors refer to this as reproducibility. Panellists will ideally agree on sample order where differences in samples exist for certain attributes, and not contribute significantly to any panellist*sample interactions.

Another element of this panel agreement stems from the attribute generation, definition and scale usage or calibration stage. If there is no consensus, or if panellists do not understand these aspects from the start, it does not matter how much training they have had, they will not be able to agree amongst themselves and produce a valid sensory profile of the samples in question. It is well worth ensuring at an earlier project stage, by conducting practice profiles or simple sample ranking tests, for example, that all panellists are in agreement prior to scaling. If the panellists do not agree on the sample order at this early stage, for example which sample is the sweetest and which the least sweet, this can also cause issues in the data analysis stage.

Another performance consideration worth mentioning here relates to scale use and consistency in scale use. Panellists in QDA-type profiling may use the scale differently from other panellists, both in range of scale used and also the portion of the scale used. However, this should not be an issue in terms of panel performance as long as they are consistent within their own scaling. Having said that, a panellist who uses the scale differently from the rest of the panel can cause issues; for example, if a panellist tends to use only part of the scale, they may find it harder to differentiate between samples, particularly when the sample differences are small. A restricted scale range can also cause issues by elevating or decreasing the panel mean score.

4.2.4 Discrimination

Discrimination is the concept that a difference is detected when a real sample difference exists.

This is another critical area for both individual panellists and the whole panel, as even one panellist who does not discriminate, when the rest of the panel is able to, can affect the whole panel's output and subsequently change project decisions. Discrimination is simply the panel's and each panellist's ability to find differences between samples in the sample set for the attributes they have generated. Obviously, for some attributes some samples may not be different and therefore the panel cannot be expected to discriminate. However, if the panel leader knows that there are differences, perhaps from discussions during the attribute generation stages, practice profiles, rankings or because the samples have been adjusted or spiked to be different, then any lack of discrimination can be highlighted.

Discrimination ability can be affected by repeatability and consistency, and so is closely linked to these performance areas. Lack of discrimination can occur through poor replication, low sensory acuity for that particular attribute or being unable to use the scale effectively. For example, if a panellist's replicates for one sample vary by a large portion of the scale (consider a sample A, with replicates of 10, 40 and 70), this may well prevent any sample discrimination, regardless of any further tighter replications for other samples (consider a second sample B with scores of 40, 42 and 44). This is because the spread of the results for sample A would encompass the results for sample B.

4.2.5 Overall Validity

This is the measure of whether the data outputs of a sensory descriptive panel can be confidently applied to business action standards.

Validity incorporates discrimination, precision and consistency. It also relates to the extent to which attributes developed during the descriptive process are relevant to business objectives, data collected are representative and analysed in a valid manner, and whether the overall conclusions taken from sensory projects relate back to the reality of the actual product set. Setting action standards and related testing objectives can go a long way to ensuring validity, as discussing plans for each possible profiling outcome can help select the correct sensory approach early on in the project.

4.2.6 Examples of Applied Panel Performance Measurement

This section outlines an approach to the overall process of panel performance measurement and gives data-driven examples, from a QDA-type profile, of possible evaluation strategies. Figure 4.2 outlines this approach which includes the following steps:

- evaluate general data quality
- evaluate repeatability;
- evaluate discrimination;
- evaluate consistency
- evaluate overall performance and validity.

A data set with five samples, two replicates and nine panellists has been selected to give worked examples for each aspect of panel performance checks.

4.2.6.1 Evaluate General Data Quality

This step involves looking at the whole data set to ascertain if the results were gathered correctly (for example, no sample presentation issues), are robust enough to use (in terms of initial assessment of panel performance), relevant to the project (meet the action standard requirements) and that the data set is

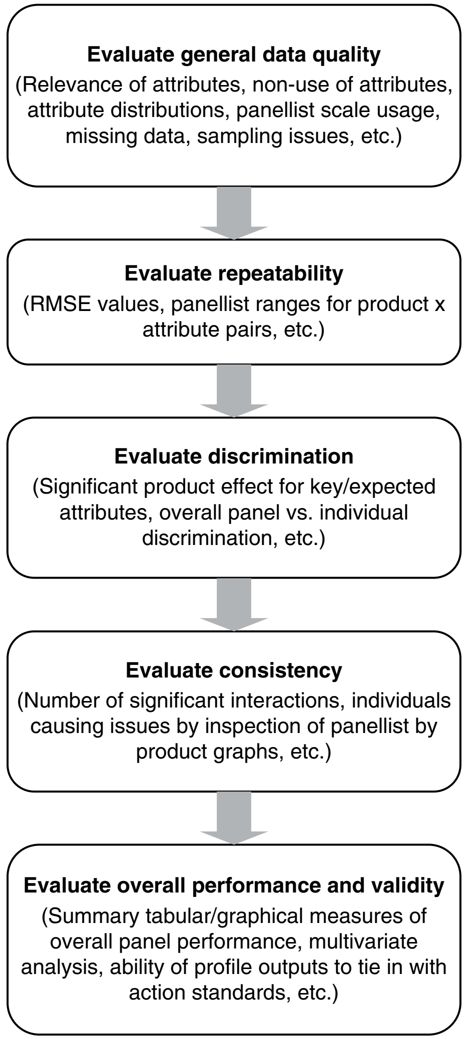


Figure 4.2 Overall process of panel performance measurement.

whole (no missing data). A first check might be to look at tables of raw data as shown in Table 4.2 (shown for one attribute), raw data distribution graphs as in Figure 4.3 (shown for one attribute) and/or a profile summary table, an extract (appearance attributes) of which is shown in Figure 4.4.

Table 4.2 gives a typical output for one attribute, 'crispiness', measured for five samples (A–E) with two replicates performed by each of nine panellists. Tables such as this are useful for examining the raw data, highlighting issues and allowing a quick overview of sample differences by replicate. For example, it is possible to see that there may have been problems with the assessment of sample A as there are several replication issues for this sample.

Table 4.2 Raw data for crispiness attribute (FIZZ).

Sample A				Sample B			Sample C			Sample D			Sample E		
Panellist	**Rep. l**	**Rep. 2**	**Mean**	**Rep. l**	**Rep. 2**	**Mean**	**Rep. l**	**Rep. 2**	**Mean**	**Rep. l**	**Rep. 2**	**Mean**	**Rep. l**	**Rep. 2**	**Mean**
1	23	42	32	64	72	68	30	64	47	27	28	27	43	24	33
2	0	18	9	13	22	18	0	36	18	8	14	11	14	10	12
3	3	46	24	25	37	31	29	10	19	30	31	30	4	1	3
4	5	60	32	10	22	16	27	34	30	30	29	30	39	9	24
5	8	40	24	4	31	18	41	17	29	13	17	15	5	5	5
6	13	41	27	20	53	36	14	40	27	38	13	25	12	25	18
7	18	21	19	29	20	24	18	28	23	23	29	26	0	11	5
8	20	20	20	20	31	25	20	21	20	21	21	21	20	21	20
9	51	12	31	51	28	39	19	40	29	4	13	8	13	0	7
Mean	16	33	24	26	35	30	22	32	27	21	21	21	17	12	14

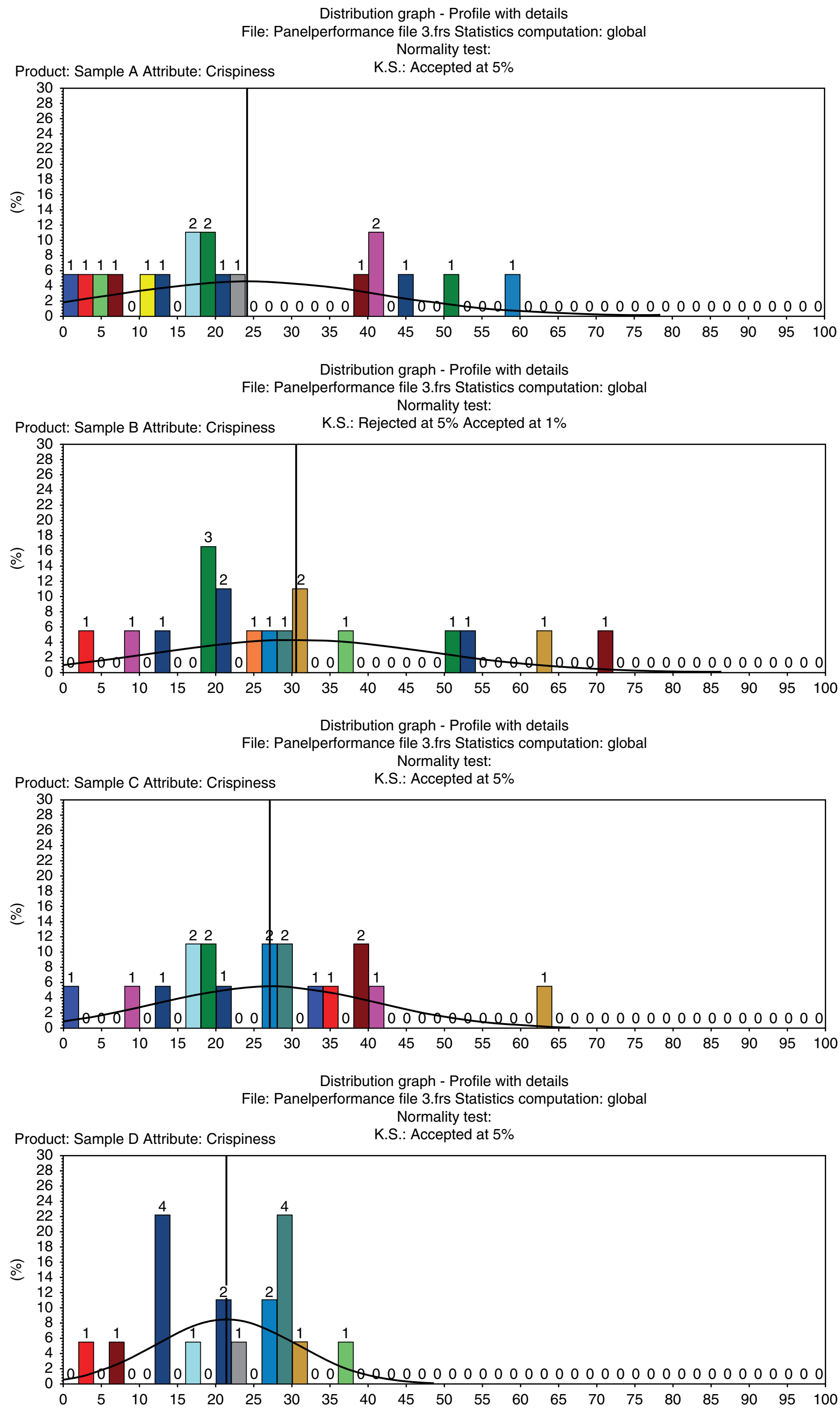


Figure 4.3 Distribution graphs for crispiness (FIZZ).

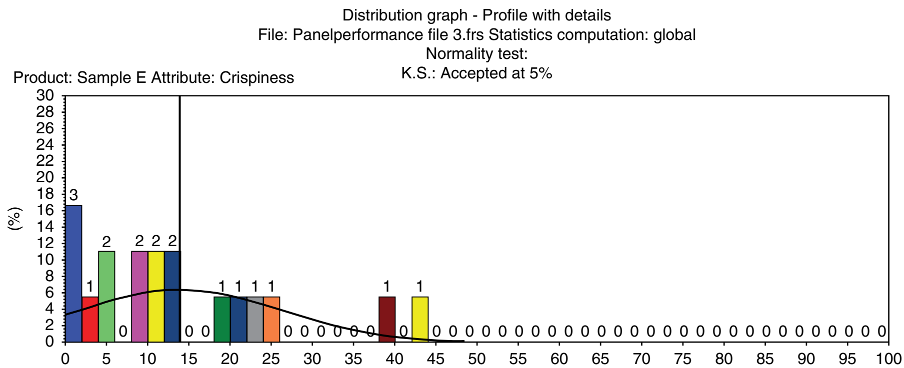


Figure 4.3 (Cont'd)

	Sample A	Sample B	Sample C	Sample D	Sample E	LSD	Prob	Scale Type	Low Scores	Interaction F-value	Interaction p-value	RMSE	Pre-Scaling
Intensity of Golden Colour	43.6	68.4	57.8	40.8	29.9	4.9	<.0001	0---100	0.0%	1.3	0.2079	6.4	No
Number of Dark Edges	44.8	80.7	77.7	70.2	38.9	8.6	<.0001	0---100	0.0%	1.8	0.0325	9.4	No
Darkness of Edges	48.4	69.8	61.4	49.2	49.5	12.2	0.0026	0---100	0.0%	2.2	0.0064	12.0	No
Number of Specks	50.9	48.3	49.8	37.8	70.8	6.1	<.0001	0---100	0.0%	1.2	0.3278	8.4	No
Darkness of Specks	64.1	63.8	56.4	44.6	78.4	6.5	<.0001	0---100	0.0%	0.7	0.8600	11.6	No
Brown Colour	3.3	1.9	4.2	0.9	8.5	4.7	0.0254	0---100	74.4%	2.7	0.0011	4.3	No
Surface Cracks	14.6	14.3	12.0	9.7	17.3	5.7	0.1020	0---100	18.9%	1.3	0.2381	7.5	No
Shininess	41.4	30.4	35.0	27.2	29.3	7.3	0.0029	0---100	0.0%	1.0	0.4455	10.5	No
Internal White Colour	41.4	43.6	37.1	26.2	23.8	6.3	<.0001	0---100	0.0%	1.5	0.0877	7.5	No

Figure 4.4 Extract from overall summary table (SenPAQ).

Distribution graphs, such as that shown in Figure 4.3, can highlight attributes that are bimodal in their scores or where panellists are using large ranges. It is worth checking how many attributes have very low scores, or where panellists have used many zeros, as this information will be valuable for the later analyses.

Figure 4.4 gives some useful overall measurements of the panel's performance. The means and the least significant difference (LSD) for each sample allow a quick assessment of how the samples were discriminated; the low scores column gives the percentage of scores under 5% of the scale value; and the root mean square error (RMSE) for the panel as a whole is a useful attribute-linked measure of repeatability: the lower the RMSE value, the better (see later for more information).

Checking the output from ANOVA can also be useful at this stage as it will give an idea of the number of attributes showing significant differences between the samples and also the amount of error involved in the assessment of each attribute. It may also be useful to look at closely related attributes to establish if each panellist uses the attributes consistently. For example, if there are odour, flavour and aftertaste attributes in the profile, one would assume that generally the samples would be differentiated in a similar manner across each modality. If there are major differences between odour and flavour particularly, this may be a concern.

Once the general quality of the data as a whole has been evaluated, specific measures of panel performance can be investigated. Some sensory analysts prefer to look at these measures from the overall panel point of view first and then highlight issues with individual panellists, while others may start with individual panellists before looking at overall panel measures. Both approaches are valid, as long as a systematic procedure is followed.

4.2.6.2 Evaluate Repeatability

The next step in looking at panel performance is to check the replication for each panellist for each attribute. Generally, in sensory profiling there are two to three replicates; in this example, there are two replicates. For a well-trained panel which works on a number of different product types, a general rule of thumb for individual panellist replication on a 100-point scale is less than 20 points difference between the replicates can be regarded as 'acceptable' replication and less than 10 as 'good' replication. Better replication may be expected in cases where the panel works on very few product types, or when using the Spectrum type approach, or when the aim of the evaluation is to pick up very subtle differences between samples.

Figure 4.5 gives an example of a simple graphical illustration of panellist repeatability within a project for the same data as given in Table 4.2. If the action standard for repeatability for these panellists is set at 20 points on the 100-point scale, only panellists 1, 7, 8 and 2 meet the criteria for sample A (panellist 2 may meet the criteria for repeatability but scoring a sample as not at all crispy in one replicate can cause issues in the data, such as deflating the mean), whereas for sample D, all panellists apart from panellist 6 meet the criteria. Panellists 1, 3 and 4 are more repeatable than panellists 2, 6 and 9 for sample D. Although panellist 8 shows no differences in their replicates (excellent replication) for both samples, this panellist appears to be showing low discrimination, which would be picked up on later when investigating that performance measure. Scrolling through these types of graphs can give a good overview of panellist repeatability and as these graphs are also easily understood by panellists themselves, they are useful for panel training and feedback sessions.

Training, possibly with spiked samples, and feedback can help panellists to improve their repeatability. Review of their own graphs, as well as sharing examples of anonymized 'good' and 'bad' repeatability, can also benefit any panellists with issues. However, if there are repeatability problems for a well-trained panellist for just a handful of attributes, one cause of this may be lack of sensitivity for these stimuli. This is a difficult area to deal with as not everyone can be equally

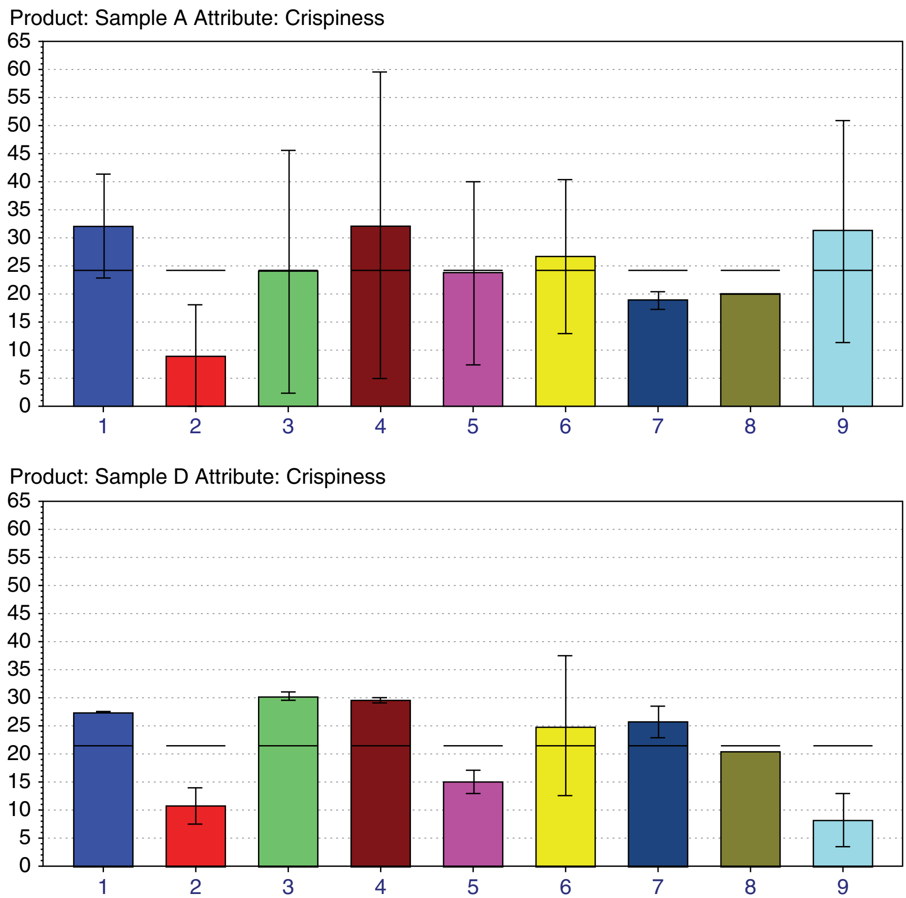


Figure 4.5 Simple graphical illustration of panellist repeatability within a project; the panel mean is represented by a black line (FIZZ).

sensitive; however, if the analyst has this information they can make decisions about data treatment from a knowledge-based point of view.

If a replication pattern was found in this analysis, for example if several panellists gave similar poor replication for the same sample(s), it may be worth checking the experimental design for errors or the preparation log for any sample preparation/presentation issues. In this example, there do not appear to be any consistent errors and those panellists who produce poor replicates do so on a number of occasions, indicating that there may be either a training requirement for 'crispiness' or there is some sample variation between replicates. However, the latter (intrasample variability) usually will become apparent during training and familiarization stages, that is, before data collection.

The mean square error (MSE) is calculated during the analysis of variance process, and is the error sum of squares divided by the error degrees of freedom. The MSE is a measure of the average variance in replicates for each panellist across

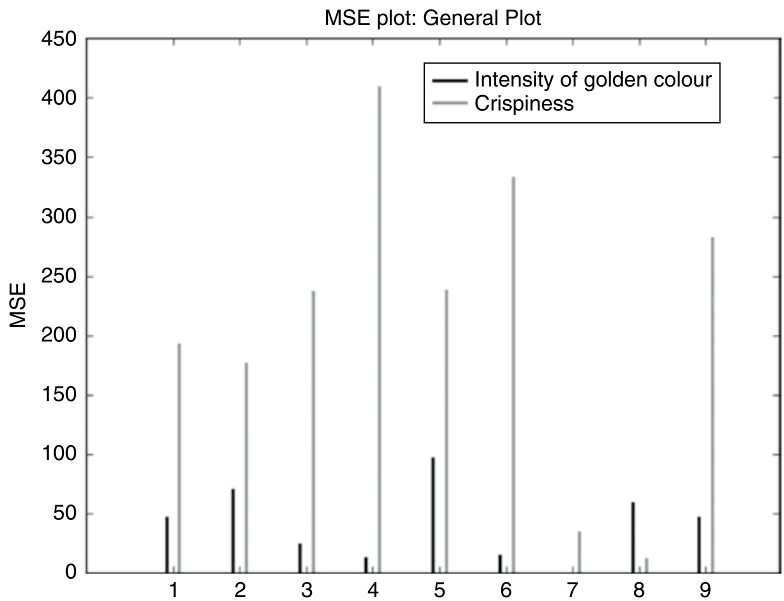


Figure 4.6 Comparison of panellist MSE values for intensity of golden colour and crispiness (PanelCheck).

the samples or, in the case of the panel, the average replicate variance across samples and panellists. The smaller the MSE the better in terms of repeatability. An example of an MSE plot for individual panellists is given in Figure 4.6 (individual panellists are indicated by number on the *x*-axis). This MSE plot summarizes the data for all the samples but the detail about which samples are being replicated poorly is lost in this plot (as samples are not depicted) and so any experimental error (such as sample presentation errors) may also be concealed. MSE values can be limited in use as action standards across projects because the value is not easily interpretable in terms of the measurement scale but they are useful for the comparison of panellists and/or attributes within a single profiling project. For example, Figure 4.6 shows that for the crispiness attribute, panellists 7 and 8 were the most repeatable, whereas panellists 4 and 6 were less so. Figure 4.6 also allows for a direct comparison of the MSEs for both intensity of golden colour and crispiness. Examining the individual panellist MSE values with reference to the raw data table for crispiness (Table 4.2), panellists 1, 2, 3 and 5 have MSEs around 200 and give replicates on average approximately 20 units apart on the 100-point line scale. Panellists 4, 6 and 9 have MSEs of around 300 or more (showing their poorer replication – sometimes as much as 55 points) and panellists 7 and 8 have MSEs of less than 50 (due to their very good replication ranges of no more than 11 points).

The RMSE is the square root of the MSE. It is the error standard deviation from the ANOVA model and has the same unit of measurement as the sensory scale, therefore giving an easy-to-use summary of the panellists' or panel's ability as a whole to replicate. The RMSE for the panel as a whole can also be easily calculated by taking the root of the MSE in an ANOVA table. It can be interpreted relative to the average panel score of an attribute and/or relative to the total measurement scale. For example, an RMSE of 5 for an off-flavour attribute with a mean of 7 shows much more relative variation than, say, the same RMSE for a texture attribute with a mean of 30. The RMSE can be determined for each panellist from an MSE plot or a table of one-way ANOVAs (including MSE values) for each panellist and each attribute.

4.2.6.3 Evaluate Discrimination

The next step for checking project panel performance is whether or not the panellists and panel have found differences between the samples when a 'real' difference exists. In most profiling type approaches, attributes are chosen based on their ability to demonstrate comparative differences between samples within the sample set, so there will be an important number of attributes in each profile for which one would generally expect overall panel discrimination between at least two samples.

One approach is to start by checking the discrimination by the panel for each attribute by looking at the ANOVA (sample F values and probability) and *post hoc* analyses. The overall panel RMSE and the panellist by product interaction value for attributes where no significant difference is found can give an indication if there are problems with repeatability and/or consistency with this attribute affecting discrimination or not. Where these problems exist, the next step will be to look at individual panellist performance for discrimination.

Simple graphs or tables of means will highlight panellists who are unable to discriminate due to poor replication or scale use, that is, too little scale use overall or too much of the scale used for a specific attribute and product. Conducting one-way ANOVA for each panellist and checking F-values or P-values for each attribute by panellist will highlight problems. The extract of the panellist performance table in SenPAQ, as shown in Figure 4.7, allows direct comparison between

Table 2b: p-values for Assessor Discrimination - Significant p-values indicate assessor can discriminate between products	1	2	3	4	5	6	7	8	9
Intensity of Golden Colour	**0.0164**	**0.0370**	**0.0019**	**0.0006**	**0.0135**	**0.0015**	**<.0001**	**0.0176**	**0.0123**
Number of Dark Edges	0.0733	**0.0003**	**0.0001**	0.0790	**0.0176**	**0.0135**	**0.0043**	**0.0201**	**0.0056**
Darkness of Edges	0.2458	0.4898	0.0707	0.9054	**0.0129**	**0.0464**	0.0993	**0.0133**	**0.0270**
Number of Specks	**0.0322**	0.0627	0.4033	**0.0212**	0.4100	0.0545	**0.0337**	**0.0019**	**0.0090**
Darkness of Specks	0.1707	0.1819	**0.0135**	0.2079	0.1282	0.2766	0.2389	0.2536	**0.0391**
Brown Colour	**0.0400**	**0.0457**	**0.0002**	0.1087	0.4291	0.5982	0.1261	0.4857	0.7392
Surface Cracks	0.4093	0.9401	0.6736	0.1435	0.3375	0.3951	0.0575	0.4745	0.1127
Shininess	0.2381	0.3491	0.2390	0.3129	0.3203	0.3260	0.5815	0.0822	0.4804
Internal White Colour	0.1311	**0.0013**	0.1174	0.1684	**0.0307**	0.0648	0.1702	0.0588	**0.0351**

Figure 4.7 Extract from panellist performance table (SenPAQ).

panellists, and the number of panellists contributing to significant overall panel discrimination can be checked. A useful action standard here is that panellists should discriminate at $p<0.10$ when the panel discriminates at $p<0.05$.

Table 4.3 gives the raw data for intensity of golden colour and Figure 4.8 gives the ANOVA output for the same attribute. Figure 4.9 gives the *post hoc* analysis showing the differences between the samples and Figure 4.10 gives the P-values for assessor discrimination. Significant P-values show that the panellist can discriminate between products for this attribute; in this example all panellists can discriminate. Figure 4.11 is the profile (or interaction) plot for intensity of golden colour which shows how the panellists compare in their average scoring of the samples.

The good discrimination performance for this attribute is probably in part related to better repeatability. Only one of the panellists, panellist 5, has any replication ranges of greater than 20 points (see Table 4.3); the rest of the panellist replication performance could be classified under 'good' (less than 20 points difference between the replicates) and several as 'excellent' replication (less than 10 points difference). See also the comparatively lower RMSE and MSE values for this attribute in Figures 4.4 and 4.6 respectively.

4.2.6.4 Evaluate Consistency

Another step in the panel performance project checks is to ascertain whether the panellists agree on the sample differences and rank order, where differentiation in samples is found. To exemplify this step, the attribute intensity of golden colour will be compared to another attribute, darkness of edges.

The ANOVA of the data from all panellists for intensity of golden colour indicated that there were differences between the products, and the *post hoc* comparisons (see Figure 4.9) show that sample B was found to be more golden in colour than all the other samples, sample E being the least golden overall. Looking at the mean scores for these two samples (see Table 4.3) for each panellist, we can see that all panellists have shown the same result as the panel as a whole. Looking more closely, we can also see that all panellists were in agreement about the slight difference between samples B and C (B slightly more golden in colour than C) as well as the larger difference between samples B and E. The lack of significant panellist by sample interaction within the ANOVA for golden colour (see Figure 4.8) confirms this good overall panel consistency for this attribute.

Figure 4.12 shows the ANOVA of the data from all panellists for the darkness of edges attribute, which indicates that there were statistically significant differences between the products, but also that there is a significant panellist*sample interaction (P-value of 0.0064). This interaction needs to be examined before any conclusions about sample differences can be reported.

Figure 4.13 shows the sample differences from the *post hoc* analysis and Figure 4.14 gives the profile or interaction plot for the darkness of edges attribute. Figure 4.14 shows that panellists 1, 2, 3 and 7 disagree about the darkness of edges

Table 4.3 Raw data for intensity of golden colour attribute (FIZZ).

Sample A				Sample B			Sample C			Sample D			Sample E		
Panellist	**Rep.1**	**Rep.2**	**Mean**	**Rep.1**	**Rep.2**	**Mean**	**Rep.1**	**Rep.2**	**Mean**	**Rep.1**	**Rep.2**	**Mean**	**Rep.1**	**Rep.2**	**Mean**
1	39	39	39	68	58	63	50	69	59	37	38	38	30	30	30
2	50	33	41	63	50	56	51	36	43	25	26	25	20	21	20
3	50	61	55	71	73	72	60	70	65	41	40	40	32	31	31
4	47	40	43	62	61	61	50	42	46	31	30	30	22	21	21
5	45	49	47	79	88	84	59	85	72	47	49	48	22	36	29
6	40	41	40	70	70	70	50	50	50	40	41	40	42	30	36
7	40	41	41	71	71	71	61	60	60	50	50	50	40	41	41
8	51	41	46	71	70	70	60	70	65	51	50	51	21	40	30
9	50	31	41	70	71	70	60	60	60	50	40	45	32	32	32
Mean	46	42	44	69	68	68	55	60	58	41	40	41	29	31	30

Intensity of Golden Colour

Source	DF	Type I SS	Mean Sq	F-Value	Pr>F
Assessor	8	3065	383.1	7.24	<.0001
Sample	4	16438	4109.5	77.67	<.0001
Assessor*Sample	32	1693	52.9	1.30	0.2079
Error	45	1835	40.8		
Total	89	23031			

Figure 4.8 ANOVA output for intensity of golden colour (SenPAQ).

Attribute posthoc Intensity of Golden Colour	**Sample B**	**Sample C**	**Sample A**	**Sample D**	**Sample E**
	68.4 a	57.8 b	43.6 c	40.8 c	29.9 d

Figure 4.9 Intensity of golden colour *post hoc* analysis. Samples with the same letters were not found to be statistically significantly different at $P < 0.05$.

Table 2b: p-values for Assessor Discrimination- Significant p-values indicate assessor can discriminate between products	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Intensity of Golden Colour	0.0164	0.0370	0.0019	0.0006	0.0135	0.0015	<.0001	0.0176	0.0123

Figure 4.10 P-values for panellist discrimination (SenPAQ).

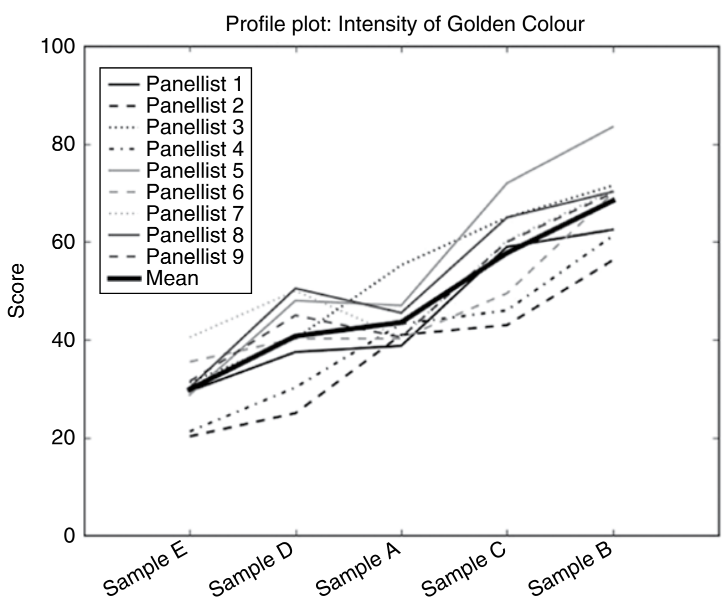


Figure 4.11 Profile plot for intensity of golden colour (PanelCheck).

Darkness of Edges

Source	DF	Type I SS	Mean Sq	F-Value	Pr>F
Assessor	8	3964	495.6	1.54	0.1823
Sample	4	6594	1648.4	5.13	0.0026
Assessor*Sample	32	10290	321.6	2.24	0.0064
Error	45	6469	143.8		
Total	89	27318			

Figure 4.12 ANOVA output for darkness of edges (SenPAQ).

Sample B	69.8	a
Sample C	61.4	ab
Sample E	49.5	bc
Sample D	49.2	c
Sample A	48.4	c

Figure 4.13 *Post hoc* analysis for darkness of edges (SenPAQ).

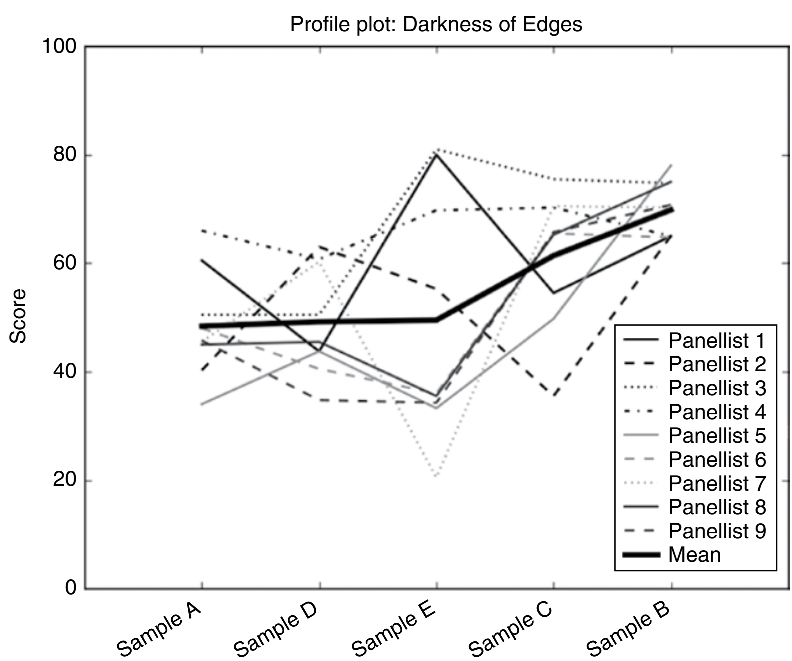


Figure 4.14 Profile or interaction plot for darkness of edges (PanelCheck).

for sample E in particular; the panellists are not agreeing on the sample order for this attribute. The *post hoc* analysis for the whole panel suggests that the samples are in the order B (most dark edges), followed by C, with E, D and A having the least dark edges, but most panellists give a sample order different to the average panel result.

Table 5b: p-values for Assessor contribution to the interaction. Significant p-values indicate poor consensus	1	2	3	4	5	6	7	8	9	*Interaction p-value*
Intensity of Golden Colour	0.7162	0.2131	0.2773	0.3632	0.0502	0.2220	0.1619	0.4622	0.7883	0.2079
Number of Dark Edges	0.6552	**0.0331**	0.0890	0.4900	0.1626	0.2403	0.1070	0.0977	0.0827	**0.0325**
Darkness of Edges	**0.0084**	**0.0163**	0.0619	0.2213	0.2153	0.5535	**0.0091**	0.4549	0.3006	**0.0064**
Number of Specks	0.7939	0.0833	0.1389	0.5904	0.5850	0.9750	0.4173	0.1978	0.1580	0.3278
Darkness of Specks	0.9299	0.6794	0.0844	0.7134	0.9040	0.4339	0.8032	0.5861	0.7217	0.8600
Brown Colour	0.2033	0.3761	**0.0057**	**<.0001**	0.9166	0.1091	0.3740	0.9041	0.2049	**0.0011**
Surface Cracks	0.8963	0.6819	0.5200	0.4223	**0.0054**	0.4509	0.8372	0.3312	0.1434	0.2381
Shininess	0.6137	0.4991	**0.0372**	0.1379	0.1977	0.9897	0.5535	0.7765	0.8632	0.4455
Internal White Colour	0.4204	0.3778	0.5580	0.2296	0.1571	0.1403	0.4781	**0.0349**	0.0911	0.0877

Figure 4.15 Extract of panellist contribution to overall interaction table (SenPAQ).

A summary table of panellist contribution to overall panel interaction confirms (as shown in the extract in Figure 4.15) the information in the profile plot. The panellists with significant contributions to the interaction can be seen: 1 (P-value 0.0084), 2 (0.0163) and 7 (0.0091). Panellist 3 is also showing a significant contribution interaction effect if the cut-off is set at 10%.

Given the nature of the attribute, and the fact that several panellists appear to have the same sample order (distinguishing samples B and C as having the most dark edges and A, D and E as having less dark edges), this interaction could be due to sample variability. The causes of the interaction should be explored before deciding how to report results.

In contrast, the interaction plot for intensity of golden colour (see Figure 4.11) shows very good agreement between the panellists with only slight differences, mainly around sample A. The lack of any panellist contributing significantly to interaction for intensity of golden colour can also be seen in Figure 4.15.

4.2.6.5 Evaluate Overall Performance and Validity

The above evaluation has shown that the overall panel and, in some cases, individual panellist performance for the intensity of golden colour attribute is considerably better in terms of replication, consistency and sample discrimination than the darkness of edges attribute, illustrating how the same panellists can show different competencies for different attributes within the same profile. This may be due to performance issues such as differences in ability to assess modalities, or understanding of, or sensitivity to, attributes. But it is important to be aware that differences in data quality may also be due to sample differences, or preparation and experimental design issues.

Some modalities, such as aroma, are known to be more difficult to rate and to demonstrate excellent panel performance, whereas others, such as appearance, can be easier, as panellists can more easily express their perception and share their thoughts with other panellists in a visual manner. For example, if there is a difference in colour, all samples can be presented in the booths to enable the panellists to informally rank them and then discuss their results. Also some attributes within a particular project can be easier for panellists to score as the definitions and anchors are well defined or have been used in previous projects.

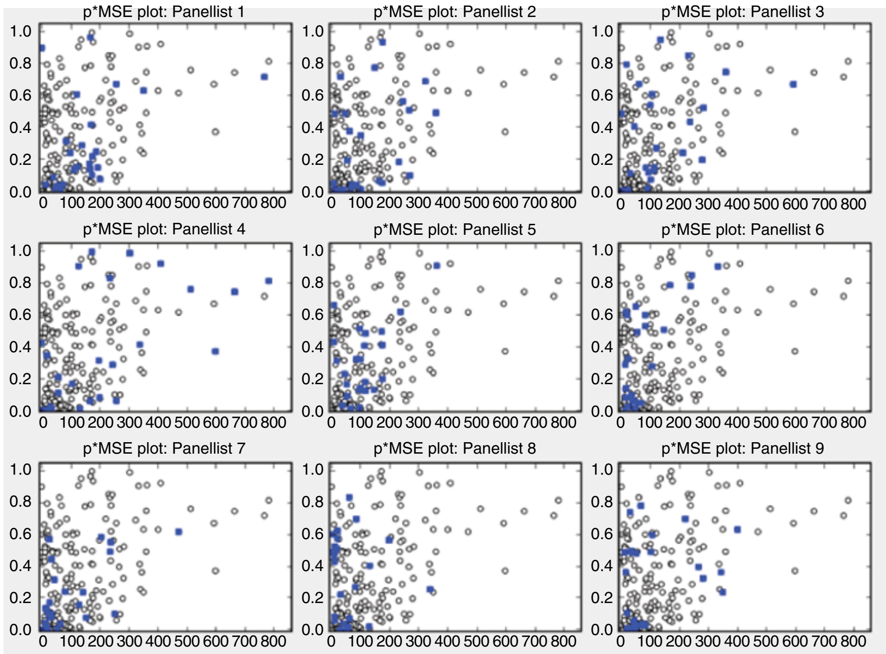
If samples are very different for a particular attribute, this can result in better panel performance as it can be easier for the panellists to grasp the differences and more easily express the sensory space. If there are several samples and a great number of attributes (more than 50), sensory adaptation or fatigue can be the cause of poor panel performance. If there are issues with sample preparation (for example, where the cooking produces non-uniform replicates) or batch-to-batch variation, these can also result in what appears to be poor panel performance; however, most of these types of errors would be highlighted during preliminary evaluations such as attribute generation sessions in QDA-type profiling methods.

Visualizing several elements in one graphic can help to give an overall summary of panel performance. One such aid in assessing panel performance is a p-MSE plot. P-values and MSE values are measures of discrimination and repeatability, respectively. Figure 4.16 shows a p-MSE plot for all panellists and all attributes, with each panellist highlighted as shaded squares. The closer a point is to the origin of the plot (i.e. low values for both p and MSE), the better the panellist has performed in these two aspects. The plot is split into four sections, based on repeatability and discrimination. The transparent circles represent all the other panellists for all the other attributes. This allows the analyst to quickly ascertain if there have been any problems with poor repeatability or discrimination within the profile overall, as well as the identification of problem attributes and panellists. In general, performance is better with respect to repeatability for this profile, as compared to discrimination.

The lower part of Figure 4.16 shows the same p-MSE plot highlighted and annotated for panellist 4. Panellist 4 showed poor discrimination between samples for crispiness but good repeatability and discrimination for intensity of golden colour.

p-MSE plots can also be created for individual attributes. The two plots in Figure 4.17 demonstrate the differences seen earlier in the two attributes intensity of golden colour and crispiness. In the intensity of golden colour plot, all panellists are very closely grouped in the 'good discrimination and good repeatability' quadrant. In contrast, for crispiness, panellists are situated further from the plot origin and five panellists are located in the 'poor discrimination between samples, good repeatability' quadrant. The results for individual panellists can be examined within these plots, and the location of data points for panellist 5 is indicated.

Other forms of overall visual assessment of performance allow for exploring repeatability and/or discrimination in terms of sample average. The 'judge performance graph' in Figure 4.18 gives a comparison of panellist performance



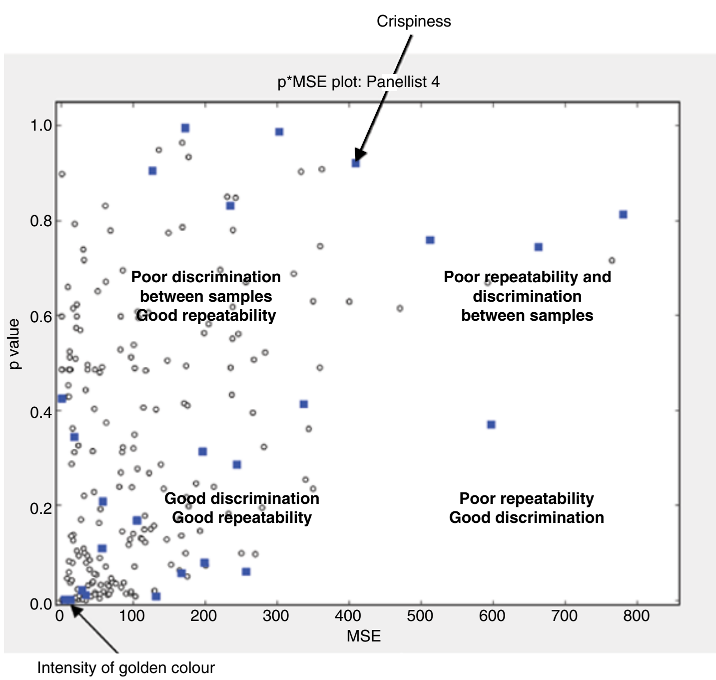
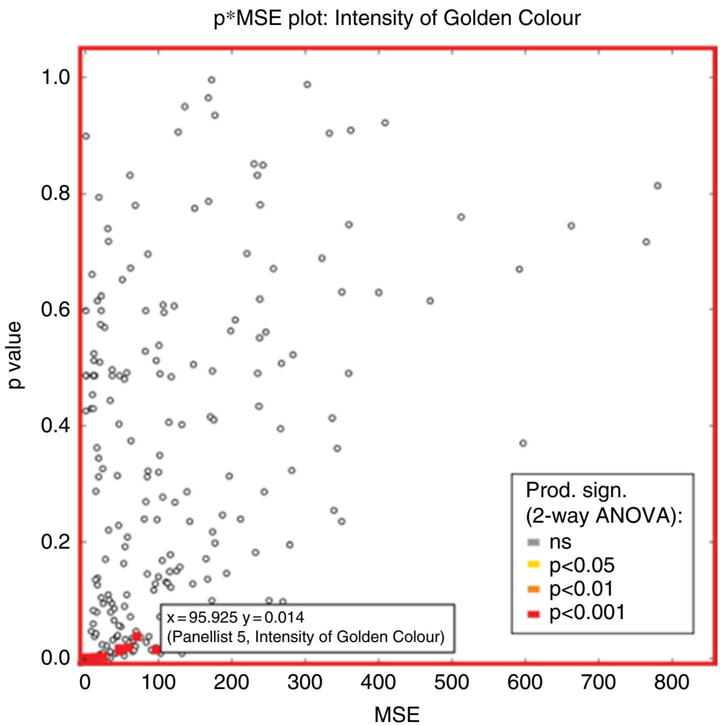


Figure 4.16 p-MSE plots by panellist for all attributes with selected attributes highlighted (PanelCheck).



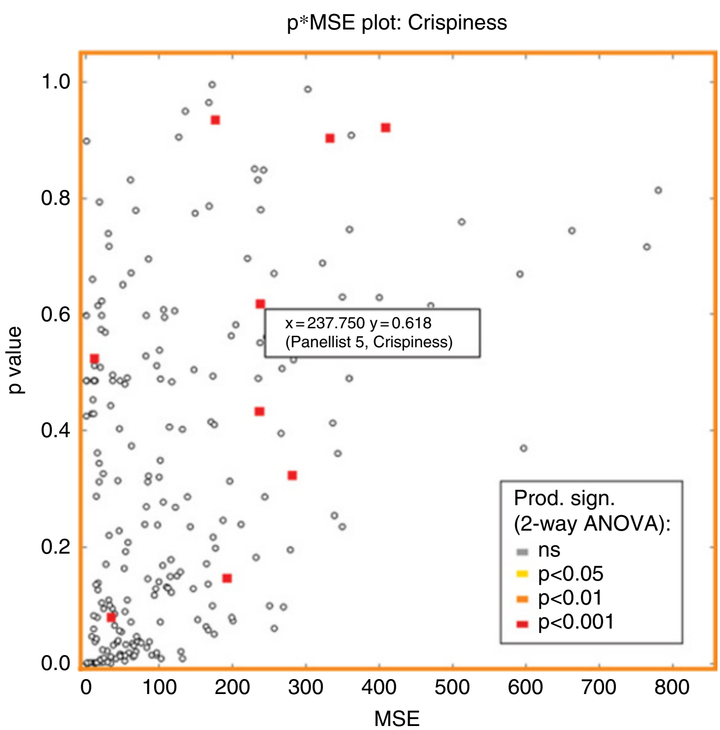
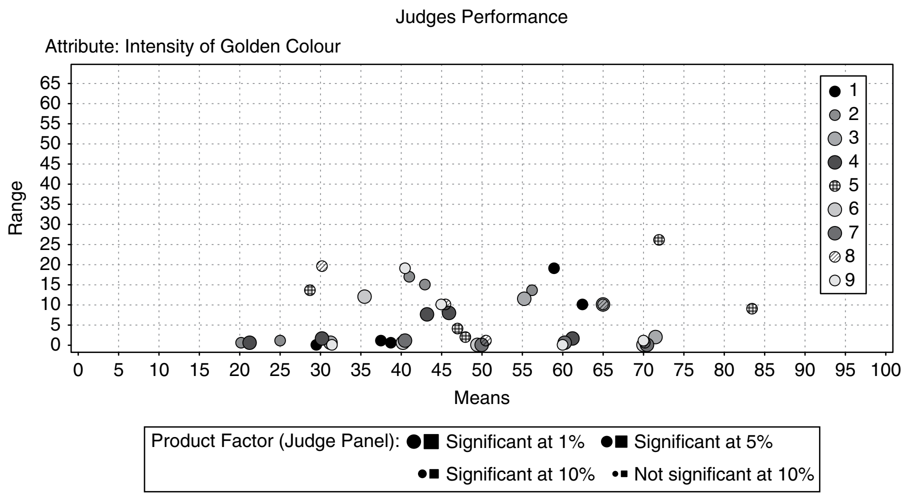


Figure 4.17 p-MSE plots for intensity of golden colour and crispiness with panellist 5 results highlighted (PanelCheck).



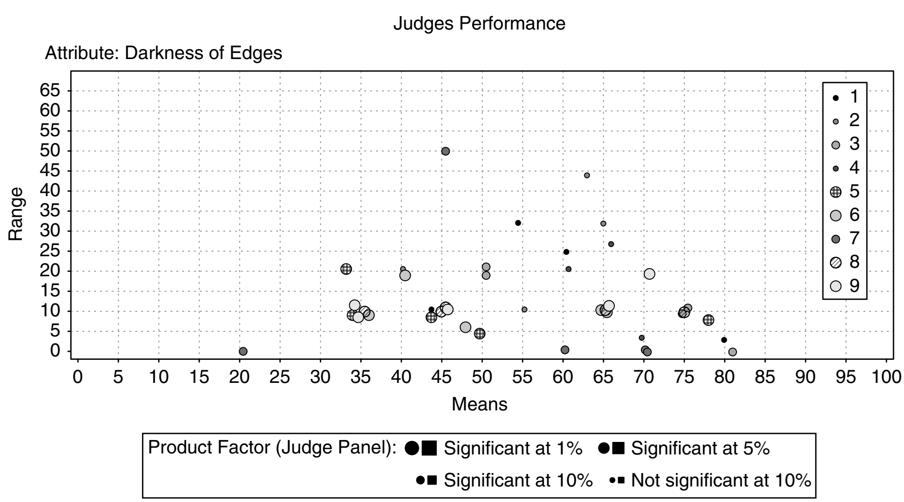


Figure 4.18 Judge performance graph for all panellists for intensity of golden colour and darkness of edges attributes (FIZZ).

for intensity of golden colour and darkness of edges to show this concept in action. Each circle on the graph represents a panellist by product combination, and when using FIZZ, one can highlight each point to identify that panellist and sample. Key performance points noted from these graphs include the following.

- All panellists are discriminating (at $P<0.05$) across the samples for intensity of golden colour but some are not for darkness of edges (circle size).
- The replication range for each panellist by sample combination is under 30 for intensity of golden colour, but for darkness of edges several panellists have sample replication ranges over 30 (placement of circle on *y*-axis).

- Most panellists appear to be using a good portion of the scale across samples for both attributes (placement of circles of same colour along x-axis; this is easier to see when the graphs are displayed in colour).

Other methods of evaluating overall performance compare each panellist to the panel overall or the panel average for major performance measures. The summary table and graph in Figure 4.19 give a quick overview of the relative performance (for all attributes in one profile) for the nine panellists participating in the profile from the point of view of discrimination, repeatability and consistency. For example, panellist 2 is 83% better at discriminating than the average panellist, and panellist 4 is 15% worse at achieving consistency than the average panellist.

Scatter plots are another method of quickly viewing panellists' performance and a combined example is shown in Figure 4.20. Each panellist's score for each replicate is shown and it is easy to see how the attributes compare in terms of replication, intensity of golden colour obviously indicating tighter replication. The plot also shows how the samples compare for each attribute and the range of scores for each sample and each attribute. It's also possible to see how far from the panel average each panellist is for each sample and each attribute.

Performance relative to panel average (positive is good)	1	2	3	4	5	6	7	8	9
Discrimination	–8.5%	83.1%	6.8%	–23.7%	–39.0%	–23.7%	6.8%	–23.7%	–22.0%
Repeatability	1.8%	4.1%	–0.5%	–23.7%	1.8%	8.7%	6.4%	1.8%	–0.5%
Consistency	1.0%	–3.7%	–1.3%	–15.4%	1.0%	8.1%	1.0%	5.7%	3.4%

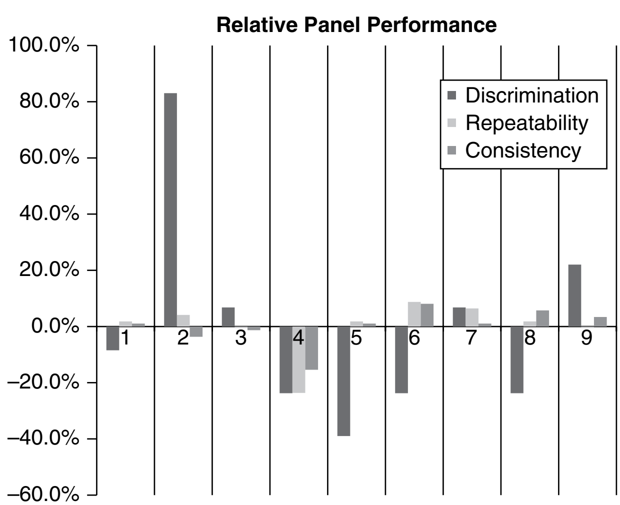


Figure 4.19 Panellist performance summary table and graph (SenPAQ).

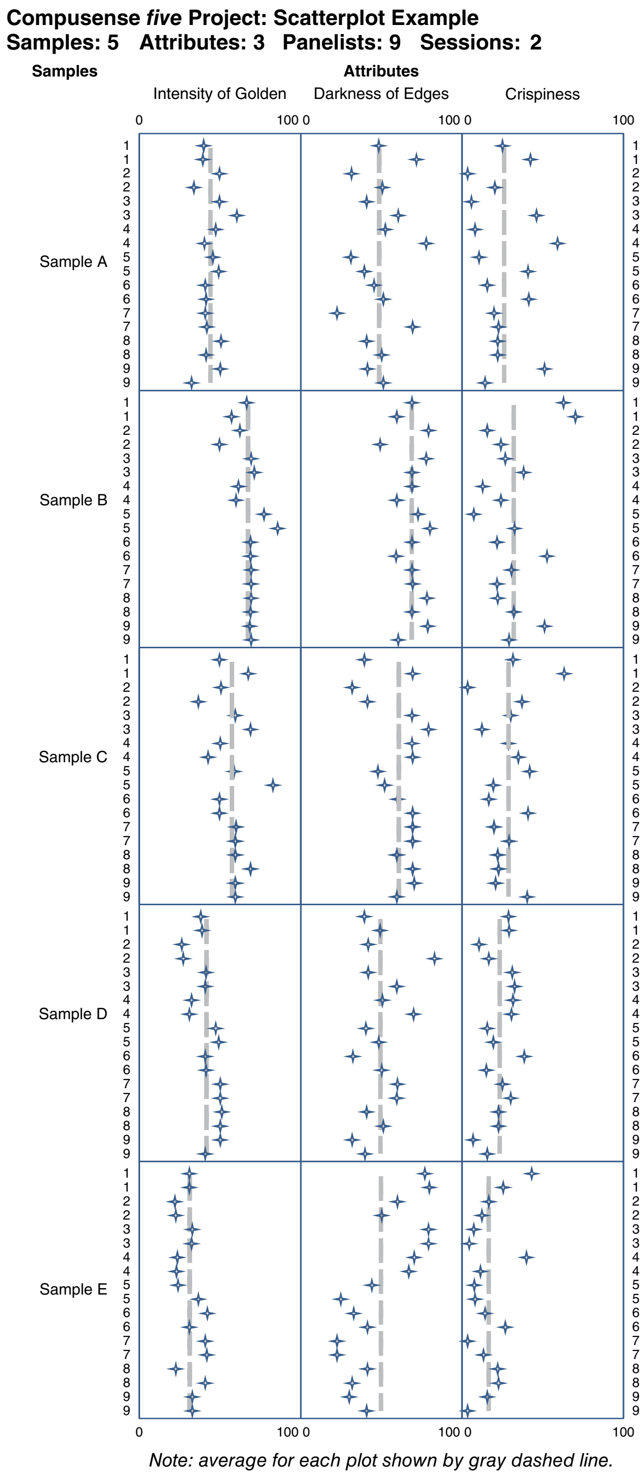


Figure 4.20 Scatterplot example (Compusense).

Issues related to the human aspects of panellists, such as motivation, behaviour, illness, personal matters or differences in perception, can also affect panel performance. These issues should be picked up during the scoring sessions, as the panellist's data will not match their earlier standards. However, it is better to have a protocol whereby all panellists must report any issues to the panel leader prior to any sensory sessions. If the panellist is causing severe problems in one project, it may be advisable to delete their complete data set on this occasion but if this is an ongoing issue, further steps will be needed to resolve the issues.

Once performance measurements have been carried out, the overall validity of a profile can be evaluated in terms of the usefulness of its output and applicability to action standards. If the data generated are not robust enough to use in the decision-making process, further training or data may be required. In some cases, the profile may need to be repeated or different sensory tests altogether may need to be conducted.

4.2.7 Dos and Don'ts

✓ Evaluate general data quality as a first step.
✓ Evaluate repeatability, discrimination and consistency for the whole panel and individual panellists.
✓ Set performance action standards (see section 4.3.2 for more information on action standards) and evaluate data against these.
✓ Give regular, considerate and timely feedback to the panellists; collecting performance data alone will not help improve the panel. Remember that panellists are human beings and deserve respect.
✓ Look at all different panel performance analysis options available and choose the one that suits the context and the panellists.
✓ Consider the overall validity of panel output with respect to project action standards and put corrective actions into place where there are problems.
× Don't report the results from the statistical analysis of data without first checking panel performance.

4.3 Panel Monitoring

4.3.1 Introduction

As explained in the introduction to this chapter, panel monitoring is the ongoing data assessment and validation, feedback and refresher training associated with individual panellist performance and overall panel performance. Panel monitoring focuses on uncovering, understanding and acting on panellist and panel performance issues over time.

Generally, panel monitoring uses specific panel performance measures to build up a picture of individual panellists' performance and overall panel performance. There are two structural approaches for achieving this.

- *Ongoing project-based monitoring*: performance measures from each descriptive panel project can be collected, evaluated, trended and monitored against targets over time.
- *Scheduled diagnostic checks*: specific monitoring tests can be developed and carried out at regular time intervals. The results are compared to targets, and trends are examined over time.

It is possible and advisable to use both of these monitoring approaches; the first will help to identify problems and address these as they arise, understand specific attribute and product-related issues, and provide outcome-based information for panellist performance reviews. The second approach can result in like-for-like comparisons of panel and panellist performance at different time points and allows for an accurate understanding of underlying performance trends. Specific monitoring tests can also be used to compare whole panel or even individual panellist performance between panels (a type of reproducibility check or proficiency testing). Both approaches require data management and evaluation strategies, relevant action standards and designed feedback loops.

All of the above assumes that the descriptive panellists and the panel have been screened and trained in an appropriate manner for the role at hand. Screening, in most cases, means regular screening for colour blindness, basic taste acuity and thresholds, odour acuity and thresholds, textural acuities and descriptive and communication abilities. More detail on appropriate screening techniques can be found in ISO 8586 (ISO 2014).

In addition, a descriptive panel should also have regular refresher training on key principles of objective assessment and for specific recognition of sensory attributes and defects in the product sector of interest. For some types of descriptive analysis, regular training may include ‘recalibration’ on universal or absolute scales.

4.3.2 Data Collection, Management and Action Standards

In all types of panel monitoring, the choice of type and frequency of monitoring measurements and evaluation will be context dependent. It is vital to collect information that gives a full picture of panellist and panel performance over time. But there is always a trade-off to be considered between the amount of information collected and analysed versus the time necessary to maintain and evaluate these records. The most sustainable systems will either pull out key and simple measures or make use of automated data collection and organization to create a clear picture from a large amount of complex information. The databases generated may be simple spreadsheets or detailed data management systems.

The overall goal is to build up and maintain a body of information that can be interrogated for a variety of purposes, including:

- to evaluate performance trends over time
- to monitor panel or panellist performance at specific time points or with respect to specific project or product types

- to choose the best panel or panellists to carry out relevant projects
- to decide when retraining should take place and in which areas and if there are specific panellists requiring adapted training
- to evaluate the effectiveness of different types of attributes with respect to different types of product evaluation or project objectives
- to inform and guide performance feedback to the panel as a whole and also for specific panellist performance reviews
- to use as an evidence base for discontinuing panellist employment/participation in a panel when necessary.

Ideally, panel monitoring systems will collect performance measures and convert these to tabular or graphical summaries and outputs that can be used for analysis and decision making.

To design and use panel monitoring outputs, it is necessary to consider and set action standards. Action standards will help to define when actions, such as retraining, should be taken. They will be necessary to define upper and lower limits for any statistical process control type procedures and graphics for visualization.

Along with action standards, a panel monitoring system will include feedback loops for the panel leader and the panellists. These are overall schema that define what sort of actions will be taken and when, based on panel monitoring outputs. Actions will be specific training sessions, trouble-shooting investigations, etc. The feedback loops also provide a mode of relaying information about the nature and detail of good and bad aspects of overall panel and individual panellist performance to appropriate stakeholders.

4.3.3 Key Measures and Criteria

The overall approach in panel monitoring is to track the sum, average or moving average of key panellist and panel performance criteria over time. The actual measurements taken at each point may be derived from standard panel performance measures. Performance measures are generally determined by attribute. These are then collated and monitored over time. The system should highlight those panellists and attributes for which there is poor performance and uncover the reasons for this. Individual panellist performance can be considered as an absolute or with respect to average panel performance. Some examples of measures that can be used are outlined in the sections below.

4.3.3.1 Repeatability

The proportion of attributes showing replicated judgements for each product above a predefined acceptable range (e.g. 10 units on a 100-point scale for the overall panel mean) can be determined and recorded for each profiling project. Another simpler approach may be to record the proportion of attributes with an RMSE/mean or overall scale ratio above a predefined level.

Qualitative information about which attributes are showing poor repeats can also be collected. Over time, this will generate a frequency list of attributes that the panel or individual panellists have problems scoring in a repeatable manner which can be evaluated by sensory analysts and explored in panel training sessions.

4.3.3.2 Consistency

The proportion of attributes showing significant interaction for each profiling project can be collected and collated. For individuals, additional measures of consistency are the proportion of attributes with:

- a low correlation of the panellist mean to the whole panel mean across samples
- high absolute deviations from the panel mean for any of the attribute by product combinations (outliers).

4.3.3.3 Discrimination

Measuring and monitoring discrimination performance is difficult because in most cases it is not possible to know ahead of time for which attributes the panel should discriminate. For diagnostic checks, samples can be formulated or chosen to show a difference for specific characters/attributes. For ongoing project work, the approach can be to look at attributes overall and assume that at least a certain percentage of them should be discriminatory, or choose key attributes for each profiling project that the panel leader or sensory analyst expect to be discriminatory and base measurements around these. Individual panellists may be expected to discriminate for those attributes for which the whole panel discriminates, albeit at a lower probability level.

Given the above considerations, the proportion of attributes for which the panel or an individual panellist does not discriminate may be recorded at each time point. A moving average or sum of the discrimination proportion can then be determined for all attributes, or only those predefined as key attributes, or any other logical selection. This calculation can be done for the panel as a whole or for individual panellists. Measurements can be charted over time to determine trends and compare to predefined action standards or targets.

Qualitative lists of attributes that the panel or each panellist discriminates or does not discriminate on can be generated from each profiling project. This can build up a picture over time of which attributes the panel understands and can discriminate on and which they cannot discriminate on. Further evaluation may be needed to decide if non-discriminating attributes actually do not vary for the samples being tested, or if the panel does not understand and agree on these attributes enough to allow for discrimination. In the first case, it might be worth considering whether these attributes even need to be measured, whereas in the second, the panel requires further qualitative and quantitative training and calibration with respect to the attribute.

4.3.3.4 Overall Validity

This is likely to be a summary rating given by the sensory analyst as to whether or not the outputs of the profiling project resulted in valid and useful information for decision making and why or why not.

4.3.3.5 Problems

This will again be a summary rating or explanation given by the sensory analyst or panel leader as to what were the major problems or issues (if any) behind poor panel performance for each project.

4.3.4 Design and Logistics of Ongoing Project-Based Monitoring

In panel monitoring involving ongoing project-based monitoring, specific outputs from each profiling project a panel completes can be measured, recorded and evaluated. In general, information should include summary measures of each of the main criteria associated with panel performance: repeatability, consistency, discrimination and overall validity, as above. There should also be measurement and recording of specific panellist, vocabulary, methodology and/or product issues.

A searchable database that produces output on queries can be built up over time. A choice needs to be made about how much information to collect and evaluate. This will depend on how much automation can be built into the system and what time is allowed for data input as well as inspection of outputs. As a minimum, the database should collect overall panel performance measures for key criteria and highlight problem areas. A possible approach is to review overall panel performance by attribute and then carry out further individual panellist evaluations for problem attributes.

Figure 4.21 shows a schematic for an ongoing project-based panel monitoring system. Figure 4.22 is an example of a simple database set up to monitor overall panel performance and highlights specific panel, attribute and product-related issues on an ongoing basis. The database shown in Figure 4.22 is designed to highlight major performance problems quickly and link these with specific attributes or panellists, as well as look at trends over time.

This type of system serves several purposes. Directly after data are collected from each project, panel performance measures carried out for entry into panel monitoring systems can be used to decide if the data are reliable enough for the outputs of those data to be reported, or if reporting should include any qualifications. The database can also be interrogated at regular intervals and trended reports produced to inform panel training and feedback. This regular reporting may also guide method development and optimization.

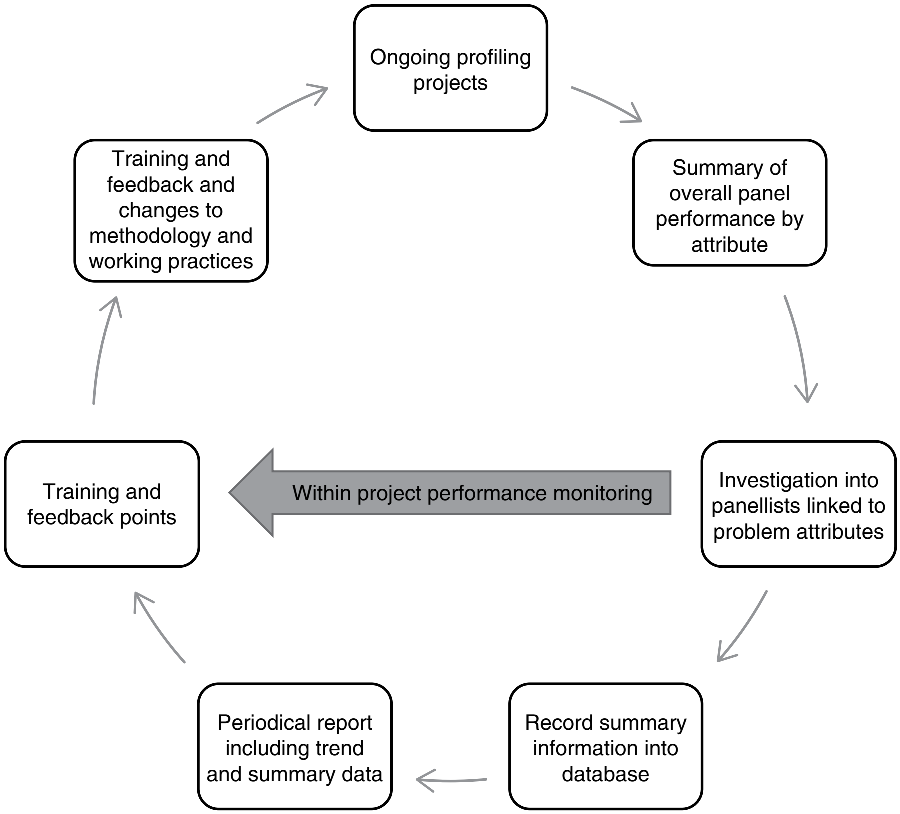


Figure 4.21 Schematic for ongoing project-based panel monitoring system.

4.3.5 Design and Logistics of a Scheduled Diagnostic Checks Approach to Monitoring

There are several ways of designing scheduled diagnostic checks. The first is to have the panel evaluate the same sample set at repeated intervals (e.g. twice a year, quarterly or monthly). This can be a form of proficiency testing and allows for true reproducibility measurement. The problem with this approach is that over time, panellists are likely to recognize the sample set and therefore there will be learning effects confounded with performance trends.

Another approach could be to use a similar type of market mapping sample set for each check point. The scope of the sample set would be similar but will not be exactly the same; for example, for a panel normally evaluating drinks, six orange juice samples, spanning the fresh and UHT market, could be used. In this way, the diagnostic check will focus on whether general performance measures are consistent over time rather than looking at actual reproducibility.

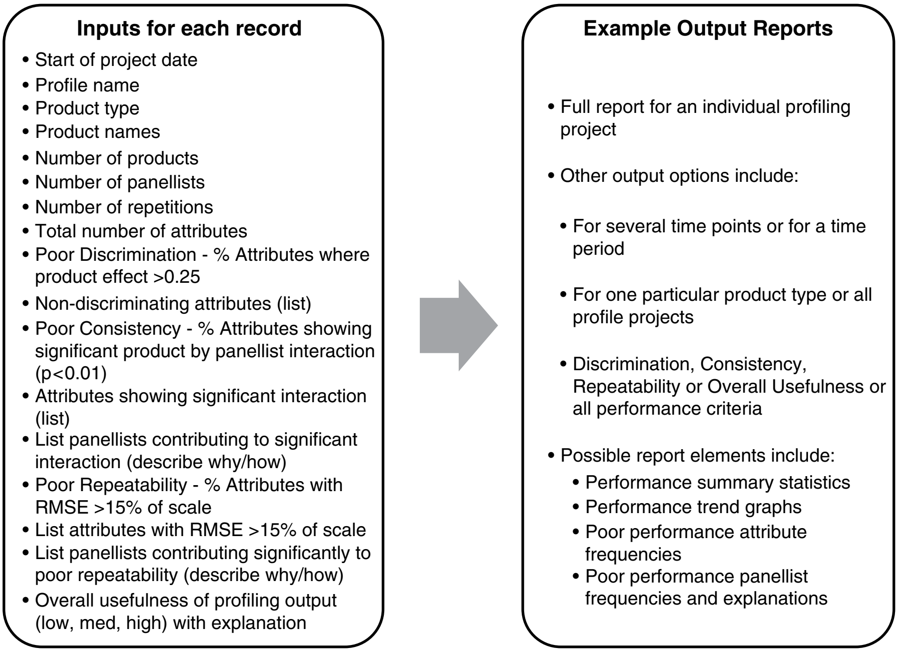


Figure 4.22 Example of simple ongoing project-based panel monitoring database architecture.

The sample set for the scheduled diagnostic check should have at least five or six samples to allow for multivariate statistical evaluations (see section 4.6.2) as well as the standard performance measures. It should show wide enough variation so that discrimination is expected between some samples but not others. Hidden repeat and engineered/designed or spiked samples are also advisable to allow for multiple methods of checking repeatability, discrimination and consistency.

Figure 4.23 shows a schematic for a scheduled diagnostic checks-based panel monitoring system. The key criteria measured and the database approach for scheduled diagnostic checks will be similar to those described above for ongoing project monitoring. Scheduled diagnostic checks are also the ideal way to look at individual panellist performance in detail if there is not time during normal ongoing project work or panel monitoring. An individual database for each panellist can be developed and updated in association with scheduled diagnostic checks. Figure 4.24 shows hypothetical statistical process control charts charting consistency over time and with respect to a target (action standard) associated with scheduled diagnostic checks for two panellists. At each diagnostic check point, the correlation of the panellist mean to the whole panel mean across samples has been calculated for attributes found to be discriminating at panel level. A target (lower control limit, LCL) has been applied so that 70% of these correlation

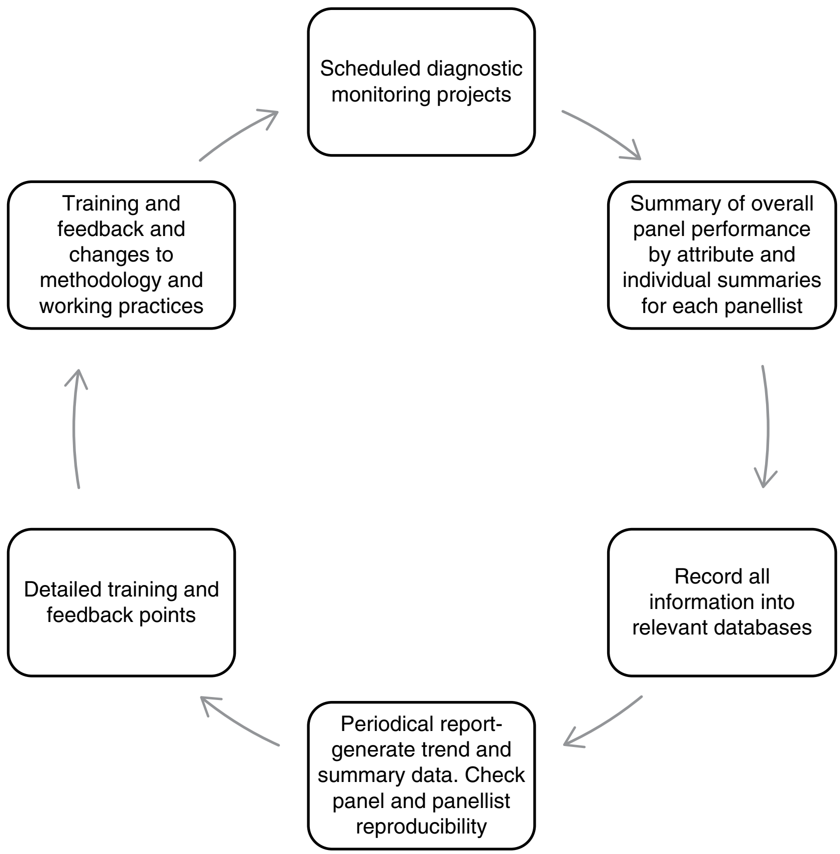


Figure 4.23 Schematic for panel monitoring system based on scheduled diagnostic checks.

coefficients should be greater than 0.7. These example charts clearly show the difference in absolute performance and performance trends over time between the two panellists.

Charts such as those in Figure 4.24 could also be created for repeatability or discrimination.

If vocabulary and products are fixed, reproducibility can be evaluated by attribute via measures such as comparing means and mean ranges, mean square errors and discrimination across diagnostic check points. This can be done for the panel as a whole or individual panellists. It is also possible to carry out a higher level ANOVA that includes panel monitoring sessions, allowing for a comparison across diagnostic check points.

For panels that develop specific vocabulary for each profile, techniques such as generalized Procrustes analysis (GPA) and principal component analysis (PCA) can be used to determine if overall conclusions are similar. Cluster analysis can also be used to check whether product groupings are similar.

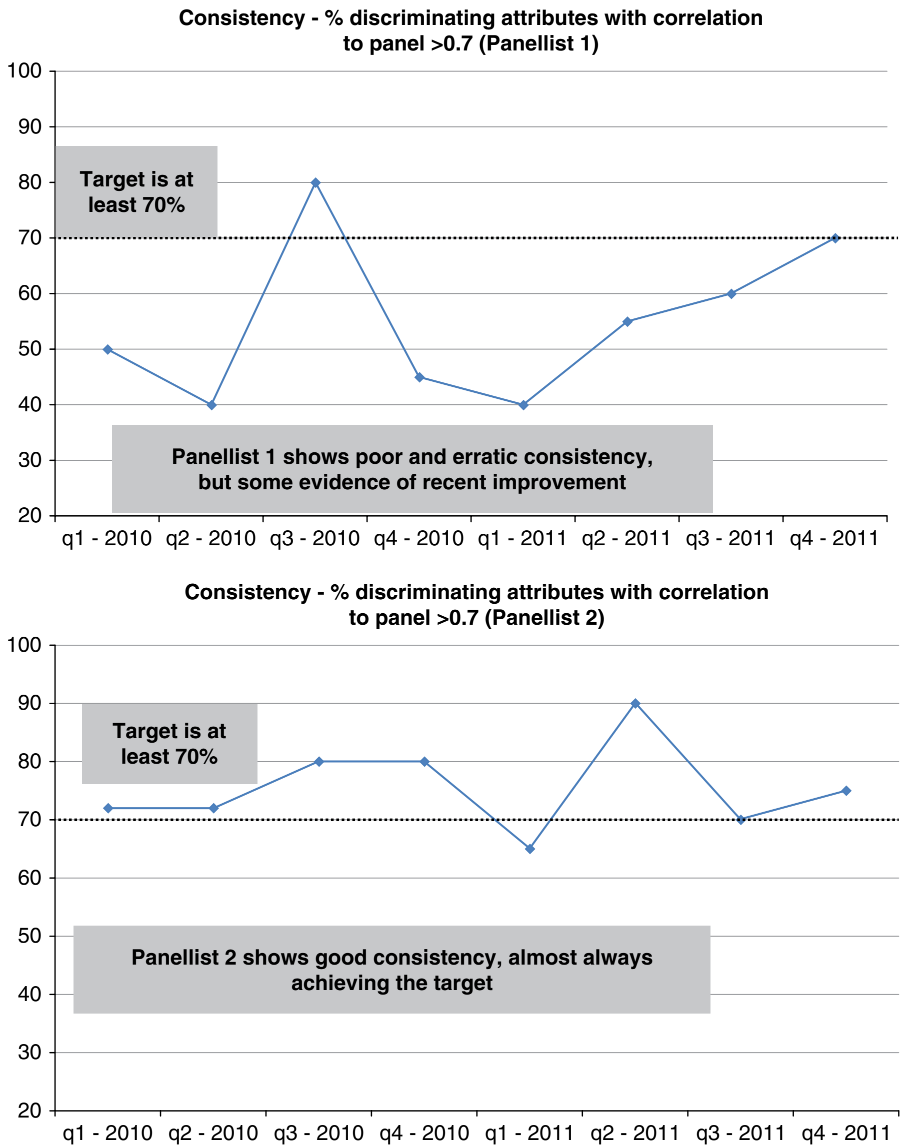


Figure 4.24 Example of hypothetical control charts for panellist consistency as measured in scheduled diagnostic checks.

If the sample set varies with diagnostic check, key criteria can still be evaluated over time as for ongoing project-based panel monitoring. The carefully selected sample set will allow for detailed individual panellist measures to be considered. Reproducibility can be checked in terms of qualitative conclusions about the range and scope of sensory profile variation within the market and similarity of overall performance statistics.

4.3.6 Panel Monitoring and Links to Training and Appraisals

A very important element of ongoing panel monitoring is its link to panellist training and performance appraisal. Collecting objective, clear and trended information allows for sharing of information, identification of key areas for attention and the ability to detect changes over time.

Setting up performance monitoring systems allows objective targets for judging panel and panellist performance. The more information that can be shared between the sensory evaluation team, sensory panellists, human resources, etc., the more robust the system will become and the more likely to result in continuous improvements.

Given the benefits of sharing targets and action standards, there are some elements of panel performance where it is advisable not to reveal too much to panellists. For example, if too much emphasis is put on repeatability, or too much information is given about replicates within presentation designs, panellists may try to find repeat samples and rate them accordingly, rather than rating each sample on its own merits with respect to the overall sample set and relevant attributes and scaling calibration.

The best systems will be two-way and also allow panellists to rate and comment on the most difficult projects, products and attributes; how methodologies and panel quality management systems affect them; and their view on specific panel leader and sensory analyst performance aspects. Involving panellists in training planning and methodology development/optimization where possible is also a way to motivate them.

It is always advisable to present and gain buy-in upfront with all stakeholders on overall strategies for panellist performance and monitoring, as well as which specific actions will be taken with respect to good and bad performance.

4.3.7 Dos and Don'ts

✓ Do set up clear systems for recording, evaluating and acting on whole panel and individual panellist performance over time.

✓ Do consider ongoing and summary assessments of repeatability, discrimination and consistency for both the panel overall and individual panellists within panel monitoring schemes.

✓ Do use ongoing project information as well as scheduled checks to monitor panel and panellist performance over time.

✓ Do provide evidence-based and clear performance feedback to panellists and the whole panel.

✓ Do link training and development initiatives with panel monitoring systems.

✓ Do remember that panellists are people and need clear and sustained support and motivation to perform well over time.

× Don't collect a lot of performance data and then never use it to look at trends over time.

4.4 Proficiency Testing

Laboratory proficiency testing involves analysis of reference samples by various proficiency group 'members'. Resultant data are analysed and compared to expected/true values to establish a laboratory's proficiency for a particular analysis method.

For sensory profiling, proficiency or ring-testing is a general term used to describe external validation which involves comparing results and performance measures from individual panellists or whole panels to reference values to establish proficiency. Proficiency testing for sensory profiling has been informed by proficiency testing schemes for chemical and physical analysis. External validation can be useful in most sensory profiling contexts, but may be particularly important for multi-panel organizations and for commercial sensory testing providers, and is an important step to show that the data provided by a sensory panel are reliable.

Hyldig (2010) described an eight-step approach to proficiency testing.

1. Selection of method
2. Selection and test of material for the proficiency test items
3. Guidelines for preparation of the proficiency test items and the execution of the test
4. Training material and guidelines for training
5. Coding and consignment of the proficiency test items to the participating laboratories
6. Proficiency test
7. Collection of data from the proficiency test round
8. Analysis of data and report of the results

For sensory profiling, choosing an appropriate sample set for testing and deciding what should constitute the reference or standard value can be difficult. The sample set needs to be broad enough to show differences, and ideally will contain spiked or designed samples for which discrimination with respect to certain attributes is expected. A sufficient amount of homogenous, consistent and stable sample material needs to be available to allow for testing by multiple laboratories, possibly at various time intervals. Issues such as batch and preparation variability need to be considered. One of the main difficulties with proficiency testing in sensory profiling is in the setting of a 'true' reference value for each panel to be measured against. Generally, the 'true' sensory rank, score or profile is considered to be the consensus generated from a subset of several panels composed of experienced, screened and trained sensory panellists. Other panels in the proficiency scheme will be judged against this based on their closeness or correlation to these values.

McEwan et al. (2002a) reported that as part of an EU-supported project, ProfiSens, 14 panels undertook ranking of sweetness on five samples of apple juice. Four panels were designated 'validation' panels, whose data were used first to establish the expected ranking results and second to set the performance criteria that a trained sensory panel would be expected to achieve. Four key measures of a panel's performance were investigated: the ability to rank the samples in the correct order; the significance level associated with differences between samples; the number of pairs of samples that a panel found to be different at a specified level of significance; and the degree of agreement between assessors within a panel. For each of these criteria, an 'expected result' was considered, as well as an overall measure of performance. The data from the remaining panels were analysed, and the level of performance was recorded for each of the stated criteria.

McEwan et al. (2002b) trialled a proficiency scheme which was based on descriptive profiling of a common sample set, and comparing each panel's results to 'expected' results from validation panels. The initial establishment and measurement procedure for panel performance included the following steps.

- Step 1 – calculate the significant number of dimensions from GPA.
- Step 2 – calculate which pairs of samples are different on each sensory dimension for each panel, and hence the number of pairs of samples that are significantly different.
- Step 3 – establish how well each panel's sensory map agrees with the 'expected' sensory map.
- Step 4 – calculate how well panellists in each panel agree with each other and the consensus.
- Step 5 – establish the level of performance each panel has achieved.

Proficiency or ring-testing schemes may incorporate simple to more complex measures and methodologies, and focus on individual panellists or the whole panel. For example, AROXA (www.aroxa.com) offers two types of proficiency testing schemes for professional taste panels. The first assesses the ability of tasters to identify flavour attributes presented at low levels in samples. Assessors have to identify the attribute which has been added to each of six samples, choosing from a list of 30–50 possible flavours. The list of flavours presented can be customized to suit the particular customer application. The second type of assessment relates to scaling ability. Assessors are presented with samples which span a range of intensities for a single flavour attribute. Replicates as well as control samples containing no added flavour are included in the test set of nine samples. Assessors have to first rank the samples in order of intensity before rating the intensity of the added flavour in the samples. In both cases, the results can be analysed by comparing output to that from other panels, within the panel or at an individual level.

Laboratory accreditation schemes often require participation in proficiency testing as part of the accreditation process. The guidance document *EA - 4/09 G: 2017 Accreditation for Sensory Testing Laboratories* (European Co-operation for Accreditation 2017) describes reproducibility/repeatability, discrimination of samples, sensitivity and interlaboratory tests as among the performance characteristics that might be relevant to validate sensory methods. It states that laboratories should participate in proficiency testing schemes which are relevant to their scope of accreditation and that in specific instances, participation may be mandatory. The guidance suggests that preference should be given to proficiency testing schemes which use appropriate matrices.

4.5 Further Panel Quality Management Concerns

Panel quality management concepts have implications for, and can be applied within, panel recruitment, training and management activities. Several important areas to consider are discussed within this section.

4.5.1 Comparing and Combining Data from Different Panels

Panel quality management techniques are used to compare performance between panels and may even be used to help decide if data can or should be combined. Typical situations where panel performance may be compared and/or data combined include:

- deciding on the use of experienced versus inexperienced panellists
- using panels from different sites to assess the same samples
- using the same panel to evaluate samples gathered at different times.

The strategy and process for comparing performance will depend on whether or not the same samples have been assessed and the same attributes have been used. There may also be issues in scaling which might need to be taken into account. The more similar the profiling processes to generate data have been, the greater extent to which these data can be compared in terms of panel performance. A summary strategy for comparing panel performance in various contexts is shown in Figure 4.25.

When comparing performance between panels or panellists where different samples have been used, generalized and standardized performance measures, such as repeatability, discrimination of spiked samples and absence of interaction, can be used. When comparing performance of panels or panellists evaluating the same sample set but using different attributes, overall sample structures can be compared using multivariate techniques such as GPA or PCA in addition to generalized performance measures. In addition, it may be possible to compare ranking of samples for common attributes such as the basic tastes. Discrimination patterns/clustering across samples for all attributes, or a subselection of those that are statistically significant, can also be evaluated. If the same sample set has been evaluated using the same descriptive vocabulary and scaling, detailed comparisons of means and differences between samples by attributes can be made.

Figures 4.26 and 4.27 show the first two dimensions of a GPA output comparing two panels which have assessed the same samples using the same

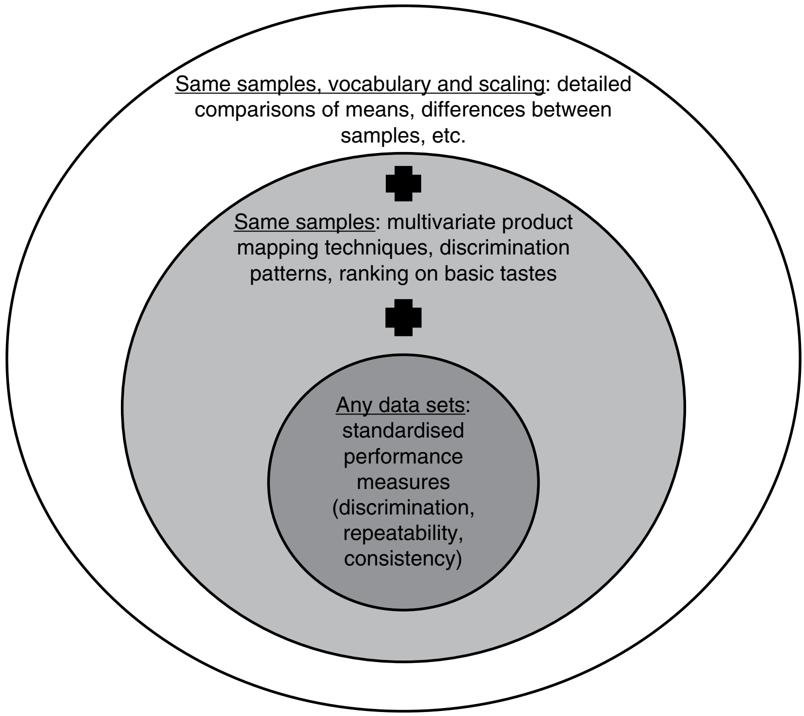


Figure 4.25 Strategy for comparing panel performance in various contexts.

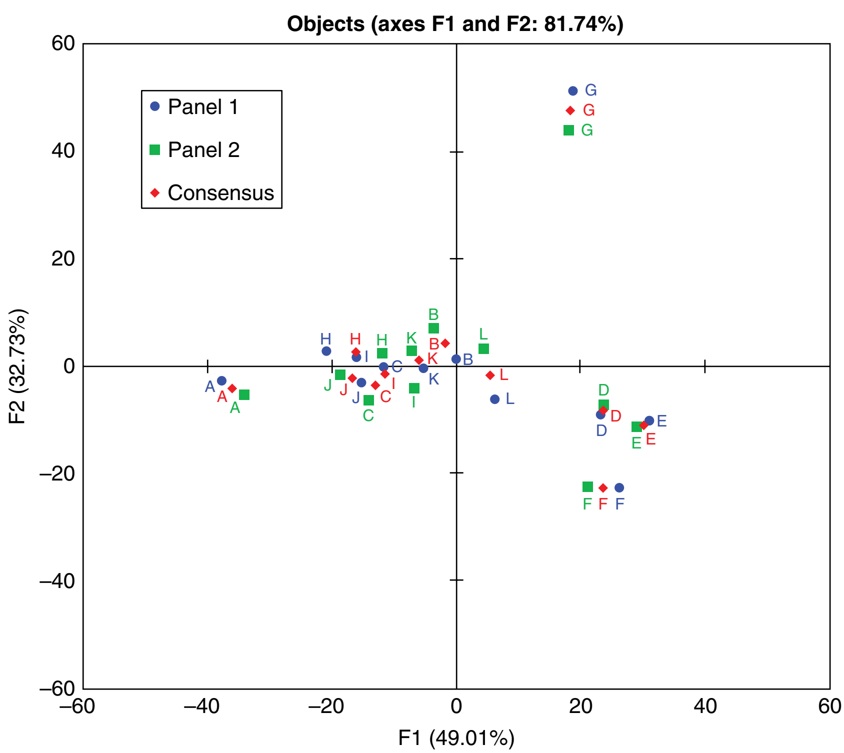


Figure 4.26 GPA analysis showing the position of each of the products A–L for both panels and the consensus (XLSTAT).

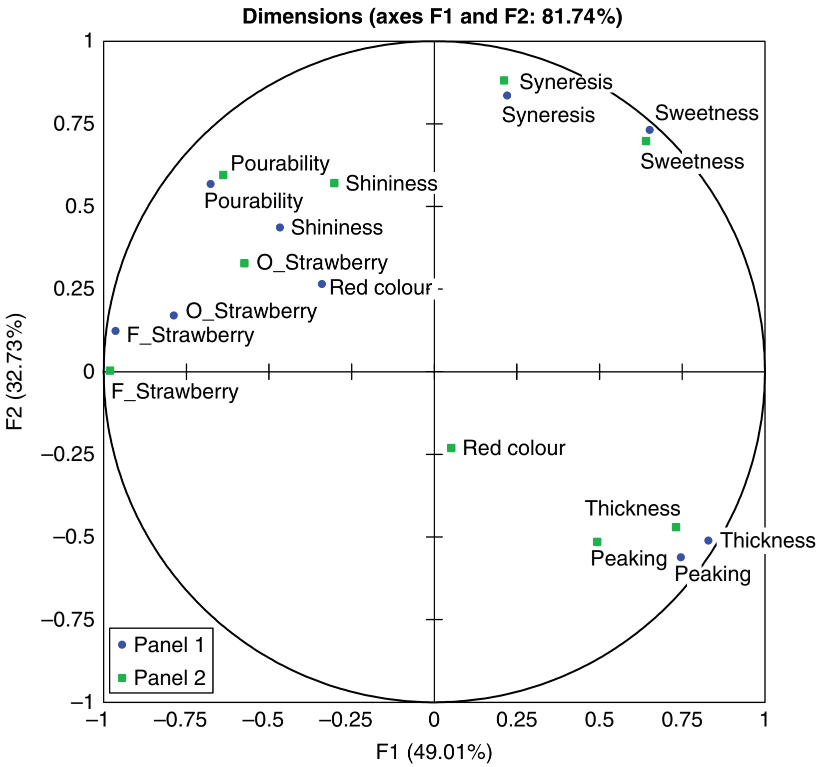


Figure 4.27 GPA analysis showing the position of each attribute for both panels (XLSTAT).

attributes. The attribute positioning in Figure 4.27 indicates that there was good agreement between the panels except for red colour. Inspection of the each panel's mean data (Table 4.4) reveals that Panel 2 ordered the samples differently compared to Panel 1, especially with respect to sample A. In addition, the range of mean values across samples for red colour was twice as large for Panel 2 as for Panel 1. Figure 4.26 shows the sample placement and the consensus for both panels and indicates good agreement overall, despite the difference in red colour perception.

If and when panel data should be combined is dependent on context and the overall sensory profiling and project objectives and process. Checks to establish equivalent or a minimum threshold level performance is just one of a set of requirements that would need to be met to confidently combine data.

4.5.2 Performance and Monitoring Issues Related to New Panellists

Panel performance measuring and monitoring systems can be developed and adapted for training and integrating new panellists. Qualitative assessments of how successfully new panellists contribute to attribute development and definition sessions can be made by panel leaders. Techniques such as visual inspection of

Table 4.4 Comparison of means from two panels for the red colour attribute.

Panel 1		**Panel 2**	
Sample name (in ascending order for intensity of red colour)	**Red colour score**	**Sample name (in ascending order for intensity of red colour)**	**Red colour score**
F	74	F	55
L	74	I	56
I	76	G	57
K	77	A	59
B	78	L	59
E	78	B	63
H	79	K	63
G	79	J	74
C	79	H	76
A	80	E	77
D	80	C	78
J	83	D	78
Range in mean scores	9	Range in mean scores	23

product/panellist profile graphs per attribute, cluster analysis, PCA and GPA can be used to understand and confirm that new panellists use attributes and scales similarly to the rest of the panel.

Agreed panellist performance targets can be used as a way of deciding when new panellist data can confidently be integrated with that from existing panellists. Giving targeted, personalized and timely feedback to trainee panellists can help speed up their training. Monitoring new panellist performance can also highlight areas for improvement if panellists have not yet reached target proficiency. For example, an action standard could be set that when an inexperienced or new panellist can personally discriminate on the majority of attributes that are discriminated at panel level, and not highly significantly contribute to panel by product interaction for any attributes for at least two consecutive ongoing projects, their data will begin to be used for project requirements. Other performance criteria or action standards might be used depending on the panel context and focus.

4.5.3 Performance and Monitoring Issues Related to Long-Standing Panellists

Well-designed and realized panel quality management systems can help to increase interest and motivation for long-standing panellists. The use of targets, data visualizations, two-way discussion and feedback, etc. will help panellists feel involved and a key part of a sensory profiling facility. Asking existing panellists to be involved in the training of new panellists can also help motivate them, and have the added benefit of facilitating new panellist integration.

Individual panellist problems, such as poor discrimination ability or lack of agreement with the rest of the panel, may develop with decreased motivation or personal issues. These should be picked up in basic panel monitoring systems, and can be brought to the attention of the panellist involved in a professional manner. The development and use of panel monitoring and feedback systems can create a culture where measurement of results and performance with respect to targets, monitoring, feedback, training and improvement is an expected and accepted part of sensory profiling panel work.

Similar to their role in helping to integrate new panellists, panel monitoring systems should highlight situations where the performance of an existing panellist is not of sufficient quality to use in project reporting and may help to explain why this is happening. This information can then be used to rectify problems.

4.6 Standards

National and international standards relating to descriptive profiling panel performance, monitoring and management are relevant and should be considered along with the concepts presented in this chapter. These standards can often provide specific statistical guidelines. In particular, ISO 11132 (ISO 2012) *Sensory Analysis – Methodology – Guidelines for Monitoring the Performance of a Quantitative Sensory Panel* is very relevant.

At the time of writing, ASTM WK8435 *New Guide for Measuring and Tracking Sensory Descriptive Panel and Assessor Performance* was under development. According to the ASTM (www.astm.org), this document will provide guidelines to assist sensory professionals in measuring performance for a given panel and assessors.

Standards such as those described above can help to clarify the objectives and design, as well as the detail, of effective panel quality management systems.

4.7 Statistics and Data Collection and Visualization

As well as univariate methods such as ANOVA, multivariate analysis of data, including techniques such as PCA, GPA, discriminant analysis/canonical variates analysis (DA/CVA) and hierarchical cluster analysis are now increasingly available in standard statistical packages. These methods can all be used to examine some elements of panel performance. In particular, multivariate methods are useful for evaluating how a panel or panellist uses attributes and views differences between samples overall, and whether or not there is a consensus between panellists or panels with respect to relevant attributes and product differences. Some typical applications of multivariate statistical techniques for panel performance measurement are shown in Table 4.5.

Table 4.5 Typical multivariate statistical techniques useful in profiling panel performance measurement.

Multivariate statistical technique	Examples of use for profiling panel performance measurement	
Principal components analysis (PCA)	• Sample space and attribute language use	• For examining panel performance overall or individual panellist performance • For comparing panels, panellists, replications
Generalized Procrustes analysis (GPA)	• Sample space and attribute language use • Use of scale	• For examining panel performance overall or individual panellist performance • For comparing panels, panellists, replications • For comparing panellists to panel consensus, panels to multi-panel consensus
Canonical variates analysis (CVA)	• Sample space, sample discrimination ability and attribute language • Agreement of panellists with respect to overall sample characterization • Identification of panellists who 'misclassify' samples	• For examining panel and panellist performance and comparing panels
Hierarchical cluster analysis	• Attribute groupings or sample groupings • Panellist groupings for an attribute	• For comparing panels, panellists, replications

An example of the use of multivariate statistical analysis is given in Figure 4.28. The first two dimensions of a DA/CVA output, showing an overall sample discrimination pattern for a profiling project with nine samples and 12 panellists and 44 attributes, is shown. Sample F is clearly discriminated from all the other samples, with all 12 panellists (symbols scattered around the consensus diamond) in reasonable agreement about its positioning. The clarity of sample F characterization compared to the other samples is also demonstrated in the confusion matrix in Figure 4.29. A benefit of the CVA plot is that it gives an overview of the relationship and discrimination between samples that is generated taking into account consistency in panellist scoring of attribute by product combinations. The DA/CVA analysis can also be used to identify panellists who 'misclassify' a high proportion of samples.

Developments in information technology and computing power and the drop in technology costs mean that now most profiling data are collected digitally and can be summarized and viewed in real time. This allows for automated evaluation and feedback systems that were not feasible in most

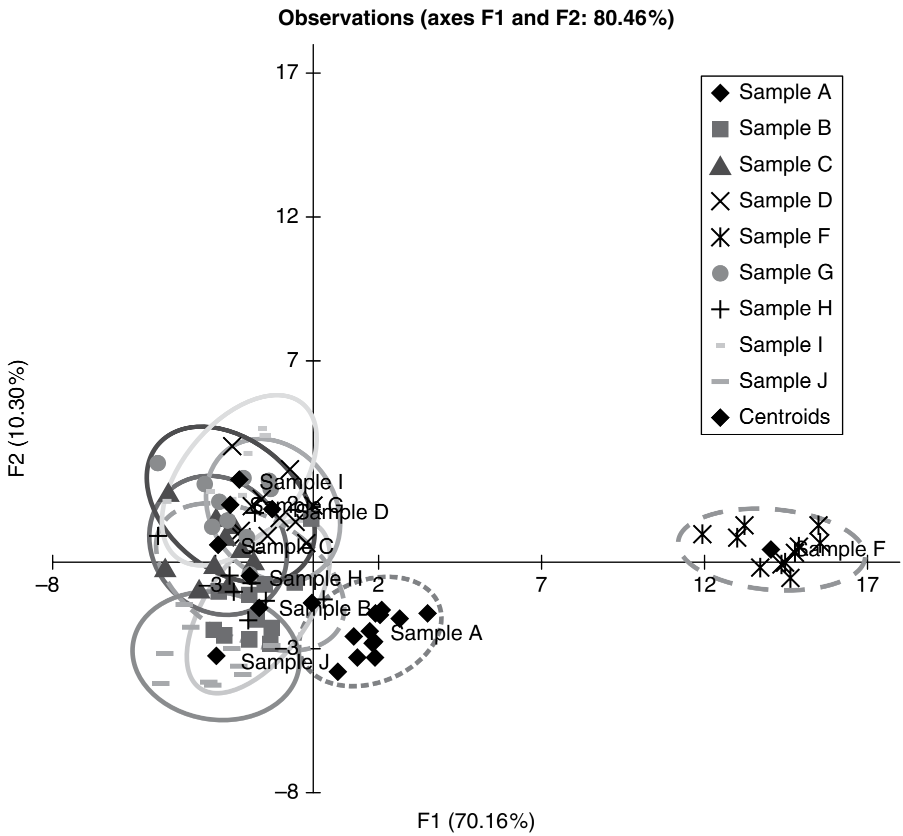


Figure 4.28 CVA/discriminant analysis plot from a data set of nine samples and 12 panellists (XLSTAT).

Confusion matrix for the cross-validation results:

from\to	Sample A	Sample B	Sample C	Sample D	Sample F	Sample G	Sample H	Sample I	Sample J	Total	% correct
Sample A	8	2	0	0	0	0	1	0	1	12	66.67%
Sample B	1	3	3	2	0	1	1	1	0	12	25.00%
Sample C	0	1	0	5	0	1	2	0	3	12	0.00%
Sample D	0	2	5	2	0	1	0	2	0	12	16.67%
Sample F	1	0	0	0	11	0	0	0	0	12	91.67%
Sample G	0	3	1	1	0	3	1	3	0	12	25.00%
Sample H	2	1	1	2	0	2	3	0	1	12	25.00%
Sample I	0	0	1	5	0	1	2	3	0	12	25.00%
Sample J	0	2	2	0	0	2	1	0	5	12	41.67%
Total	12	14	13	17	11	11	11	9	10	108	35.19%

Figure 4.29 Confusion matrix from CVA/discriminant analysis (XLSTAT).

sensory evaluation facilities just a few years ago. The output of these feedback systems and statistical analyses can be relevant for panel leaders, sensory scientists and sensory panellists.

The development of statistical software that is adapted for sensory evaluation applications has also resulted in a multitude of possible ways for analysing panel data and calculating and using performance and monitoring measures. In particular, there is a large range of data acquisition and/or statistical analysis packages available for descriptive analysis. Many of these include features designed specifically for evaluating panel or panellist performance measurement. A few examples of available applications are listed.

- SenPAQ version 6 (www.qistatistics.co.uk) automatically evaluates and produces tables and summary graphics related to panellist performance within a profile in terms of discrimination, repeatability and consistency, comparing individual performance to that of the whole panel. SenPAQ also produces discrimination versus consistency and discrimination versus repeatability graphics comparing panellist performance for each attribute. Some examples of the tables and graphs from SenPAQ are used within this chapter.
- PanelCheck V1.4.0 (www.panelcheck.com) offers a range of easily accessible plots to visualize a wide range of panel performance measures, many of which could be used for panel feedback exercises. These include p-MSE and eggshell plots. Some PanelCheck plots are used as examples within this chapter.
- FIZZ (www.biosystemes.com/) includes extensive graphical and tabular panel performance evaluation options within its calculations module and some of the output has been shown in this chapter. In particular, judge performance graphs (shown in the panel performance section) may be based on repeatability indices of standard deviation, range or coefficient of variation for product by attribute data for individual panellists and the panel as a whole. Another available index is called the CV ANOVA which is the RMSE divided by overall mean for an attribute across products. The CV ANOVA is calculated for an individual panellist and can also be averaged across panellists for a CV ANOVA measure for the panel as a whole. Data points that incorporate product effect significance, for all attributes and panellists, and the panel as a whole, can be combined on one graph and plotted against attribute mean, range, coefficient of variation or CV ANOVA. An example of a summary judge performance graph incorporating panellist and panel information for three attributes is shown in Figure 4.30. It is possible to also create customized tables of these performance measures; so for example, a table can be generated displaying the calculated range of the replicates of every attribute by product pair for each judge (panellist).
- As well as featuring different options for panel performance checks (scatterplots were shown earlier), Compusense (www.compusense.com) offers the feedback calibration method (FCM) facility, which is designed to help with scaling calibration. The FCM process uses preset or derived attribute targets and ranges and is described as providing panellists with ‘immediate, individualized visual feedback to help them learn attributes and how to scale them’. This feedback method can be used during training and generates reports on hits and misses, distance from target measurements and calibration charts that categorize panellist performance into ‘excellent’, ‘moderate’ and ‘unsatisfactory’ groupings. This is a good example of how performance management can be linked directly to training, as well as how real-time automation options are now an important part of the overall panel quality management landscape.

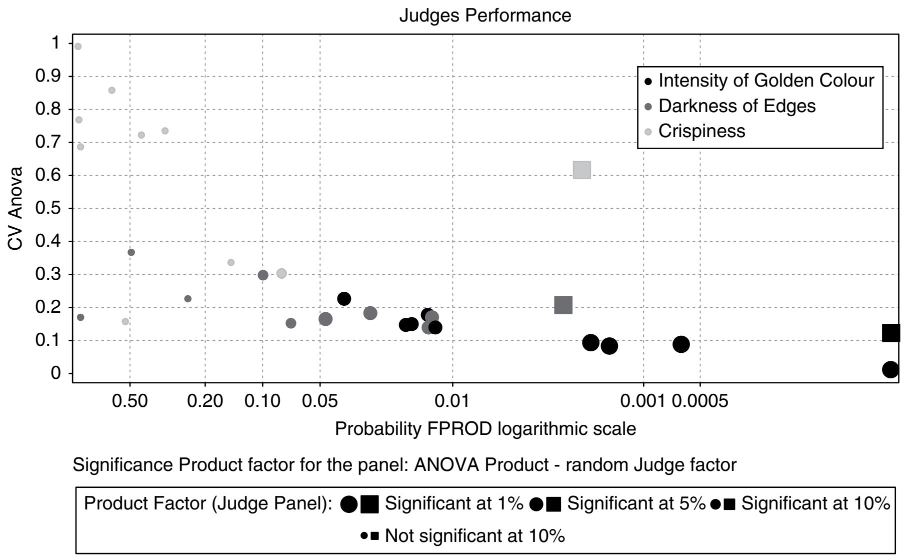


Figure 4.30 Overall judge performance plot incorporating information from three attributes (FIZZ).

- Other newer packages that have useful functions for panel performance include RedJade powered by Tragon (www.redjade.net) and EyeopeneR from EyeQuestion (https://eyequestion.nl/).

The scope for automated data collection and analysis to support panel quality management is large. The above examples only demonstrate elements of what a particular application can do and are only a selection of the applications available. Using tailored systems can allow much easier development of a workable panel performance and monitoring system. In some cases, there is the possibility of creating macros or databases within or that tie in with these systems to allow for customized automation of panel monitoring.

In addition to systems designed in particular to measure panel performance, there are many other statistical package options that can support panel quality management. These include XLSTAT (www.xlstat.com) which is an add-on to Excel and has many analysis options for sensory data (with some specifically designed to help evaluate panel performance), and other statistical packages such as Minitab (www.minitab.com), SensoMineR (http://sensominer.free.fr/) and SAS (www.sas.com/en_us/software/stat.html).

4.8 Summary

In the same way that it is essential that instrumentation is serviced and its output verified, carrying out checks on performance is critical for every type of sensory panel and panellist. This chapter has introduced the concept of panel quality management, defined by the authors as consisting of three areas of performance, monitoring and proficiency. Four main quality measures for panel quality management, repeatability, reproducibility, consistency and discrimination, have been described alongside key statistical criteria and tests, practical suggestions, dos and don'ts, real data examples and stepwise approaches to assessing and correcting issues.

Panel quality management is a vital part of any sensory science function but will also require other standardized approaches to be fully effective, such as training programmes, use of sensory standards (ISO or ASTM), procedures, protocols and correct project design. It is also important to note that any issues discovered while assessing the sensory panel data may not be associated solely with the panellists' performance; they may also be due to preparation errors, sample variability, experimental design, etc. Approaches to panel quality management will also vary depending on context.

4.9 Future Developments

The challenge and opportunity for the future of panel quality management is the creation of effective yet easy-to-use systems that geographically dispersed stakeholders can engage with. Developments in information management, graphics and data visualization mean there are emerging options that will be available to the sensory scientist to visualize, investigate and summarize complex sensory data and panel performance trends in an interactive manner. Interactive dashboards of performance data could be presented and queried in many different ways to look at the data from different angles – for example, panellist performance over time or comparing individual panellists' performance.

In addition, rapid profiling methods are becoming an important tool in the sensory toolbox. These will require adaptations and additions to more traditional panel quality management techniques. Rapid panel performance methods may incorporate elements of real-time feedback to the panellists themselves to optimize outputs in an efficient and timely manner.

There are likely to be developments in these areas over the coming years which will make panel quality management more flexible, easier and more creative.

References

Catchalian, M.M., de Leon, S.Y. & Yano, T. (1991) Control chart technique: a feasible approach to measurement of panellist performance in product profile development. *Journal of Sensory Studies*, **6**, 239–254.

European Co-operation for Accreditation (2017) *EA – 4/09 G: 2017 Accreditation for Sensory Testing Laboratories*. Paris: European Accreditation.

Findlay, C.J., Castura, J.C. & Lesschaeve, I. (2007) Feedback calibration: a training method for descriptive panels. *Food Quality & Preference*, **18**, 321–328.

Hyldig, G. (2010) Proficiency testing of sensory panels. In: Kilcast, D. (ed.) *Sensory Analysis for Food and Beverage Quality Control*. Cambridge: Woodhead Publishing, pp. 37–48.

International Organization for Standardization (ISO) (2012) *Sensory Analysis – Methodology – Guidelines for Monitoring the Performance of a Quantitative Sensory Panel*. ISO 11132. Geneva: International Organization for Standardization.

International Organization for Standardization (ISO) (2014) *Sensory Analysis – General Guidelines for the Selection, Training and Monitoring af Selected Assessors and Expert Sensory Assessors*. ISO 8586. Geneva: International Organization for Standardization.

Kemp, S.E., Hollowood, T.H. & Hort, J. (2009) *Sensory Evaluation: A Practical Handbook*. Chichester: Wiley-Blackwell.

Kuesten, C.L., McLellan, M.R. & Altman, N. (1994) Computerised panel training: effects of using graphic feedback on scale usage. *Journal of Sensory Studies*, **9**, 413–444.

McEwan, J.A., Heinio, R.L., Hunter, E.A. & Lea, P. (2002a) Proficiency testing for sensory ranking panels: measuring panel performance. *Food Quality & Preference*, **14**, 247–256.

McEwan, J.A., Hunter, E.A., van Gemert, L.J. & Lea, P. (2002b) Proficiency testing for sensory profile panels: measuring panel performance. *Food Quality & Preference*, **13**, 181–190.

Naes, T., Hirst, D. & Baardsetii, P. (1994) Using cumulative ranks to detect individual differences in sensory profiling. *Journal of Sensory Studies*, **9**, 87–99.

Pfeiffer, J.C., Kapparis, E. & Gilbert, C.C. (2008) *Guideline No. 58: Practical Guidelines for Monitoring On-Going Job Performance of Sensory Descriptive Panellists*. Chipping Campden: Campden & Chorleywood Food Research Association.

Pritchett. Mangan,. PA. (1992) Performance assessment of sensory panellists. *Journal of Sensory Studies*, **7**, 229–252.

Stone, H. & Sidel, J.L. (1985) *Sensory Evaluation Practices*. Orlando: Academic Press.

Tormod, N., Per, B.B. & Tomic, O. (2010) *Statistics for Sensory and Consumer Science*. Chichester: Wiley.

Wilkinson, C. & Yuksel, D. (1997) Modelling differences between panellists in use of measurement scales. *Journal of Sensory Studies*, **12**, 55–68.

CHAPTER 5

Statistical Analysis of Descriptive Data

Anne Hasted

5.1 Introduction

5.1.1 Objectives

Statistical analysis is a key component of descriptive sensory data collection; the data collected are subject to uncertainties due to assessor-to-assessor variation in assessment, sample variation and inherent measurement variation in the interpretation of scales. The attribute-by-attribute sensory data analysis with which we start addresses two important questions.

- How do the samples differ in their sensory aspects? Which differences can be reported as real repeatable differences and which may be only due to chance variations?
- How are the panel performing, both as a whole unit and as individuals?

The first of these questions is addressed in this chapter, the second has been covered in Chapter 4.

Once we have determined which attributes show real differences between the samples, we can then use the sample means on these attributes as input into multivariate statistical techniques which will help us see the bigger picture of sample differences and the inter-relationships between the sensory attributes.

5.1.2 Data Collection

The structure of the data collection must be carefully designed to ensure that biases in the data due to assessor fatigue or carry-over effects from one sample to another are minimized. If practically possible, the test should be replicated so that each assessor sees each sample twice or three times in different sessions. The presentation order within a session should, if possible, be randomized for each assessor. Replication is an important component of panel performance assessment; in the context of this chapter, it adds to the robustness of the data by representing an assessor's score for a sample as an average across attribute replicates, rather than a single score. Statistical analysis to investigate sample differences can be carried out

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

with data from a single replicate or multireplicated design. Modern software copes with missing values in the data due to an assessor being absent from a session and adjusts the results, presenting adjusted means (usually referred to as least square means) which are model-based estimates of the mean scores which would have been obtained if the data were complete. If the data are so incomplete that there is not at least one score on each sample for a particular assessor, then it is advisable to remove that assessor from the data set before analysis.

The calculations outlined in section 5.2 assume complete data.

5.1.3 Example Data

The data used to illustrate the statistical analysis in this chapter are from a large study on apple varieties:12 assessors testing 10 unpeeled apple varieties for texture on a 0–100-line scale in three replicates.

5.1.4 Preliminary Analysis

Before modelling the data, preliminary data checks should be made to check for the number and structure of missing values using cross-tabulations (either provided by the data collection software or by using EXCEL pivot tables). The scoring ranges for each sensory scale should also be evaluated; this can be done either graphically using histograms or boxplots or numerically using descriptive statistics (count, mean, median, range, standard deviation).

5.2 Analysis of Variance

The first statistical task is to determine whether differences between samples are identified for each sensory attribute. Sample differences are characterized by the sample mean scores across the panel and the task can be expressed as 'Is the variation in the sample means greater than we would expect due to panel variation?'. This question is addressed by the technique called analysis of variance (ANOVA).

5.2.1 Panel Variation

Table 5.1 shows the raw data for four of the assessors scoring the same apple variety for the texture attribute juicy.

There is variation in the data between assessors; this can be seen by comparing their mean scores; assessor number 3 tends to use a higher scoring range to score juicy than the other three assessors. There is also variation in the three replicate scores for each assessor; this is most probably scoring variation in each assessor's perception of juiciness but could also be due to inherent variation in the juiciness of each apple slice tasted. Comparing this within assessor repeatability variation for different assessors is a key component of panel performance.

The other important source of variation is in differences between the assessors in how they perceive the sample differences. This is termed the assessor by sample interaction.

Table 5.1 Raw data for four assessors scoring juicy for the same sample in three blind replicates.

Assessor	Session	Juicy	Mean score
1	1	49	
1	2	59	58.0
1	3	66	
2	1	59	
2	2	51	52.3
2	3	47	
3	1	73	
3	2	81	73.3
3	3	66	
4	1	57	
4	2	67	55.3
4	3	42	

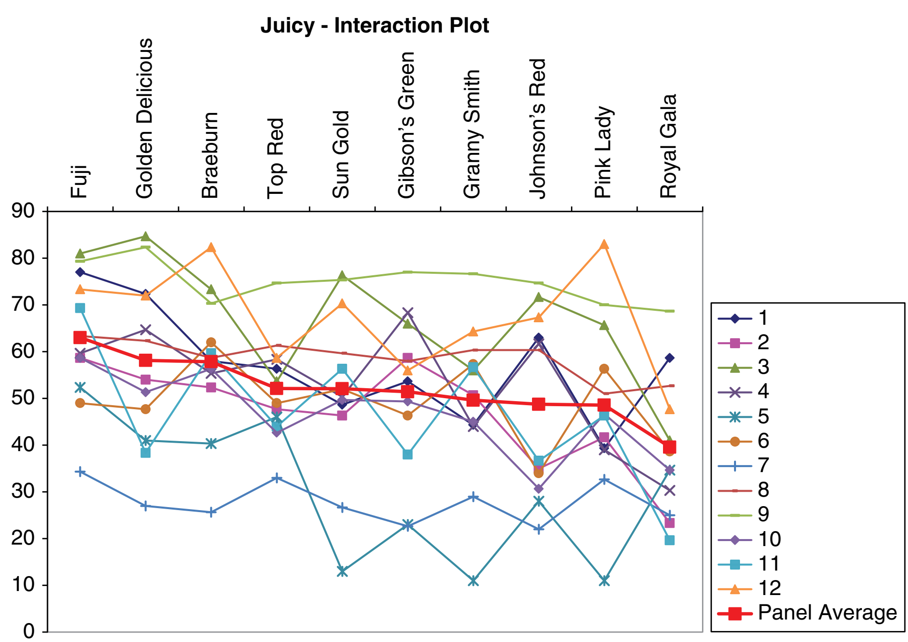


Figure 5.1 Assessor by sample interaction plot for juicy.

Figure 5.1 shows a plot of the sample mean scores across replicates for each assessor. If all assessors were perceiving the sample differences in the same way, but with their own average scale reference, then we would expect the traces to be parallel. The greater the differences in patterns across the samples between assessors, the higher the assessor by sample interaction. The interaction can either be due to scaling effects, some assessors using the scale more widely than others, or to cross-over effects where assessors perceive the sample differences in different ways. It is this assessor by sample interaction variation that is the key measure in determining whether there is a panel consensus on the sample differences and it is used as the baseline measure of panel 'noise' variation in the statistical tests for product differences that follow.

5.2.2 ANOVA Models

The analysis of the data is based on fitting a statistical model to the data to explain the variations in the scores for each sensory attribute. This model allows us to separate out and quantify the variation observed in the data, attributable to the products we are testing (samples) and the panel variation (both scale effects for assessors and assessor by sample interactions).

For replicated sensory tests, the additive ANOVA model for a sensory attribute Y in a design with p samples, q assessors and r replicates is:

$$Y_{ijk} = Intercept + Sample_i + Assessor_j + Sample \times Assessor_{ij} + Error_{ijk}$$

where:

$Sample_i$ ($i = 1$ to p) measures the sample effects – differences in average sensory scores between products

$Assessor_j$ ($j = 1$ to q) measures the assessor effects – differences in average sensory score between assessors

$Sample \times Assessor_{ij}$ measures the interaction effects – the variation in perception of sample differences between assessors

$Error_{ijk}$ measures the replication variability within assessors and samples.

This gives an ANOVA table.

Source	df	Sum of squares	Mean square
Sample (S)	p-1	SS(S)	MS(S)
Assessor (A)	q-1	SS(A)	MS(A)
Sample × Assessor (S × A)	(p-1)(q-1)	SS(SxA)	MS(SxA)
Error	pq(r-1)	SS(Error)	MS(Error)
Total	pqr-1	SS(Total)	

where df stands for degrees of freedom. This measures the amount of information available to measure variation at the level of the source.

The sample sum of squares SS(S) is a scaled measure of the summed variation in the sample means. The greater the differences in the sample means, the greater the sum of squares. Similarly, the assessor sum of squares SS(A) measures the variation in the assessor mean scores (across all the samples).

The assessor by sample sum of squares SS(S × A) measures the deviation in the pattern of assessor by sample scores from a pattern of parallel traces which would indicate that assessors were perceiving the sample differences in exactly the same way. The larger the sum of squares, the higher the deviation in the individual assessor mean scores from a consensus. It is this variation that should be used as the panel 'noise' variation measure in testing for differences in sample means.

The error sum of squares SS(Error) measures the total variation within the p by q sample by assessor replicated tests.

The column labelled Mean square is calculated as the sum of squares/df.

The model is defined as a mixed model with fixed sample effects and random assessor effects; this ensures that the sample effects are tested against the assessor by sample interaction (for details of mixed models, see Appendix 5.1 at the end of this chapter). If a standard fixed effect model is fitted then by default, all model terms are tested against the error mean square which will result in inappropriate significance tests.

To test for sample differences: the variance ratio $F = MS(S)/MS(S \times A)$ is tested against an F distribution with (p-1),(p-1)(q-1) degrees of freedom.

The null hypothesis is that there are no sample differences and so the variation observed in the sample means is purely due to random variations in the panel, in which case the calculated F would be close to 1. The larger the variance ratio F, the less likely this hypothesis is to be true. The tabulated F distribution gives us the probability of obtaining the observed F or higher if the null hypothesis was true. This is the significance level of the test.

To test for sample × assessor interaction: $F = MS(S \times A)/MS(Error)$ is tested against an F distribution with (p-1)(q-1)/pq(r-1) degrees of freedom

Example:
Apple profile data: Table 5.2 gives the analysis of variance of the attribute juicy. Do we have evidence of a difference in average juiciness between the samples? From the ANOVA table F (samples) = 6.48 with $P < 0.0001$.

We have strong evidence of differences in average juiciness between the samples. The P-value tells us that if we report that the apples differ in average juiciness, the chance that we are wrong in making this statement (and the differences have occurred by chance) is less than 0.0001 (a 1 in 10 000 chance).

The F test for the interaction ($F = 1.27$, $P = 0.07$) indicates that we do not have strong evidence that the interaction variation is greater than would be expected due to the level of replication variation (measured by the error mean square).

Table 5.2 Analysis of variance of the attribute juicy.

Source	DF	Sum squares	Mean square	F-value	Pr>F
Assessor	11	66721	6065.6	26.16	<.0001
Sample	9	13521	1502.3	6.48	<.0001
Assessor*Sample	99	22952	231.8	1.27	0.070
Error	240	43665	181.9		
Total	359	146859			

The error mean square =181.9; this is the average variance of the assessor replicate scores – taking the square root of this value gives the average replication standard deviation of 13.5 units on the 100-point scale, which is relatively high.

5.2.3 Comparison of Means

The analysis of variance gives us an indication of differences in sample mean scores for an attribute. It is a very general test, testing the overall variation in the sample means. To interpret the results, the mean scores are tabulated and pairwise comparisons made using various comparison tests. The underlying metric for any comparison test is the *standard error* of the mean which measures the precision of the mean, allowing for panel measurement uncertainty measured by the assessor by sample interaction and the size of the test (number of assessors by number of replicates)The standard error of mean is calculated by the formula:

$$s.e.\left(\bar{X}\right) = \sqrt{\frac{S^2}{n}}$$

where S^2 = sample by assessor interaction mean square with (p-1)(q-1) degrees of freedom and n = number of measurements in each average = rq. The software you use may not print out the value of the standard error but it will use it in any comparison test that you specify.

The fundamental pairwise comparison is made using a t test. Provided there are no missing values, the formula for the t test is:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{2} \times s.e.\left(\bar{X}\right)}$$

which is compared to tabulated t values with (p-1)(q-1) degrees of freedom to obtain the statistical significance of the test. (If there are missing data, the t test can still be carried out although the equation is slightly more complicated.)

Alternatively, if only a rough measure of which differences are statistically significant is required, a *least significant difference* (LSD) measure can be calculated. This gives the smallest difference in means that will give a significant difference at the prescribed significance level.

$$LSD\left(5\%\right) = T_{(p-1)(q-1)}\left(5\%\right) \times \sqrt{2} s.e.\left(\bar{X}\right)$$

Example:
Comparison of means for juicy.

Using the results from the ANOVA table in Table 5.2:

$$LSD\left(5\%\right) = T_{99}\left(5\%\right) \times \sqrt{2} s.e.\left(\bar{X}\right) = 1.98 \times \sqrt{2} \times \sqrt{\frac{231.8}{36}} = 7.1$$

Means are then presented in rank order. A common convention is to code the differences so that means that are not significantly different share the same

Table 5.3 Mean scores for the attribute juicy with LSD (least significant difference) (5%) pairwise comparisons.

Variety	Mean	LSD (5%) groupings
Fuji	63.0	a
Golden Delicious	58.1	ab
Braeburn	57.9	ab
Top Red	52.1	bc
Sun Gold	52.1	bc
Gibson's Green	51.4	bc
Granny Smith	49.6	c
Johnson's Red	48.8	c
Pink Lady	48.6	c
Royal Gala	39.6	d

letter. Table 5.3 shows mean scores for the 10 apple varieties with differences coded using the least significant difference

5.2.4 Multiple Comparison Tests

The simple LSD method of comparison, described above, is based on a t test; the 5% significance threshold has the property that for each pairwise test carried out, there is a 5% risk of wrongly reporting a difference in means as significant when the true difference is not detectable. The more samples in a test, the more chance there is of detecting spurious differences between pairs of samples. So in a multiproduct trial, where all pairwise comparisons are of interest, it may be appropriate to reduce the risk of finding spurious differences by using a test that adjusts for this by setting the 'family' error rate across all pairwise tests to 5%. Many tests and approaches to this problem have been devised, all giving slightly different results. The most widely used multiple comparison test is the Tukey HSD test (HSD standing for Honestly Significantly Different). For this test the threshold value for significance of:

$$t = \frac{\bar{X}_A - \bar{X}_B}{\sqrt{2} \times s.e.(\bar{X})}$$

is tabulated and depends on both the degrees of freedom of the assessor by sample interaction and the number of samples in the test. Table 5.4 shows the significance of pairwise comparisons using the Tukey test; comparing this with Table 5.3, we can see that the test is much more conservative as fewer significant differences are flagged.

The question, of course, is, what method should you use for your data? If you are testing samples with specific pairwise comparisons planned before you start the analysis then it is considered safe to use the simple LSD test. If, however, you are carrying out a category appraisal using many products and you are

Table 5.4 Mean scores for the attribute juicy with Tukey honestly significantly different (HSD) (5%) pairwise comparisons.

Variety	Mean	Tukey HSD (5%) groupings
Fuji	63.0	a
Golden Delicious	58.1	ab
Braeburn	57.9	ab
Top Red	52.1	ab
Sun Gold	52.1	ab
Gibson's Green	51.4	ab
Granny Smith	49.6	bc
Johnson's Red	48.8	bc
Pink Lady	48.6	bc
Royal Gala	39.6	c

interested in any differences that you can find then a multiple comparison test will be more appropriate.

Which multiple comparison test to use is a question frequently posed to statisticians. Basically, the tests fall into two types: those which adjust the threshold for significance of pairwise tests to take into account the number of products tested (e.g. Tukey HSD, Bonferroni) and those which adjust the threshold for significance based on the ranked position of the mean scores (e.g. Duncan, Student Neuman Keuls, REGWQ). This author would recommend Tukey HSD for studies where there is no structure to the samples.

Another common sample structure is when several samples are tested together with a control. If the only comparisons of interest are between each of the test samples and the control (and not between the test samples themselves) then Dunnett's test should be used in preference to the Tukey HSD test.

Some statisticians will advise you not to carry out any pairwise comparisons unless the F test for sample differences in the ANOVA is significant at $P = 5\%$. In practice, this means that you are unlikely to detect as significant an attribute which gives very similar scores for all but one sample, for example sample mean scores $A = 15.1$, $B = 15.5$, $C = 14.9$, $D = 15.4$, $E = 27.1$. The F test is testing the average variation over the samples whereas the pairwise comparison tests are likely to detect sample E as significantly different to all other samples. Even if the F test is not significant, it is still worth casting an eye over the pairwise comparisons to check that there are no significant pairwise differences.

When worrying about which test to use, it is important to remember what statistical significance is and is not measuring.

5.2.5 Statistical Significance

Statistical significance measures the risk in reporting signals of differences in sensory perception between your samples measured by the panel mean scores. In the framework of statistical hypothesis testing of pairwise differences, it measures the chance of reporting two samples as different in their mean scores when the true difference is zero. It has become accepted practice in sensory science to set this risk level at 5%, but to some extent this is an arbitrary choice.

Informally, you can interpret this significance measure as a measure of the repeatability of your result. If you were to perform the test again using the same number of assessors and the same number of replicates under exactly the same conditions then it is the chance that you would not get repeat directional results.

It is important to understand that statistical significance does not necessarily imply commercial importance. If you have a large, highly trained panel and carry out several replicate sessions, you may find very small differences in mean scores on the sensory scale are flagged as statistically significant. On the other hand, a poorly trained or small panel may fail to detect relatively large differences in mean scores as statistically significant and if these differences were detected, they would be commercially or scientifically important.

Too often, sensory scientists base all of their decision making on statistical significance without reference to the difference in product quality measured by the sensory scale.

5.2.6 Visualizing Sample Differences

Sample differences measured by mean scores across attributes can be visualized in a number of different ways.

The star diagram plots the mean scores of samples by attribute to highlight similarities and differences between the samples. However, it can be difficult to show significant differences in this format. A tape plot plots the attribute means linearly by sample with ±1/2 LSD or HSD (Tukey) bands; at each attribute point, if the bands do not overlap the samples are significantly different for the criterion used. Figure 5.2 shows a star diagram for two apple samples, Fuji and Gibson's

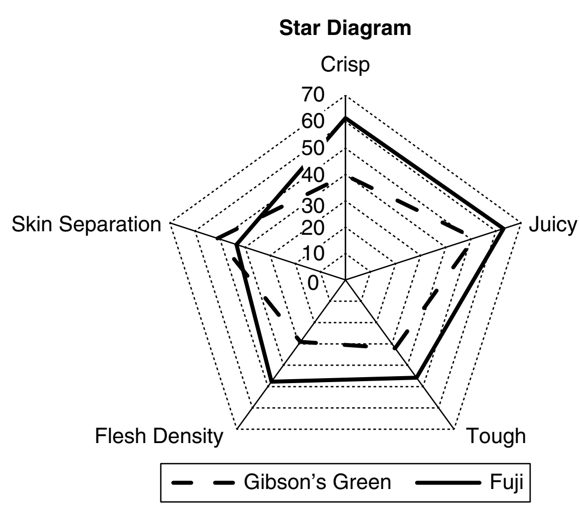


Figure 5.2 Star diagram of apple texture attributes.

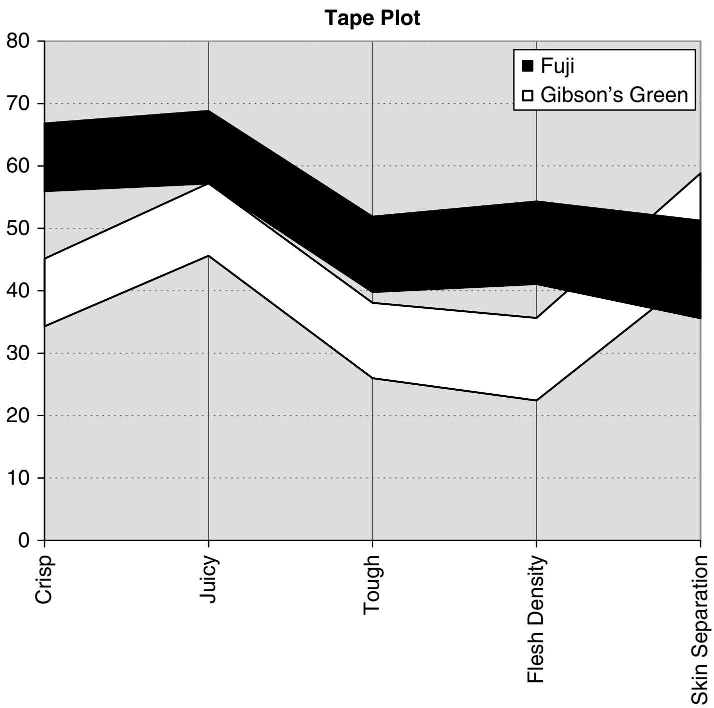


Figure 5.3 Tape plot of apple texture attributes with Tukey HSD (5%). Tapes are centred on attribute mean scores with tape width ±1/2 HSD from centre.

Green. Fuji is less crisp, less juicy, less tough and has less density of flesh and may be higher in skin separation. The tape plot in Figure 5.3 confirms the significant difference for all selected attributes apart from skin separation, since the tapes only overlap for skin separation so these means are not significantly different using the Tukey HSD criterion.

If many attributes are to be compared simultaneously then a simple line plot of the average attribute scores by sample will be easier to read but harder to overlay information on significant pairwise differences.

5.2.7 Assumptions

The analysis of variance model assumes that both the assessor by sample interaction variation and the error variation in replicates follow a normal distribution with constant variance across assessors and samples. The method has been shown to be reasonably robust to departures from this normality assumption, particularly for the size of raw data set typically encountered in analysis of sensory data. It must be emphasized that these assumptions are made on components of the panel variation and not on the raw data themselves, so testing or assessing normality of the raw data is not appropriate. In analysis of sensory panel data, if problems with the assumptions occur, it is usually in the

Table 5.5 Mean scores for flavour attribute – unripe apple.

Mean scores	Braeburn	Fuji	Gibson's Green	Golden Delicious	Granny Smith	Johnson's Red	Pink Lady	Royal Gala	Sun Gold	Top Red	LSD (5%)
Flavour unripe apple	5.6	3.4	8.3	3.3	26.0	0.2	12.6	0.2	18.8	0.4	**8.3**

LSD, least significant difference.

case where many of the attribute scores for some samples are zero or close to zero but the attribute is identified in other samples. Table 5.5 shows the mean scores and LSD(5%) for the attribute unripe apple flavour where 79% of the individual scores are under 5 (on a 100-point scale). Clearly, the LSD(5%) value is inappropriate for comparing means. The assessor by sample variation will be close to zero for samples with low detection of the attribute and higher for samples where the attribute is detected and so the assumption of homogeneity of the assessor by sample variation is violated.

The author's recommendation is that for attributes of this type, results should be presented as mean scores without any pairwise significance tests based on the analysis of variance approach. If statistical significance measures are absolutely necessary then non-parametric methods of analysis can be employed. The appropriate tests are as follows: for tests with more than two samples, Friedman's analysis of the assessor by sample mean scores; for profiles with only two samples, the Wilcoxon signed rank test can be used.

5.3 Multivariate Data Display

The objective of analysis of descriptive sensory data is to quantify the differences between the samples using panel mean scores for attributes which are shown to differ across the samples and for which there is a reasonable consensus across panellists. The level of discrimination and consensus is evaluated using the univariate (one variable at a time) ANOVA explained in section 5.2. The results are presented as a table of mean scores together with measures of significance. However, the univariate analysis does not evaluate or take into account the inter-relationships between the sensory variables. This can be achieved using a technique called multivariate analysis of variance (MANOVA) which is briefly covered in section 5.3.3 but first, we will look at the most widely used technique for obtaining graphical displays which help to illustrate overall sample differences and inter-relationships between the sensory variables.

Principal component analysis (PCA) is based on a matrix decomposition of either the covariance or correlation matrix of the sensory data. Before we look at the method in detail, we need to look at the measures of covariance and correlation.

We will illustrate the multivariate techniques using the texture profile of the apples; the panel mean scores for attributes where the panel found significant differences in the sample means are given in Table 5.6. The table also shows the variance of the samples over each attribute. The apples differ most in flesh density (variance = 217.2) and least in the attribute fibrous (variance = 17.8).

5.3.1 Covariance and Correlation

We are now interested in measuring the strength of relationship between sensory variables. Although the panel may have defined 30–40 attributes to describe product differences, it is very unlikely that these are 30–40 completely independent measures. For example, the attributes fibrous and tough are likely to distinguish the apples in similar ways.

The two related measures of association between attributes are covariance and correlation. Covariance measures the linear association between a pair of variables taking into account the scale used for each variable and is scale dependent. Correlation is independent of scale

If we consider any pair of sensory variables X and Y with measurements (X_i, Y_i) on a set of samples, the covariance between X and Y is defined as:

$$Cov(X,Y) = \frac{\sum_i (X_i - \bar{X})(Y_i - \bar{Y})}{n-1}$$

If X and Y tend to increase together the covariance will be positive. If as X increases Y decreases, the covariance will be negative. The stronger the trend in the data, the higher the covariance.

Table 5.7 shows the covariance matrix of the sensory texture measures on the apples. The diagonal terms of the matrix give the variance of each attribute. The variable with the highest positive covariance with crisp is flesh density (covariance = +163.2) and the variable with the highest negative covariance with crisp is floury (covariance = -84.9). Crisp and flesh density are positively associated, as apples which are crisper will tend to have higher scores of flesh density, Crisp and floury are negatively associated, as apples which are crisper will tend to be less floury. The covariance is scale dependent but because all our variables are on the same scale, we can compare the values (this would not be the case if you were dealing with analytical data where each attribute was measured on a different scale).

Because of the issues of scale, a more widely used measure of linear association is the Pearson's correlation coefficient r.

$$r(X,Y) = \frac{Covariance(X,Y)}{SD(X) \times SD(Y)}$$

Table 5.6 Panel mean scores for texture attributes.

Sample	Crisp	Juicy	Skin separation	Tough	Flesh density	Fibrous	Floury	Pulpy	Fibrous bits	Skin bits
Gibson's Green	38.7	51.0	52.2	32.8	29.4	4.2	13.3	18.1	3.6	51.4
Johnson's Red	38.7	47.1	51.8	33.5	28.2	3.0	18.9	7.6	4.9	44.8
Golden Delicious	54.9	56.9	40.7	42.4	43.5	9.4	3.4	7.4	9.2	44.3
Granny Smith	59.9	48.3	53.5	58.5	58.9	15.2	0.3	5.2	17.0	55.3
Pink Lady	66.9	45.4	38.6	61.0	67.5	10.2	0.6	3.7	12.1	50.2
Fuji	60.6	62.1	45.9	47.3	45.9	10.7	1.4	6.5	9.6	46.6
Top Red	47.1	51.5	61.5	38.9	38.0	7.0	10.0	10.4	6.4	56.2
Braeburn	59.7	55.6	48.4	49.2	49.6	7.1	2.9	6.8	10.3	43.7
Royal Gala	37.6	38.9	42.8	37.2	32.8	3.4	18.2	4.7	3.4	35.2
Sun Gold	67.7	50.4	43.4	62.5	67.9	13.7	0.2	3.2	12.8	56.3
Variance	138.4	42.2	48.9	126.1	217.2	17.8	56.8	18.6	19.3	45.8

Table 5.7 Covariance matrix of apple texture variables.

Variables	Crisp	Juicy	Skin separation	Tough	Flesh density	Fibrous	Floury	Pulpy	Fibrous bits	Skin bits
Crisp	**138.4**	29.3	−33.6	122.9	163.2	42.5	−84.9	−30.1	45.0	33.4
Juicy	29.3	**42.2**	3.4	3.7	6.5	8.3	−25.2	5.5	6.2	8.7
Skin separation	−33.6	3.4	**48.9**	−30.9	−43.4	−5.2	16.3	15.9	−5.5	21.0
Tough	122.9	3.7	−30.9	**126.1**	164.2	41.2	−72.4	−32.9	44.7	34.0
Flesh density	163.2	6.5	−43.4	164.2	**217.2**	53.0	−96.3	−41.2	57.0	46.6
Fibrous	42.5	8.3	−5.2	41.2	53.0	**17.8**	−28.4	−9.0	17.4	17.0
Floury	−84.9	−25.2	16.3	−72.4	−96.3	−28.4	**56.8**	13.7	−29.1	−23.8
Pulpy	−30.1	5.5	15.9	−32.9	−41.2	−9.0	13.7	**18.6**	−10.8	4.4
Fibrous bits	45.0	6.2	−5.5	44.7	57.0	17.4	−29.1	−10.8	**19.3**	14.9
Skin bits	33.4	8.7	21.0	34.0	46.6	17.0	−23.8	4.4	14.9	**45.8**

here SD(X) = standard deviation (X), SD(Y) = standard deviation(Y). This measure is scaled to be between +1 and −1 and is independent of measurement scales.

Table 5.8 gives the correlation matrix between the attributes. Typically, pairs of variables with correlation coefficients numerically greater than 0.7 can be identified as showing a linear association. Crisp and tough show a strong positive linear association ($r = +0.94$), crisp and floury a strong negative linear association ($r = -0.96$) and crisp and skin bits show no strong linear relationship ($r = +0.42$). Correlations can be visualized using scatter plots; these are useful to confirm linearity in the association between variables. Scatter plots of the three pairs of correlations with the attribute crisp are shown in Figure 5.4.

It is recommended that a minimum of 5–6 samples is required to allow investigation of the correlation structure of sensory variables. Although in theory, the minimum number of samples to calculate a correlation is three, with this very small sample size we would tend to overestimate the correlations between variables and would certainly not be providing enough data to characterize the sensory space.

Examination of the correlation matrix is a good starting point for any multivariate analysis as it gives us an idea of the number of underlying dimensions in the sensory space required to characterize product differences. If all the correlations between the sensory attributes are numerically low, this suggests that the attributes are independent, each one quantifying different product differences, and it is unlikely that we will be able to reduce the dimensionality of the sensory space.

5.3.2 Principal Component Analysis

Principal component analysis is the most widely used multivariate technique. It aims to give a low dimensional representation of the sample differences in the multidimensional sensory space by deriving new components or directions in the overall sensory space that capture the main sensory variation in the samples.

The first principal component (PC1) is defined as the direction passing through the centroid (attribute means) of the multivariate sample space which explains as much of the variation in the samples as possible. The second component (PC2) is orthogonal to the first and is in the direction that captures the maximum amount of remaining variation. A graphical representation of the technique for two variables, X1 and X2, is shown in Figure 5.5.

The amount of the total variation explained by each component is measured by its eigenvalue, and the relationship between the derived components (PC1, PC2, PC3, etc.) and the original sensory variables is most easily described by their intercorrelations. For further details of the mathematics behind the technique, see Chatfield and Collins (1981).

Example: apple texture profile

The analysis is carried out on the table of panel mean scores (see Table 5.6). Attributes for which there are no significant differences between the sample means are usually best excluded from any multivariate analysis unless there is a particular requirement to include them.

Table 5.8 Correlation matrix of apple texture variables.

Variables	Crisp	Juicy	Skin separation	Tough	Flesh density	Fibrous	Floury	Pulpy	Fibrous bits	Skin bits
Crisp	**1**	0.38	−0.41	0.93	0.94	0.86	−0.96	−0.59	0.87	0.42
Juicy	0.38	**1**	0.08	0.05	0.07	0.30	−0.51	0.20	0.22	0.20
Skin separation	−0.41	0.08	**1**	−0.39	−0.42	−0.18	0.31	0.53	−0.18	0.44
Tough	0.93	0.05	−0.39	**1**	0.99	0.87	−0.86	−0.68	0.91	0.45
Flesh density	0.94	0.07	−0.42	0.99	**1**	0.85	−0.87	−0.65	0.88	0.47
Fibrous	0.86	0.30	−0.18	0.87	0.85	**1**	−0.89	−0.50	0.94	0.60
Floury	−0.96	−0.51	0.31	−0.86	−0.87	−0.89	**1**	0.42	−0.88	−0.47
Pulpy	−0.59	0.20	0.53	−0.68	−0.65	−0.50	0.42	**1**	−0.57	0.15
Fibrous bits	0.87	0.22	−0.18	0.91	0.88	0.94	−0.88	−0.57	**1**	0.50
Skin bits	0.42	0.20	0.44	0.45	0.47	0.60	−0.47	0.15	0.50	**1**

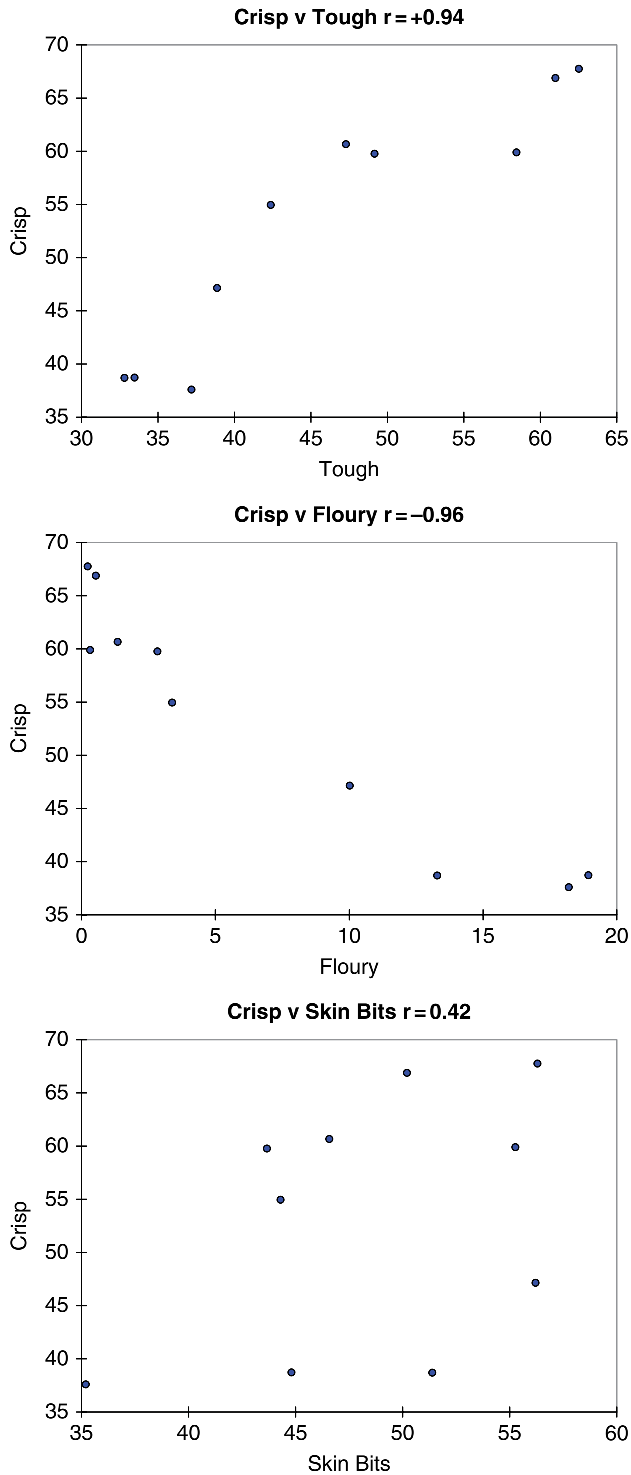


Figure 5.4 Scatter plots of attributes.

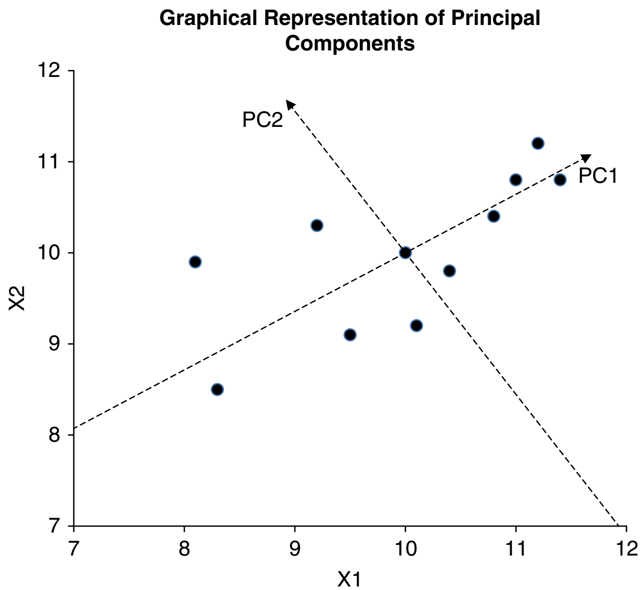


Figure 5.5 Graphical representation of principal components.

Table 5.9 Eigenvalues principal component analysis of apple texture.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Eigenvalue	572.6	79.5	55.0	13.0	5.3	3.8	1.0	0.7	0.1
Variability (%)	78.3	10.9	7.5	1.8	0.7	0.5	0.1	0.1	0.0
Cumulative %	78.3	89.2	96.7	98.5	99.2	99.8	99.9	100.0	100.0

The first piece of output to look at is the table of eigenvalues which gives you the percentage of variation explained by each principal component. Our main interest is in visualizing the sample differences and to achieve this we hope to find a high percentage of the overall variation explained in two or three dimensions. Table 5.9 gives the eigenvalues for each component, which are most easily interpreted by the row of the table which gives the percentage of variation explained. Here we see that 89.2% of the total variation can be explained by the first two principal components. This means that we can get a good picture of the sample differences by plotting the scores of the samples (apple varieties) on the first two principal components. These are plotted in Figure 5.6. Apple varieties that are close together on the plot are perceived to be more similar in texture than apples which are further apart. Apples close to the centre are average in texture relative to the other apples in the data set. Fuji and Braeburn are similar in texture, Royal Gala is different in texture to other varieties, particularly Granny Smith and Top Red. Notice that the spread of the samples is much greater on PC1 than on PC2, reflecting the amount of variation explained in each dimension.

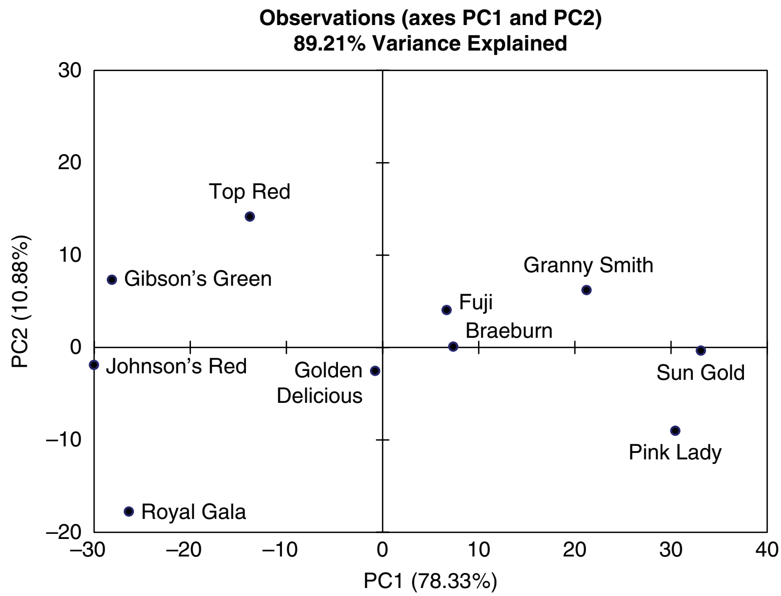


Figure 5.6 Apples texture samples plotted on first two principal components.

Care must be taken in selection of software for PCA as some of the available packages normalize the principal component scales before plotting them on the two components, thus distorting the relative distances between the samples. This can be checked by calculating the variance of the sample scores on each component; if the data are correctly scaled, the variance of scores on a component should equal the eigenvalue of that component.

The correlation plot in Figure 5.7 shows the relationship between the principal components and the sensory variables displayed as vectors representing the correlation of each sensory variable with the principal component axes. Sensory vectors in the same direction will tend to be highly correlated and the directions of the principal component axes are interpreted by their correlations with the sensory variables. Samples with high scores on the first principal component will tend to be crisp, fibrous and with high density of flesh; in contrast, samples with negative scores on the first principal component will tend to be more floury and pulpy and less crisp and fibrous. Samples with positive scores on the second principal component will tend to have more skin separation and skin bits.

It is conventional to draw the vectors from the origin in a positive direction (attribute increasing) but the full vector would extend in both the positive and negative directions.

Another important point in interpreting this plot is that the length of the sensory vector indicates how well the variable projects onto the principal component space.

The sensory variable juicy is not well represented in the plot of the first two principal components but a correlation plot of the first and third dimension

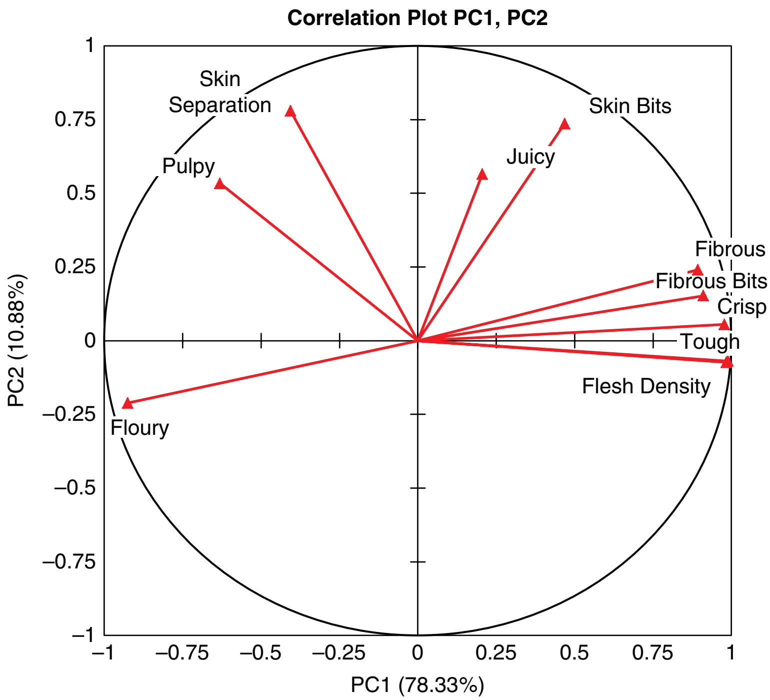


Figure 5.7 Apples texture principal components correlation plot PC1 versus PC2.

(Figure 5.8) indicates that juicy is represented by the third principal component. Investigation of the samples plot in Figure 5.9 shows that Fuji, Braeburn and Granny Smith, which are positioned close together on the first two components (see Figure 5.6), are well separated on the third dimension, Granny Smith being less juicy. Most software packages also offer a biplot which shows samples (points) and variables (vectors) on the same plot. There are several problems with this plot.

- It can be hard to read if there are many variables and samples.
- Some software packages do not automatically give an ideal scaling to enable visualization of both samples and variables.
- There are different biplots that can be drawn depending on whether the focus is on the samples (distance biplot) or the correlation structure of the variables (correlation biplot). The differences are usually slight but it is an issue for consideration.

Remember always that the PCA is a visualization of the sample differences, with a quality of representation measured by the percentage of variation explained, so any statement of difference should be cross-referenced with the pairwise test results from the analysis of variance.

Principal component analysis can be carried out using either the covariance or correlation between the sensory attributes as a basis of the analysis.

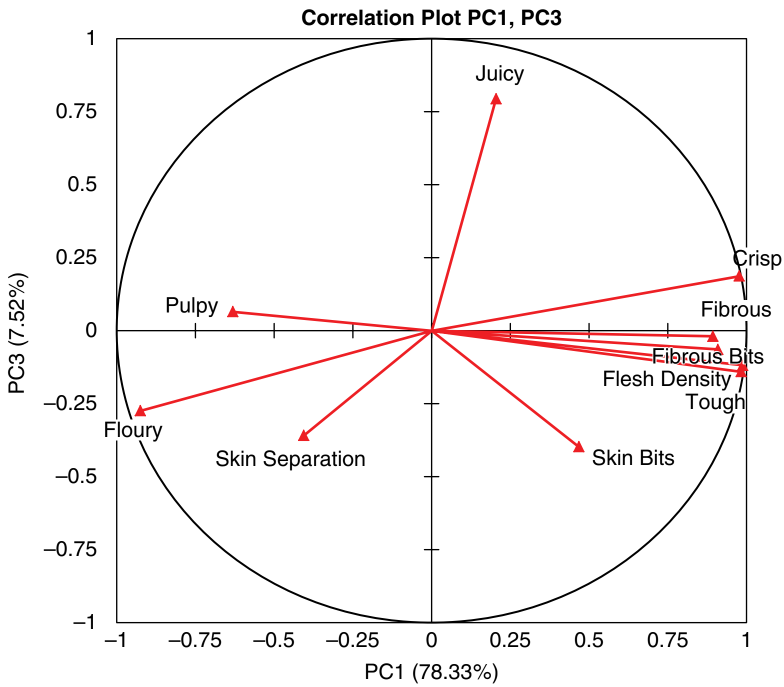


Figure 5.8 Apples texture principal components correlation plot PC1 versus PC3.

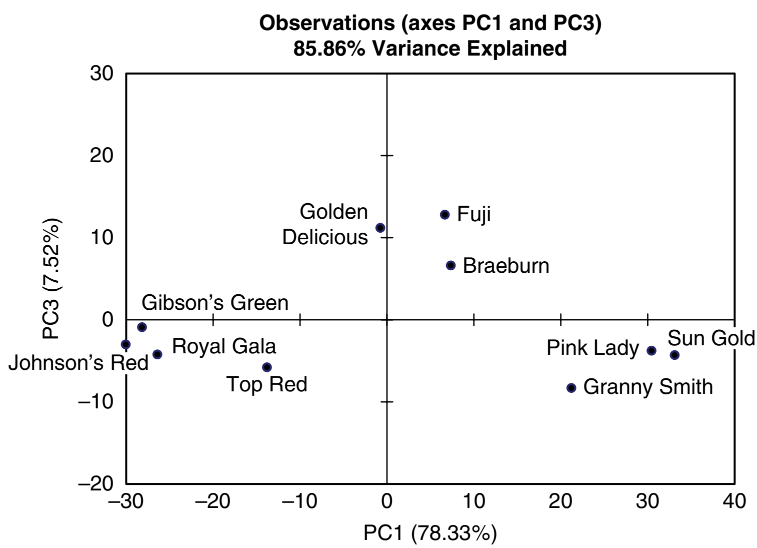


Figure 5.9 Apples texture samples plotted on first and third principal component.

A covariance-based analysis represents the samples in multivariate space using the actual data scales, so attributes with larger variances across the samples will tend to weight heavily on the first few components. If you use the correlation matrix then this is equivalent to standardizing all the sensory scales to the same variance and then representing the multivariate sample space using these standardized scales. If your panel have been trained to use the scales using reference standards then you will most probably want to base your analysis on the covariance matrix, since attributes with higher variance should reflect greater product differences. If the panel are untrained then a correlation-based analysis may be more appropriate.

Using PCA as a data reduction technique, we hope that we can characterize the key product differences in the first two or three principal components which we can then plot. Ideally, we would like to explain at least 70% of the total sensory variation in these plots. For complex profiles with many sensory attributes, principal component mapping is often most successful when carried out separately for each sensory modality. As with correlation analysis, a minimum of five or six samples is required to carry out the analysis.

5.3.3 Canonical Variates Analysis

A multivariate analysis of variance (MANOVA) models product and assessor variation over all the sensory variables simultaneously, allowing for the correlation structure between the variables. Canonical variates analysis (CVA) is based on the MANOVA approach and calculates canonical components which maximize the F statistic of between-sample differences relative to the assessor by sample interaction.

Whereas data for PCA are typically the matrix of panel mean scores, data for CVA are the table of assessor by sample mean scores (with the assessors zero centred).

Canonical variates analysis maps the samples with the objective of giving the 'best' view of the sample differences relative to the assessor by sample variation. As part of the analysis, confidence ellipses for the sample means can be overlaid on the plot to provide a graphical view of the significance of differences between the samples over the sensory space; samples whose confidence ellipses do not overlap are 'significantly' different in their mean positions, taking into account assessor variation. Conventionally, software packages that offer this analysis scale the axes so that the uncertainty regions are represented by circles. A CVA plot of the apple textures is given in Figure 5.10. The plot shows that the apple varieties Pink Lady, Sun Gold and Granny Smith are discriminated from the other varieties; they are crisper and more fibrous. Apple varieties with overlapping circles are not distinguished; for example, Royal Gala and Johnson's Red are not discriminated in texture. The CVA plot on the first two dimensions explains 78.3% of the variety differences relative to the interaction variation.

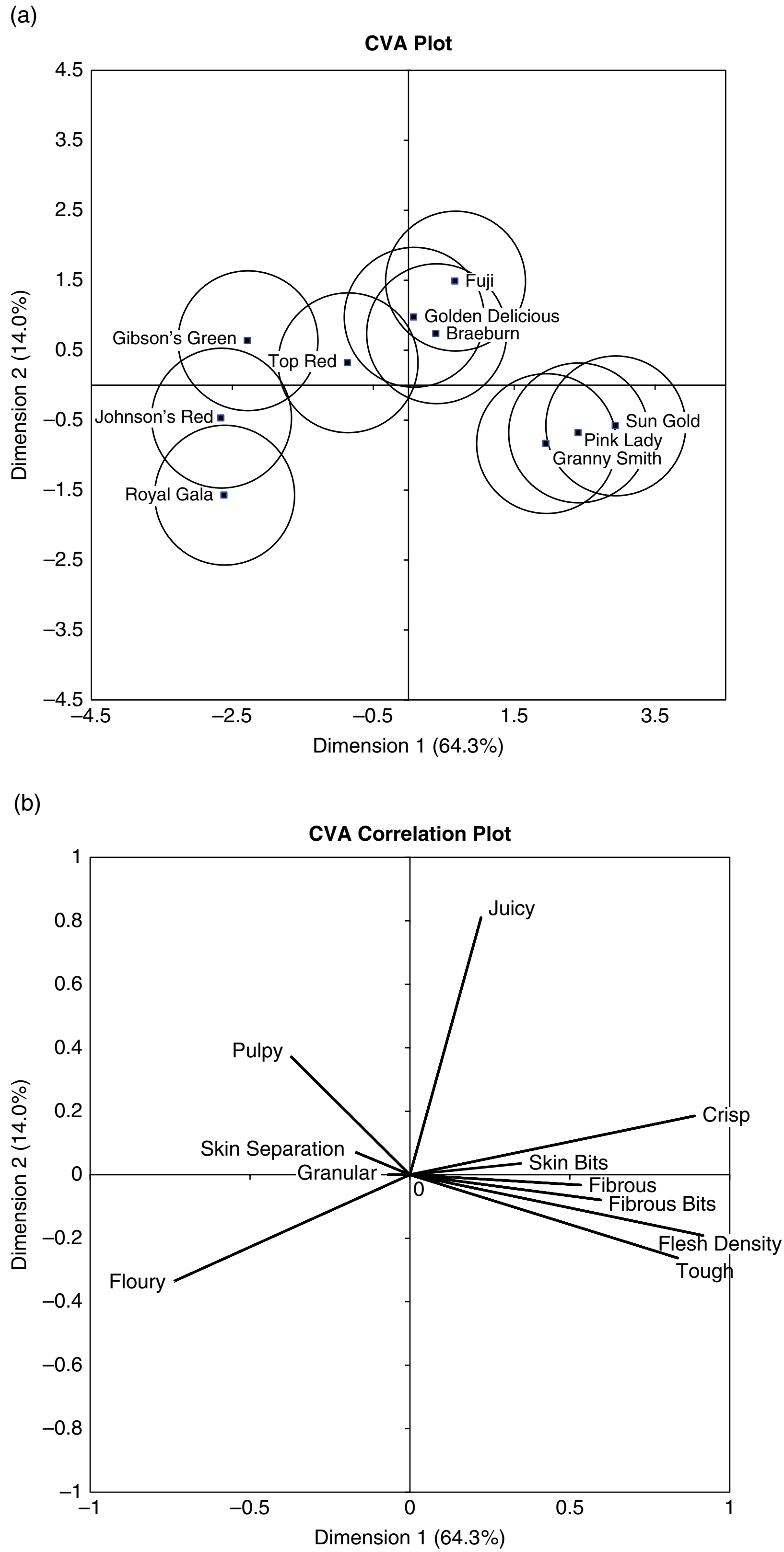


Figure 5.10 (a) Canonical variates apple texture first and second component. (b) Canonical variates loadings plot first and second component.

This analysis can be useful for panel performance, as a byproduct of the analysis is the ability to detect panellists whose scores are not typical. Some further discussion of the application of this technique to panel performance is given in Chapter 4.

5.3.4 Cluster Analysis of Samples

Cluster analysis techniques aim to find groupings of objects, or in our context samples, based on measures of similarity or dissimilarity derived from quantitative or qualitative descriptors of the objects. Many of the techniques available, for example k means clustering, latent class clustering, fuzzy clustering, require a large number of objects to provide stable solutions and are appropriate for clustering consumers to investigate market segmentation.

For sensory data projects, a hierarchical clustering algorithm can be used to investigate grouping of samples based on their sensory characteristics. This can be useful if the objective of the profiling exercise is to determine groupings of samples and also to provide graphical visualization of the samples in situations where the sample space is too multidimensional for PCA to provide an adequate display in two or three dimensions. Hierarchical clustering methods start with all the objects as individuals; the two closest objects are then merged into a cluster and the process is repeated, merging objects into clusters and merging clusters until all objects are in one group. The merging process is displayed graphically in a dendrogram and is used to determine an appropriate number of clusters based on the grouping patterns in the data.

Two options have to be set:

- how similarity or dissimilarity between pairs of objects is to be measured
- how an aggregate measure of similarity/dissimilarity between clusters is to be measured.

For descriptive sensory data, the most commonly used options are to use Euclidean distance (a multidimensional version of Pythagoras' theorem) to measure dissimilarity between objects (samples) and the average of all pairwise distances between objects in the different clusters to measure the distance between clusters; for further details on other agglomeration methods see Everitt (2001). As with PCA, the analyst must decide whether differences between samples should be calculated using the actual data scales (in which case scales with larger sample variance will have more weight in the solution) or by first standardizing the data to remove scaling effects.

Unlike PCA, cluster analysis does not give a unique solution and the decision on the number of clusters is based on the clustering pattern and the analyst's knowledge of the samples tested and the objective of the analysis.

A minimum of eight objects is recommended to carry out this analysis. The number of attributes used will depend on the objective of the analysis (as in PCA, attributes for which there are no significant differences detected are usually excluded from this analysis).

Example: cluster analysis of apples using the texture profile
The analysis is carried out on the table of panel mean scores (see Table 5.6).

Software which offers this analysis will ask you to specify the measure of similarity/dissimilarity and the clustering (agglomeration) criterion. The dendrogram in Figure 5.11 is based on Euclidean distance to measure dissimilarity between pairs of objects and pairwise average linkage to measure distance between clusters.

Examination of the dendrogram suggests three cluster groups.

- Cluster 1 – Gibson's Green, Johnson's Red, Top Red, Royal Gala
- Cluster 2 – Golden Delicious, Fuji, Braeburn,
- Cluster 3 – Granny Smith, Pink Lady, Sun Gold

The interpretation of the group differences is made by calculating the cluster mean scores for each attribute (Table 5.10). Apples in cluster 1 are floury and pulpy in texture with higher skin separation. Apples in cluster 3 are tougher, more fibrous, with higher flesh density. Apples in cluster 2 are more juicy but less tough and fibrous than those on cluster 3.

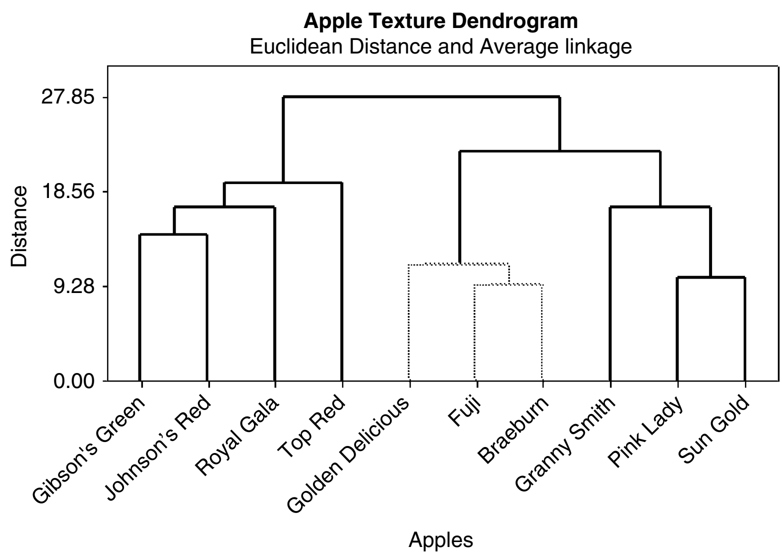


Figure 5.11 Dendrogram cluster analysis for apples.

Table 5.10 Cluster means for apple texture.

Cluster	Crisp	Juicy	Skin separation	Tough	Flesh density	Fibrous	Floury	Pulpy	Fibrous bits	Skin bits
1	40.5	47.1	52.1	35.6	32.1	4.4	15.1	10.2	4.6	46.9
2	58.4	58.2	45.0	46.3	46.3	9.1	2.5	6.9	9.7	44.9
3	64.8	48.0	45.1	60.7	64.8	13.0	0.4	4.0	14.0	53.9

In this case, the cluster analysis gives a very similar interpretation of sample groupings and differences to that found in the PCA. Cluster analysis can be useful to validate the PCA map and will provide a view of sample groupings in situations where PCA is not successful at mapping the samples in two or three dimensions due to the complexity of the samples tested.

5.3.5 Cluster Analysis of Attributes

Cluster analysis can also be applied to the attributes to provide another way of visualizing the correlation structure in the profile. Again, this can be particularly useful when PCA does not provide an adequate representation in two or three dimensions. As we are now clustering variables, we use correlation as a measure of similarity. The dendrogram in Figure 5.12 shows how the attributes cluster based on this similarity measure. The vertical axis labelled Similarity is in fact the average correlation between groups multiplied by 100. We can see that the attributes crisp, tough and flesh density form a group at a high measure of similarity and so these are terms that the panel are scoring in a similar way for the set of samples profiled.

If we think further about the typical correlation structures we see in profile data, we often have variables that are highly negatively correlated because they are measuring sample characteristics which will always be opposites. Using the correlation coefficient as a measure of similarity, these attributes (for example floury and crisp correlation = −0.96) will be seen as most dissimilar; as can be seen in Figure 5.12, they are clearly in different clusters. If we cluster the attributes using the size of the correlation coefficient, ignoring its sign, that is,

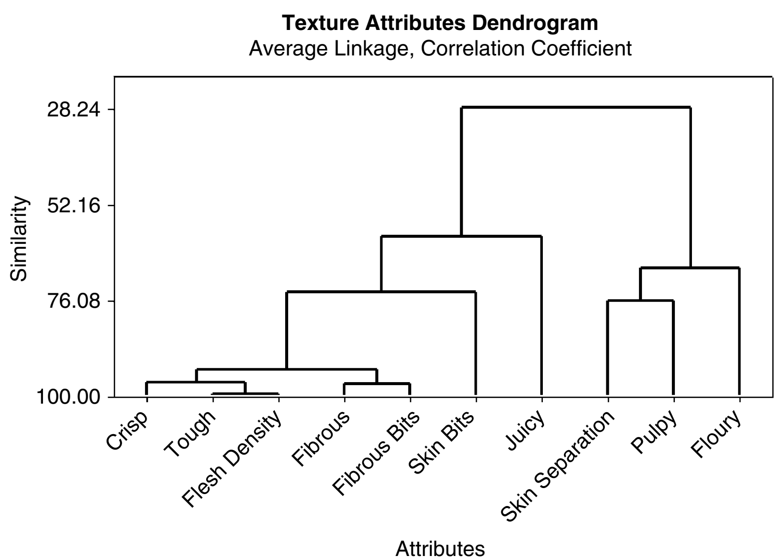


Figure 5.12 Clustering of attributes using correlation coefficient.

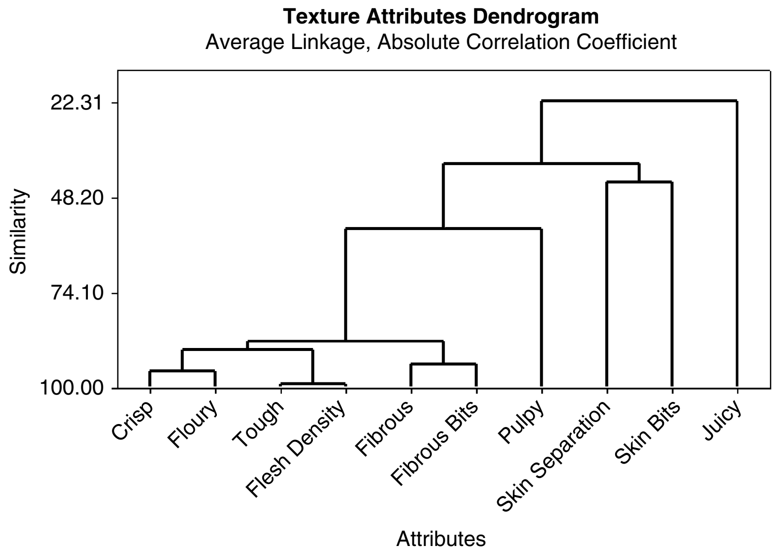


Figure 5.13 Clustering of attributes using absolute value of correlation coefficient.

the absolute value, a different cluster pattern is seen (Figure 5.13). Now floury is clustered with the three attributes crisp, tough and flesh density since it is discriminating the apples in a similar way but opposite in direction (the most floury apple will tend to be the least crisp). This approach can be useful when trying to understand how panellists are using attributes in a new profile, for objectively identifying redundancy in sensory terms or when trying to reduce the number of variables profiled for, say, planning a long-term shelf-life study.

One word of warning, though. Be careful to ensure that the broader conclusions you make about the similarity in attributes take into account any accidental correlations; for example, will skin separation always be associated with pulpy, floury apples (see Figure 5.12) or is this chance due purely to the samples selected for the test?

5.4 Modelling Relationships between Variables

The multivariate analyses we have looked at so far consider the variables as all of equal importance in the sense that we are interested in understanding the inter-relationships between all of them. We now look at a different approach where we consider our variables differently – we are interested in using some of our variables (predictors) to help understand or predict the behaviour of other variables (responses).

In analysing sensory descriptive data, we may want to investigate whether relationships exist between our sensory descriptors and other quantitative information collected on the samples. These could be average liking scores from consumer tests, processing conditions, analytic measures on our samples or, if we are carrying out shelf-life studies, the age of the sample. Calculating correlations will indicate the strength of the linear association between variables but it does not specify *how* they are related; to do this, we need to fit appropriate models to our data.

5.4.1 A Further Thought on Correlation

One very important point to understand is that correlation measures the degree of linear association between variables. A famous data set developed by Anscombe (1973), which has become known as Anscombe's Quartet, illustrates this point. The four pairs of X and Y variables plotted in Figure 5.14 all have the same correlation coefficient $r = 0.82$. It is only the first set X1 v Y1 for which the correlation measure is a really meaningful numerical summary of the data, indicating an underlying linear trend that is not absolutely determined by the data. For the other three sets (X2 v Y2, X3 v Y3, X4 v Y4), the correlation is misleading. Anscombe's example highlights the importance of plotting the data before embarking on any statistical modelling.

5.4.2 Simple Linear Regression

In simple linear regression, the value of a single dependent variable Y_i is predicted as a linear function of the descriptor variable X_i using a linear model of the form $Y_i = a + bX_i + \varepsilon_i$. A model of this form is shown graphically in Figure 5.15. The parameters of the model are a (the intercept), b (the gradient of the line)

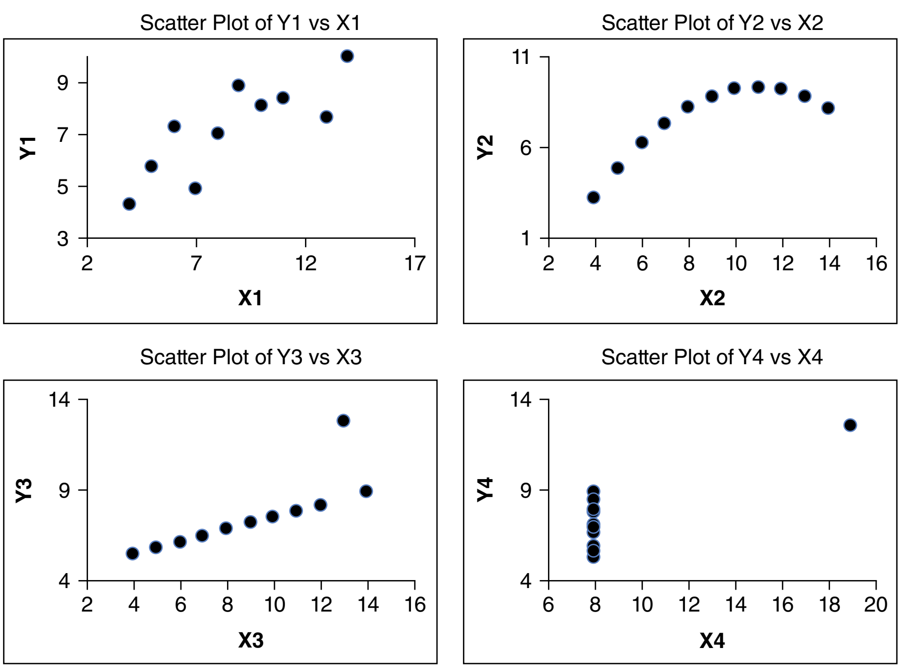


Figure 5.14 Anscombe's Quartet.

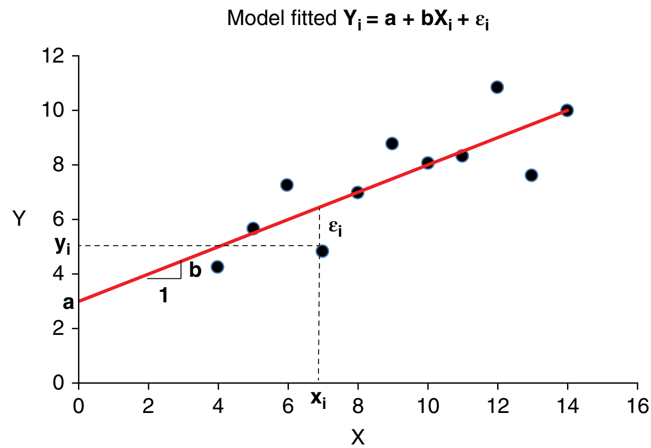


Figure 5.15 Linear regression model

and ε_i, the deviations of individual data points from the fitted line; these are called errors or residuals. For example, Yi might be the weight of a child and Xi the height. There will be a tendency for average weight to increase linearly with height. The model assumes that there is an underlying linear trend in the data but individual children will deviate from the average trend line; for example, a tall, skinny child will weigh less than the average for their height as predicted by the model – this child will have a large negative residual. The parameters of the model are estimated from the data; a is the intercept which is the predicted value of Y when X = 0 (this is rarely of interest in itself) and b is the gradient of the line, which measures the rate of change of Y with X. The residuals ε_i are the deviations between the observed values y_i and the predicted values, and give us an inherent measure of the quality of the fit of the model.

The parameters a and b are estimated using a technique called least squares which minimizes the sum of squares of the residuals $\sum_i \varepsilon_i^2$. This is the mathematical equivalent to finding the best fitting line by eye; we would draw the line that minimized the deviations of the individual data points from our line.

The most widely used summary measure of the quality of fit of the model is measured by a statistic called R^2 or R-squared which measures the proportion of the total variation in the response data that is explained by the regression model fitted. For simple regression on one variable, this is explained graphically in Figure 5.16; the total variation in the data is measured by the total sum of squares:

$$Total\ SS = \sum_i (Y_i - \bar{Y})^2$$

and the residual sum of squares:

$$Residual\ SS = \sum_i \varepsilon_i^2 \text{ and } R^2 = \frac{Total\ SS - Residual\ SS}{Total\ SS}\%.$$

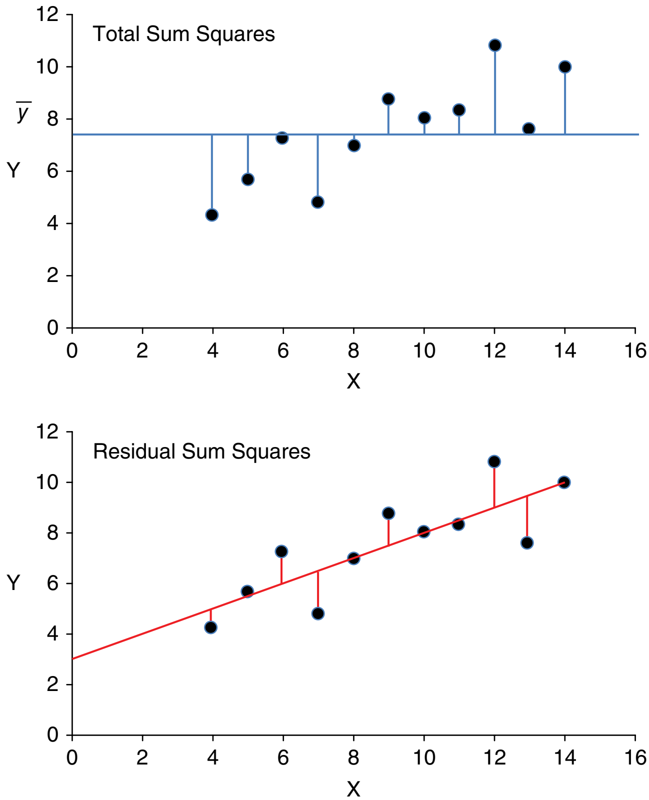


Figure 5.16 Visualization of R^2.

This breakdown of the variation in the data is summarized in an analysis of variance table.

Source	df	Sum of squares	Mean square	F
Regression	1	Reg SS	Reg MS	Reg MS/Resid MS
Residual	n-2	Resid SS	Resid MS	
Total	n-1	Total SS		

The significance of the regression is tested using either an F test to test whether a significant amount of variation has been explained by the model or a t test on the slope parameter b, testing whether the slope is significantly different to zero.

A frequent question is 'How large does R^2 have to be?'; like so many questions in statistics, the answer depends on the application. An analytic chemist using regression to build calibration relations will be looking for very precise models with R^2 values over 95%. A sensory scientist investigating the relations between

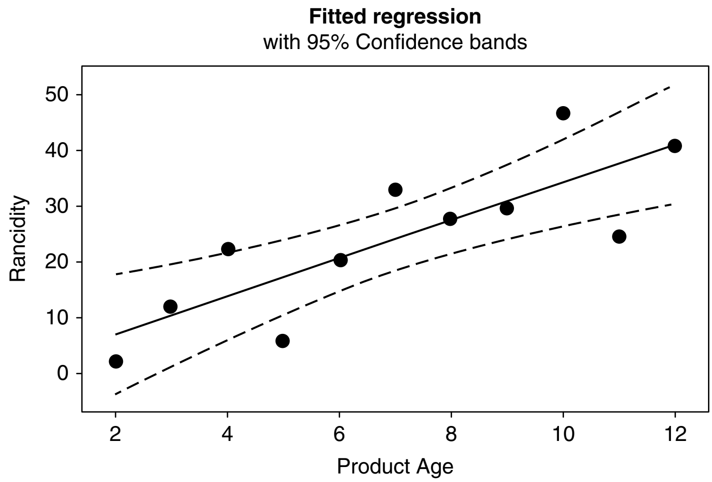


Figure 5.17 Fitted regression with 95% confidence bands.

sensory attributes and average consumer acceptability may find a model with an R^2 around 50% useful, since it indicates that 50% of the variation in average acceptability can be explained by a linear trend of liking with sweetness.

One way to evaluate the quality of the model is to plot the fitted model together with 95% confidence bands for the average prediction; this will provide a visual assessment of the precision of the model fit – an example is given in Figure 5.17. The R^2 for the model fit is 67%. If we use the model to predict the rancidity of a sample aged 6 weeks, we can be 95% confident that the true average rancidity will be in the range 14.6–26.6 units; this may or may not be precise enough depending on the context of the modelling. We certainly have evidence that rancidity increases with sample age but we cannot estimate the relationship with high precision.

Another important part of the modelling task is to assess the adequacy of the fit of the model; are we fitting the appropriate model? Do we have evidence of outliers? In simple regression where we are only modelling one descriptor, this can be done visually but as we move into higher dimensional models, we will need to assess the model fit using residual plots.

If the model fit is adequate the residuals should:

- be relatively small
- be randomly distributed around the fitted model
- have reasonably constant variation
- show no pattern of curvature.

Residuals should be plotted against the predicted values of the response to investigate their behaviour. Outliers are best detected by investigating the residuals on a standardized normal scale (scaled to mean 0 and standard deviation 1). On this scale, we would expect 99% of standardized residuals to be within the range ±2.5, so any observations with standardized residuals outside this range warrant investigation as possible outliers. In particular, any outliers detected at the extremes of the data range may be influencing the model fit and may be indicators that a straight line model is not appropriate for the data.

The interpretation of the analysis is explained in the following example.

Example: predicting liking from apple flavour and texture

A consumer test on the apples indicated a clear segmentation between consumers in their liking of apple texture; about two-thirds consumers preferred crisp hard apples and one-third preferred the softer textured apples. The average liking scores for the first segment are presented in Table 5.11 together with selected sensory flavour and texture attributes.

Correlations and scatter plots of the data indicated that the variable acidic was most highly correlated with liking. Fitting a regression model will indicate how well average liking can be predicted by this sensory attribute.

The fitted equation is Liking = 3.37 + 0.065 × Acidic with $R^2 = 51\%$. We estimate that with every 10 unit increase in acidic on the sensory scale, liking increases by 0.65 hedonic units.

Table 5.12 gives details of the fitted equation. The t test is calculated as:

$$t = \frac{Estimate}{Standard\,Error}$$

and is tested against a t distribution with degrees of freedom equal to the degrees of freedom of the residual (in this case df = 8). The t test tests the null hypothesis that the slope is zero against a two-sided alternative (slope not equal to zero). Both the intercept and the slope are significantly different to zero at $P < 0.05$.

As discussed earlier, the intercept is rarely of interest (it is the predicted liking of an apple with zero acidity on the sensory scale) and we focus on the slope parameter for acidic. We have strong statistical evidence that this parameter is not zero and so can state that acidic is a positive 'driver' of liking. The lower and upper bounds give us a 95% confidence interval for the true value of the parameter, and this begins to give us an indication of the precision of our model fit. A 10-unit increase in acidity may give an increase in liking between 0.13 and 1.17 hedonic units with 95% certainty.

The analysis of variance table (Table 5.13) gives an alternative test for the fit of the model, testing the average variation explained by the model against the average residual variation using the F test. The F test F = Model MS/Residual MS = 8.296 which is significant at $P < 0.021$. Note that for simple regression with only one predictor, the F test and t test will always be equivalent.

The analysis of variance also illustrates the calculation of R^2: R^2 = proportion of total variation explained by the model = Model SS/Total SS = 8.418/16.536 = 51%.

Although the analysis of variance is important in understanding how regression models are fitted and the limitations of such modelling, in itself it does not give any additional information on the quality of fit of the model.

Table 5.11 Consumer average liking and texture/flavour sensory mean scores for apples.

Sample	Crisp	Juicy	Skin separation	Fibrous	Pulpy	Fibrous bits	Skin bits	Sweet	Acidic	Bitter	Watery	Unripe	Average liking
Gibson's Green	38.7	51.0	52.2	4.2	18.1	3.6	51.4	29.4	38.0	12.8	26.9	8.9	**5.1**
Johnson's Red	38.7	47.1	51.8	3.0	7.6	4.9	44.8	49.6	10.3	10.3	15.6	0.1	**4.6**
Golden Delicious	54.9	56.9	40.7	9.4	7.4	9.2	44.3	35.3	23.9	13.0	24.2	3.6	**6.9**
Granny Smith	59.9	48.3	53.5	15.2	5.2	17.0	55.3	20.1	51.9	24.8	19.5	28.4	**6.3**
Pink Lady	66.9	45.4	38.6	10.2	3.7	12.1	50.2	36.5	45.5	12.2	6.2	13.6	**6.0**
Fuji	60.6	62.1	45.9	10.7	6.5	9.6	46.6	36.1	21.9	14.6	19.7	3.7	**4.5**
Top Red	47.1	51.5	61.5	7.0	10.4	6.4	56.2	38.1	14.1	25.4	27.1	0.5	**2.9**
Braeburn	59.7	55.6	48.4	7.1	6.8	10.3	43.7	34.1	30.8	9.6	13.6	5.9	**5.9**
Royal Gala	37.6	38.9	42.8	3.4	4.7	3.4	35.2	41.4	16.6	15.0	9.8	0.2	**3.9**
Sun Gold	67.7	50.4	43.4	13.7	3.2	12.8	56.3	30.6	46.7	14.9	6.5	17.5	**7.1**

Table 5.12 Parameters of fitted model liking versus acidic.

Source	Value	Standard error	t	Pr>|t|	Lower bound (95%)	Upper bound (95%)
Intercept	3.37	0.75	4.49	0.002	1.64	5.09
Acidic	0.07	0.02	2.88	0.021	0.01	0.12

Table 5.13 Analysis of variance table regression of liking on acidic.

Source	DF	Sum of squares	Mean squares	F	Pr>F
Model	1	8.42	8.42	8.30	0.02
Residual	8	8.12	1.01		
Corrected total	9	16.54			

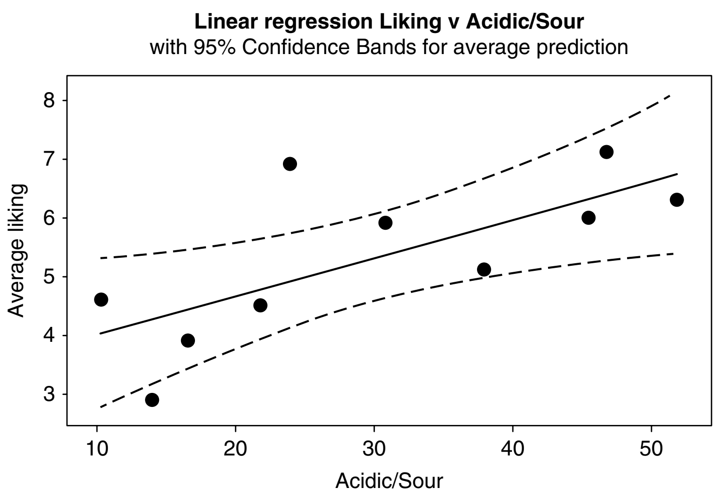


Figure 5.18 Fitted regression liking versus acidic with 95% confidence bands.

A plot of the fitted model with 95% confidence bands is given in Figure 5.18. If we use the model to predict the average liking of an apple with an acidic score of 20 on the sensory scale, we can be 95% confident that the average liking will be between 3.8 and 5.6 hedonic units. This is not very precise! I think we would be very surprised if we could predict liking adequately from one sensory attribute. However, we can say that acidity is a positive linear driver of liking for this segment of consumers.

The residuals are tabulated in Table 5.14. Residual = Observed Liking-Predicted Liking. The smaller the residuals, the higher R^2 will be and the better the fit of the model. Golden Delicious is the least well predicted by the model

Table 5.14 Residuals for regression of liking on acidic.

Observation	Acidic	Average liking	Predicted liking	Residual	Standardized residual
Gibson's Green	37.97	5.1	5.84	−0.74	−0.74
Johnson's Red	10.33	4.6	4.04	0.56	0.56
Golden Delicious	23.94	6.9	4.93	1.97	1.96
Granny Smith	51.85	6.3	6.75	−0.45	−0.44
Pink Lady	45.45	6.0	6.33	−0.33	−0.33
Fuji	21.85	4.5	4.79	−0.29	−0.29
Top Red	14.09	2.9	4.29	−1.39	−1.38
Braeburn	30.82	5.9	5.38	0.52	0.52
Royal Gala	16.64	3.9	4.45	−0.55	−0.55
Sun Gold	46.73	7.1	6.41	0.69	0.68

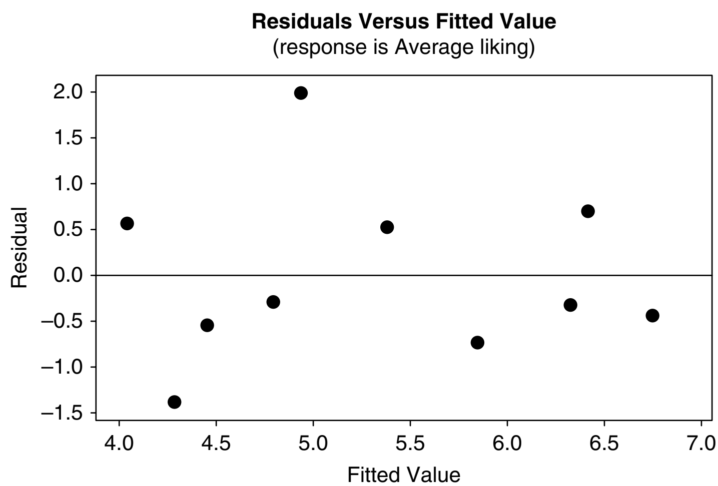


Figure 5.19 Residual plot regression of liking on acidic.

with a residual of +1.97 – its average liking is much higher than would be predicted by its acidic score.

The standardized residuals are rescaled to a standard normal scale with mean zero and unit standard deviation. These can be used to identify outliers; observations with standardized residuals outside the range ±2.5 should certainly be investigated for errors in the data or inappropriate inclusion in the modelling data set. Here we have no indication of outliers.

We should also check that the model we are fitting is appropriate to the data; this can be assessed visually by plotting residuals against predicted/fitted values. We hope to see a random scatter of residuals – any curvature in the plot will indicate that a straight line model is not appropriate. A plot of the residuals for this model is given in Figure 5.19 and suggests an adequate model.

5.4.3 Multiple Regression

The modelling can be extended to include more than one descriptor variable. These can be quadratic terms to model curvature (as opposed to linear relationships), for example: $\text{Liking} = b_0 + b_1\text{Sweet} + b_2\text{Sweet}^2$ will fit a quadratic curve to allow estimation of an optimum sweetness level. A typical fitted data curve is shown in Figure 5.20.

Or we may want to fit more descriptors, for example: $\text{Liking} = b_0 + b_1\text{Sweet} + b_2 \text{Smooth} + b_3\text{Milky} + \cdots$.

With a typical profile of 20–30 sensory attributes, how do we decide which ones to fit? We do not want to use descriptors which are highly correlated with each other as they are unlikely to improve the quality of the model fit but we want to add variables that will improve the prediction of our response variable. Statistical software comes to our aid here as it offers a range of algorithms which are designed to build a model which maximizes the fit using a limited number of descriptors. The most widely used is stepwise selection which adds terms into the model sequentially until no significant improvement in the fit can be made. At each step it also re-evaluates the contribution of variables already in the model and drops any variables that have become redundant due to their correlation with the latest variable added.

Example: stepwise regression of liking on apple texture and flavour

We run a stepwise selection procedure, allowing the software to select from the 12 sensory descriptor variables the 'best' model to predict liking. We set the criterion to include any descriptors which are significant at $P = 0.1$ (this is an acceptable criterion when dealing with consumer test data).

The algorithm selects two descriptors: acidic and skin separation. Fitting these two variables gives an R^2 of 67%, a reasonable improvement on the value from the single variable regression on acidic of $R^2 = 51\%$.

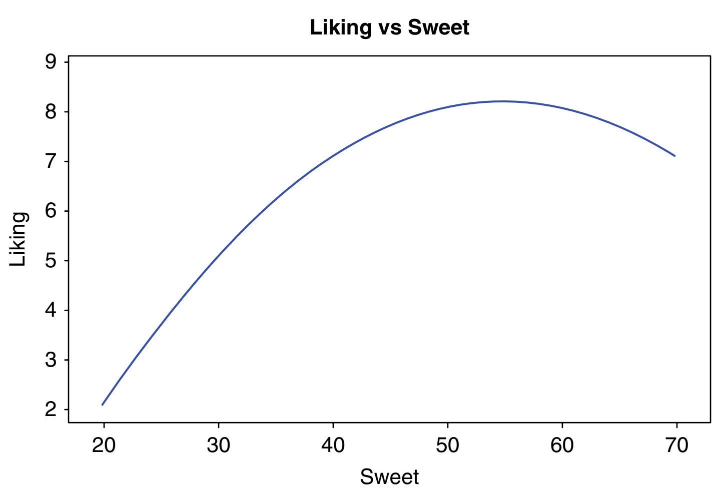


Figure 5.20 Liking versus sweetness.

Table 5.15 Fitted model stepwise selection of liking on sensory descriptors.

Source	Value	Standard error	t	Pr>|t|
Intercept	7.46	2.33	3.20	0.01
Skin separation	–0.08	0.04	–1.94	0.09
Acidic	0.06	0.02	2.73	0.03

Table 5.16 Fitted model prediction of liking from crisp and acidic.

Source	Value	Standard error	t	Pr>|t|
Intercept	1.97	1.60	1.23	0.26
Acidic	0.05	0.03	1.57	0.16
Crisp	0.04	0.04	0.99	0.36

The fitted equation (Table 5.15) indicates that both parameters are significant at $P < 0.1$.

Comparing the fitted equation with the single variable equation (see Table 5.12), we see another issue. The value of the acidic parameter changes from 0.065 to 0.056 when we include skin separation in the model. This is a consequence of the intercorrelations between the descriptor variables which can lead to traps for the unwary in the interpretation of the model. For example, if we fit crisp and acidic as descriptors, we find that neither parameter is significant (Table 5.16) and yet $R^2 = 57\%$. This does not make sense – we know that acidic is a significant predictor and the R^2 value is suggesting reasonable predictive power. The reason for the apparent anomaly is as follows. The t tests for each model term are testing the significance of adding parameters conditional on all other parameters already in the model. Because crisp and acidic are reasonably highly correlated ($r = 0.64$), adding crisp into a model that already contains acidic does not give a significant improvement and vice versa.

Using a stepwise algorithm will guard against these problems as crisp will not be added as a predictor to a model that already contains acidic. We will always be limited to the number of sensory variables that can be included in a classic regression model and this is a fundamental weakness of the technique as we would like to be able to build a model using both the variables crisp and acidic as we believe they may both influence consumer acceptability and we would like to consider the combined effect of all the variables in our profile on consumer acceptability.

Multiple regression is a useful technique for modelling data from designed experiments where the correlations between the descriptor variables are controlled by the experimental design or in situations where we are modelling a limited number of reasonably independent descriptors. However, for sensory descriptive data where we almost always have more variables than samples and these variables are themselves highly correlated, we need to employ more sophisticated techniques such as principal component regression or partial least squares regression.

5.4.4 Principal Component Regression

Principal component regression (PCR) combines the techniques of principal component analysis and multiple regression. The sensory profile data are first transformed to principal components and a judgement made as to the number of components required to explain the main structure of the sensory data. One of the properties of the principal components is that they are uncorrelated and so the problems we encountered in using regression with the correlated sensory variables disappear. The principal component scores are then used as input as descriptor variables (X) into a stepwise multiple regression to predict the response. Curvature can be considered by including quadratic terms. The fitted model will be in terms of the principal components but as each component is a linear combination of the sensory variables, the final equation can be back transformed to be expressed in terms of the sensory variables.

Example: principal component regression liking versus apple flavour and texture data

The first step in the analysis is to carry out a principal component analysis on the 12 sensory variables in Table 5.11. The eigenvalue analysis (Table 5.17) indicates that four principal components explain 95% of the sensory variation. The scores of the samples on these four components are then used as predictors in the regression. Stepwise regression selects the first and fourth principal component to predict average liking (Table 5.18), giving a model $R^2 = 65\%$. The equation of the fitted model is Liking = 5.32 + 0.046 PC1 −1.56 PC4 and as PC1 and PC4 are linear combinations of the original sensory

Table 5.17 Principal component analysis variance explained.

	PC1	PC2	PC3	PC4
Variability %	58.9	18.4	11.4	5.9
Cumulative %	58.9	77.3	88.8	94.7

Table 5.18 Stepwise selection principal component regression.

No. of variables	Variable added	R^2
1	PC1	53%
2	PC4	65%

variables, the prediction equation is dependent on the sensory mean scores of the samples across all sensory variables.

The principal components are ordered by the amount of sensory variation in the samples that they explain. In the example above, although the second principal component explains 18% of the sensory variation in the apples, it is not selected as a predictor of liking. The fourth component which only explains 5.9% of the sensory variation is, however, selected as a predictor variable. This may happen as there is no guarantee that the main sensory differences in the samples detected by the panel will be the main predictors of the sample acceptability and this is a downside of the approach.

However, the principal component approach is easily extended to include square and cross-product terms of the components to allow for non-linear response patterns. These ideas form the basis of external preference mapping techniques, allowing the researcher to investigate regions of the sensory space where liking may be optimal.

5.4.5 Partial Least Squares Regression

In principal component regression, the principal components are derived independently of the response variable as directions in the sensory space that explain maximal sensory variation. Partial least squares (PLS) regression is a technique which takes the middle ground between principal component regression and multiple regression. The mathematics behind the technique is complex but an intuitive understanding of the method is probably all that is required.

At the first stage, the technique finds directions (PLS components) in the sensory space which both explain sensory variation in the products *and* at the same time are correlated with the response variable. As such, the direction of the first PLS component(Figure 5.21) will be close to the first principal component of the sensory space (i.e. the angle θ will be small) but should also correlate with the response variable Y; further components are extracted using the same criteria. The second stage of the analysis uses multiple regression to predict the response as a linear function of the PLS components.

The method also allows simultaneous prediction of multiple response variables. In our context, these may be analytic measures on the samples or average consumer acceptability for different groups of consumers. The technique provides visualizations of the correlation structure of the sensory variables (X), the correlation between the responses (Y) and the intercorrelation between the variables in X and Y.

The number of PLS components to be included in the model is determined using cross-validation techniques which guard against the problem of overfitting. The predictive ability of the model is assessed with just one component, then the model with two components, etc. When the predictive ability of the model no longer improves by adding in extra components then the optimal model is found. The predictive ability is measured by a cross-validation statistic called Q^2.

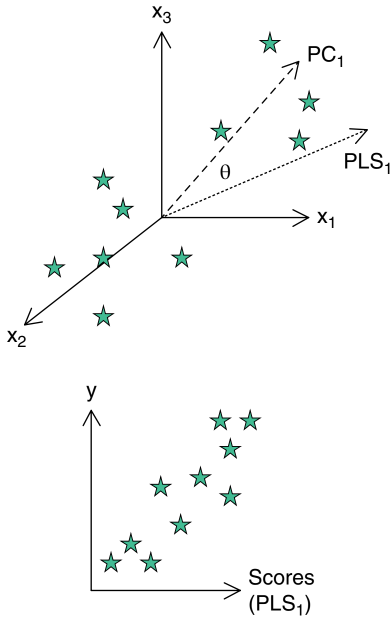


Figure 5.21 Geometric illustration of partial least squares.

For more technical details on this and other regression techniques and their applications in sensory science, see Naes et al. (2010).

Example: partial least squares regression of liking versus apple flavour and texture data
Using cross-validation, a two-component model was selected to predict liking from the sensory profile data. The sensory variables were then screened and those which did not contribute to the prediction of liking were excluded. The correlation plot (Figure 5.22) indicates that liking is positively correlated with crisp and acidic and negatively correlated with skin separation. The multiple regression model predicting liking from the first two PLS components (t1 and t2) has an $R^2 = 78\%$ and the fitted equation is presented in terms of the original sensory variables (Table 5.19). Figure 5.23 shows the standardized coefficients, which are rescaled taking into account the differing variation in each sensory scale; this gives an easy-to-interpret visualization of the relative size and direction of each sensory variable as a 'driver' of liking. Acidic and crisp are the strongest positive drivers, the higher the better, and skin separation and bitter are negative drivers, the lower the better. The model fitted is linear in all drivers and residual plots should be examined to look for any curvature in the response that has not been explained by the model.

Figure 5.24 plots standardized residuals versus predicted values. From this, we can see no evidence of outliers which would have standardized residuals outside the range ±2.5, or patterns of curvature in the data, so we can accept the model fit. Notice that the R^2 value is higher than that obtained using classic

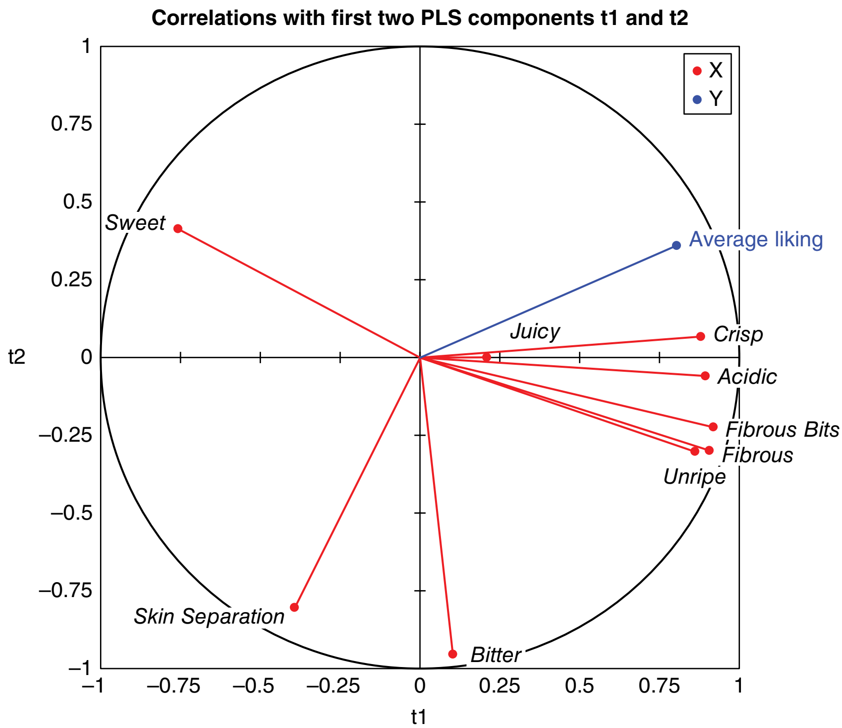


Figure 5.22 Correlation plot of first two PLS components.

Table 5.19 Partial least squares regression equation parameters.

Variable	Parameter
Intercept	6.751
Crisp	0.016
Juicy	0.011
Skin separation	–0.056
Fibrous	0.034
Fibrous bits	0.044
Sweet	–0.015
Acidic	0.018
Bitter	–0.070
Unripe	0.021

multiple regression on the individual sensory measures, 78% versus 67% for the stepwise multiple regression model fitting only acidic and skin separation. The PLS approach has exploited the correlation between the sensory descriptors to derive the PLS components whereas the classic multiple regression was restricted by the correlation structure in the data.

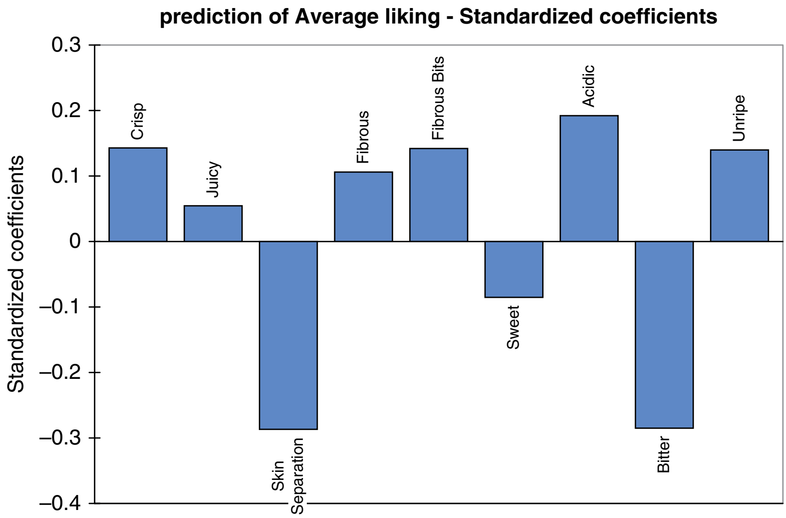


Figure 5.23 Standardized coefficients PLS regression.

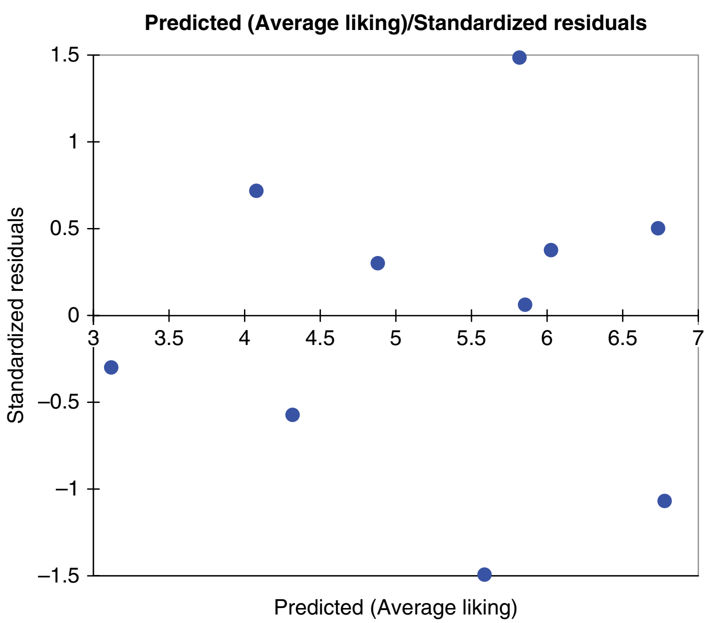


Figure 5.24 Residual plot partial least squares regression.

5.4.6 Other Approaches

Recent software developments have made available interactive graphical tools for building path models which allow the researcher to hypothesize a possible hierarchical causal path to a response. There are two approaches: structural equation modelling, which models the covariance structure but requires large

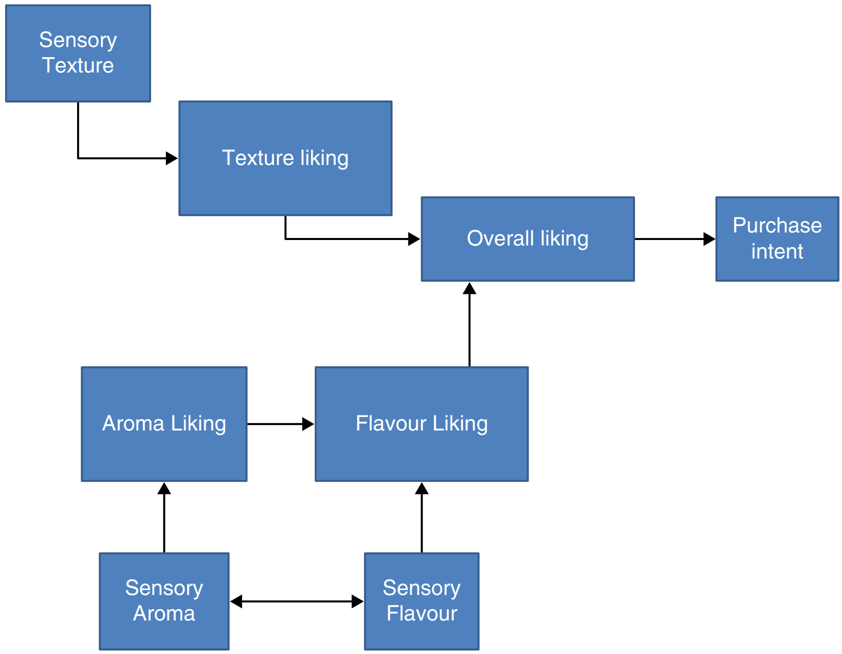


Figure 5.25 Hypothetical path model for purchase intent.

data sets with many observations, and path PLS modelling which uses a component-based approach and is more suited to consumer test data where the number of observations may be limited by the number of products tested. For example, we might propose the causal model in Figure 5.25 to predict purchase intent. If we have sensory, analytic and consumer data on an appropriate set of products, we can test the model and validate the proposed links. These models are beginning to be applied in the sensory field and may have advantages in demonstrating causal links in the data. An example is given in Tenenhaus et al. (2005).

5.5 Summary

Statistical analysis plays a key role in understanding and presenting the results collected by descriptive profiling methods. However well trained a panel is, there will always be variation in the data collected due to variation in perceptions within and between assessors, product variation and scaling and scoring variation. For this reason, it is critically important that the assessment of sample differences is made taking into account these variations using appropriately constructed statistical tests. Multivariate methods provide powerful graphical tools to visualize overall sample differences across the sensory profile. Modern software packages bring these analysis tools within the reach of all sensory professionals.

5.6 Future Developments

Work has recently been published on the MAM model, which is an analysis of variance mixed model which separates the assessor by sample interaction into components due to assessor scaling effects and cross-over interactions. Currently, this is being implemented as part of panel performance assessment but may lead to more precise estimation of product differences (Peltier et al. 2014). Multiple factor analysis, a method for combining tables of data on products which may be of different types, is increasingly being used to characterize products over a range of different data (sensory, instrumental, consumer) and so understand the synergies between the information held in the different tables (Escofier & Pagès 1994).

References

Anscombe, F.J. (1973) Graphs in statistical analysis. *American Statistician*, **27**, 17–21.

Chatfield, C. & Collins, A.J. (1981) *Introduction to Multivariate Analysis*. London: Chapman & Hall.

Escofier, B. & Pagès, J. (1994) Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis*, **18**, 121–140.

Everitt, B.S., Landau, S. & Leese, M. (2001) *Cluster Analysis*. Oxford: Wiley-Blackwell.

Naes, T., Brockhoff, P.B. & Tomic, O. (2010) *Statistics for Sensory & Consumer Science*. Oxford: Wiley-Blackwell.

Peltier, C., Brockhoff, P.B., Visalli, M. & Schlich, P. (2014) The MAM-CAP table: a new tool for monitoring panel performances. *Food Quality & Preference*, **32**, 24–27.

Searle, S.R., Casella, G. & McCulloch, C. (2006) *Variance Components*. Oxford: Wiley-Blackwell.

Tenenhaus, M., Pagès, J., Ambroisine, L. & Guinot, C. (2005) PLS methodology to study relationships between hedonic judgements and product characteristics. *Food Quality & Preference*, **16**, 316–325.

Appendix 5.1 Mixed Models

$$Y_{ijk} = Intercept + Sample_i + Assessor_j + Sample \times Assessor_{ij} + Error_{ijk}$$

where:

$Sample_i$ (i = 1 to p) measures the sample effects
$Assessor_j$ (j = 1 to q) measures the assessor effects
$Sample \times Assessor_{ij}$ measure the interaction effects

Source	df	Sum of squares	Mean square
Sample (S)	p-1	SS(S)	MS(S)
Assessor (A)	q-1	SS(A)	MS(A)
Sample × Assessor (S × A)	(p-1)(q-1)	SS(S × A)	MS(S × A)
Error	pq(r-1)	SS(Error)	MS(Error)
Total	pqr-1	SS(Total)	

Mixed model parameterization gives:

Sample effects S_1, S_2, S_3 ⋯are fixed model effects.

Assessor effects are randomly distributed normal variables with mean 0 and variance σ^2_A.

Sample × Assessor effects are randomly distributed normal variables with mean 0 and variance σ^2_{SA}.

$Error_{ijk}$ are randomly distributed normal variables with mean 0 and variance σ^2.

Source	Expected mean square
Sample (S)	$\frac{qr}{p-1}\sum S_i^2 + r\sigma^2_{SA} + \sigma^2$
Assessor (A)	$pr\sigma^2_A + r\sigma^2_{SA} + \sigma^2$
Sample × Assessor (S × A)	$r\sigma^2_{SA} + \sigma^2$
Error	σ^2

So to test the null hypotheses:

H_0: No sample differences – all $S_i = 0$

We test F = Mean square(S)/Mean square (S × A). If sample effects are negligible then F will be close to 1. A large F with P <0.05 will indicate significant differences in sample means.

H_0: No assessor differences – $\sigma^2_A = 0$

We test F = Mean square(A)/Mean square (S × A). If assessor variation is negligible then F will be close to 1. A large F with P <0.05 will indicate significant variation between assessors in their average scoring level.

H_0: No sample by assessor interaction – $\sigma^2_{SA} = 0$

We test F = Mean square (S × A)/Mean square (Error). If sample by assessor interaction variation is negligible then F will be close to 1. A large F with $P < 0.05$ will indicate significant assessor by sample interaction variation.

This is the mixed model formulation applied in section 5.2.2 (termed the unrestricted mixed model formulation). There is an alternative formulation (termed the restricted formulation) which gives the same F tests for samples but tests for assessor variation using the ratio Assessor MS/Error MS. There is no consensus among statisticians as to which of these formulations is better. For more details see Searle et al. (2006).

SECTION 2
Techniques

CHAPTER 6
Consensus Methods for Descriptive Analysis

Edgar Chambers IV

6.1 Introduction

Consensus refers to joint decision making by a group. Thus, consensus methods of sensory descriptive analysis refer to those methods that use some sort of consensus procedure for determining sensory characteristics, intensities, or both. Consensus is differentiated from individual averaging or interpretation of individually determined attributes by the simple concept that agreement must be reached. That concept, where a group must discuss and reach a group decision on attributes, intensities, or other descriptive characteristic, that is, a group consensus measure, is different from individual evaluations used to calculate mean values, commonly known as group averaging. Consensus methods also differ from descriptive sensory procedures that allow individuals to use any words for description and then determine potential similarities in meaning using statistical multivariate methods such as Procrustes analysis.

Consensus methods of sensory evaluation date to the first published descriptive method, the 'flavor profile'*, as defined by scientists at Arthur D. Little Inc. (Cairncross & Sjöström 1950; Caul 1957). Those authors defined a process whereby a panel was extensively trained to determine flavour properties and able to make decisions about the attributes and intensities as they tasted. Later, the 'texture profile' method, based on concepts of the flavor profile, was developed that used a similar consensus approach initially (Brandt et al. 1963). Numerous studies were published in those early days that used consensus methods on a variety of food products (Cairncross & Sjöström 1948; Caul & Sjöström 1951; Kurtzman & Sjöström 1964) and other products such as pharmaceuticals (Caul & Rockwood 1953) and the odor of motor oil (Fiero 1958). Over many years, studies using

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

consensus methods have been published in a variety of journals and other publications (Berry et al. 1980; Caul & Vaden 1972; Chang & Chambers 1992; Prell & Sawyer 1988; Sanchez & Chambers 2015; Suwonsichon et al. 2012; Vara-Ubol et al. 2006).

In recent years consensus methods have been used for decision making in other fields of study. For example, Murphy et al. (1998) discuss the use, desirable aspects and pitfalls of using consensus methods to help make medical decisions. Not surprisingly, many of the advantages and disadvantages of using consensus for making medical decisions are the same as those suggested for consensus sensory methods.

Various psychological and sociological aspects of consensus judgements have been studied for many years. In a study of group dynamics with college students, Asch (1951) found that although students exhibited a high degree of independence there was still a trend to want to agree with the group. Hicks et al. (1966) reaffirmed this affect for soldiers participating in a made-up group with no prior exposure. However, Festinger (1954) indicated this was true primarily when there was no physical method to validate individual beliefs. That suggests that the potential for this bias may be less in sensory studies where panellists are actually experiencing a psychophysical phenomenon. Discussion was shown to be helpful in a study by Henry (1993) who reported that after discussion, group members were able to determine which members were most probably 'correct' in various assessments and changed their responses based on that discussion, a positive move that improved decision making. Brooks and Koretsky (2011) noted that discussion also resulted in learning that enhanced future abilities of the individuals in the group.

Recent studies also have shown that trained professionals tend to improve their decision making when they use group discussion to come to consensus. For example, Gabel et al. (2010) found that medical outcomes of dementia diagnoses were improved dramatically using group discussion and consensus compared to multiple individual diagnoses, particularly when there was disagreement on the diagnoses or details of treatment. Lin et al. (2014) showed that tourism experts were better able to forecast future tourism growth in various regions using group decision rather than averaged individual estimates. In addition, those authors reported that no bias existed in the group tasks compared to individual surveys. Song (2009) reported high levels of trust among group decision makers with little expectation that they would be 'paid back' (i.e. I helped you now you help me) by their colleagues. Those studies suggest that groups can work well together to produce information that is unbiased and potentially more actionable than individual averaged assessment.

Consensus decisions have been compared to 'staticized' decisions in psychology and other fields as well. 'Staticized' refers to a group where the participants make independent judgements and those judgements are then analysed statistically to come up with a group 'mean' or decision. Stevens et al. (2001) compiled a list of studies comparing various consensus and staticized techniques for various tasks such as estimation, forecasting, ranking, judgement, idea generation, etc. In that list, the staticized decision was never an improvement over the consensus methods. However, Murphy et al. (1998) indicated that when little depth of analysis is required, staticized groups work better and perhaps more efficiently than consensus groups, but in complex situations interaction and consensus methods may work better. Thus, multiple research projects in multiple fields of study suggest that consensus methods have value, perhaps especially when the problem is such that it needs in-depth understanding for decision making.

6.2 Method Overview

Consensus sensory methods may be full or ‘all-inclusive’, determining both the attributes and intensities by consensus, or they may be individualized where panellists only determine attributes by consensus for further rating or where the attributes are determined and the group must reach consensus on the intensity of those attributes. The full consensus method is most similar to the concepts espoused by the flavor profile method; the use of consensus-driven attribute development is used at some level in many trained descriptive methods such as the Spectrum™ or quantitative descriptive analysis (QDA) methods, and determination of consensus intensity measures on predetermined attributes is most commonly noted in quality control where a group determines if a product or process is ‘in or ‘out’.

Common to all consensus descriptive methods is the idea that individuals evaluate products and then discuss the attributes and/or intensities to determine an agreed-on description of the product in question. Key to this concept is that the panel must be trained to use the consensus approach as a way to come to agreement rather than simply ‘average’ individual information. This process is aided by the use of references (i.e. products, chemicals, pictures) that are determined during the project orientation phase and may be modified by agreement during the testing phase. A process for full consensus (Figure 6.1) after the panel has been selected is as follows.

1 A panel is trained to (a) describe and evaluate attributes, (b) determine intensities of attributes, (c) evaluate other components as needed (e.g. order of appearance of attributes or wholistic concepts such as amplitude, balance, blended), (d) learn to discuss differences in a manner that provides unconditional positive regard for the beliefs and opinions of others. This training often, though not always, is rather general in nature and is designed to help the panel understand the process rather than particular aspects of a product category. The training for many consensus methods is extensive, with more than 100 hours of training over a 4–12-month period. Generally, training starts with products that are easy to differentiate

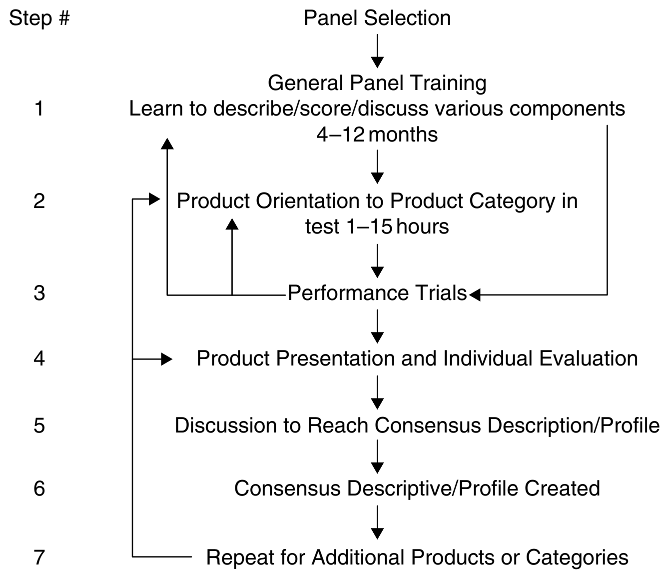


Figure 6.1 Consensus descriptive panel training flowchart.

and describe and then moves to more complex and difficult to describe products. For example, a first set of products during training may be a simple fruit sauce with and without sugar and other seasonings added. This simple exercise helps panellists understand that differences may exist and allows them to experience simple differences that they can easily describe using common terms or ingredient names. The descriptors for this simple exercise might include the fruit, a variety name, ripe, green, cooked, sweet, sour, various spices that have been added, smooth and pulpy. A complex set of samples later in the training may be a set of butter cakes where the list of attributes contains few ingredient names and may have upwards of 50 terms, including common terms such as eggy, sweet and firm, and terms that are less obvious such as butyric, sulfur, popcorn and dense. During training, descriptive comparative tests are often conducted with individuals and with the panel as a whole to ensure that training is progressing properly and to determine remedial training needs. This is similar to the training procedures described for the development of scales and lexicons by Muñoz and Civille (1998).

2 Additional orientation and training is provided specific to products in the category of interest. References are determined by product experts, panellists or both to help guide both the identification and intensity measurement of characteristics. Those references usually are chosen from multiple options available in the literature, ones that have been used in other studies or the personal insights of experts or panellists. Intensities of those references usually are determined by the panel or experts and set to allow

panellists to recognize both attributes and intensities within the range of products to be evaluated. The defined category may be broad, such as dairy products, or may be more specific, such as cheddar cheese, as defined by the panel managers and other scientific and marketing staff.

3 Trials are conducted to ensure the panellists (individually) and the panel (consensus) are able to repeatedly evaluate and differentiate samples.

4 The panel is provided a single product according to a predetermined experimental plan and each assessor evaluates that product.

5 At the end of the individual evaluation, the panel leader, who is usually a member of the panel, leads a discussion to come to consensus on the attributes, intensities and any other aspects (such as the order of the attributes) that the panel is expected to convey in their product description. This process may take minutes or hours depending on the training of the panel, the complexity of the product and the skill of the panel leader. Usually, during this process, products and references are retested (and new samples of the same product presented if needed), and panellists may discuss why variations, if any, in panel perceptions have been found. Additional attributes are added as needed (with appropriate references if the panel leader and panel believe it to be necessary). The process continues until the panel reaches consensus (usually) or until the panel determines that consensus is not possible and notes about the differences are added to the sensory 'profile'.

6 Once the panel has reached consensus on the product, a final sensory descriptive 'profile' is created in a form that is useful to researchers.

7 This process is repeated with additional products until all products have been completed.

Various consensus methods are in use. The flavor and texture profiles as originally published were both consensus methods, but were limited to single sensory modalities. More recent work has adapted them for use for multiple sensory phenomena. A recent industrial adaptation harnesses the ability to track order of appearance of attributes in consensus studies to determine what textures and flavours are occurring simultaneously. Similarly, consensus methods are being used for time intensity studies to overcome the problems with averaging data from individual curves. In all cases, the number of scale points used can vary and is up to the researcher. The original four-point (flavor profile) and five- or seven-point (texture profile) scales are rarely seen and 10- or 15-point scales are more common.

A 'typical' consensus description of a product can take many forms. It is often tabular and looks exactly like any descriptive analysis with the exception that consensus scores are used instead of mean values (Chambers et al. 2010). Some studies have presented data in a traditional flavor profile format that includes 'order of appearance' data (Chambers & Robel 1993). Data can be presented in standard radar charts as with mean values from other descriptive methods (Sanchez & Chambers 2015) and multiple studies have used profile data in

Table 6.1 Tabular consensus scores for six production time periods.

Time period	1	2	3	4	5	6
Flavour						
Buttery	5.5	5.5	5.0	4.5	5.0	5.5
Brown sweet	8.5	8.5	7.5	7.0	8.0	8.0
Molasses	6.0	5.0	4.5	4.0	5.0	6.0
Sweet taste	5.0	5.5	5.0	4.5	5.0	5.0
Salty	1.5	1.5	1.5	1.5	1.5	1.5
Texture						
Surface roughness	10.0	10.0	10.0	10.0	10.0	10.0
Hardness	6.0	6.0	5.5	6.0	6.0	5.5
Crispness	7.5	6.5	6.5	7.0	7.0	7.0
Sustained crispness	7.5	6.5	6.0	6.0	6.5	6.5
Cohesivenss of mass	4.0	4.0	4.0	4.0	4.0	4.0
Roughness of mass	7.5	7.0	7.0	7.0	7.0	7.0
Particles	7.0	7.0	7.0	7.0	7.0	7.0

Scale: 1–15, none to extreme with 0.5 increments.

complex mapping such as the 'tree' map (Prell & Sawyer 1988) or principal components maps and deviations (Oupadissakoon et al. 2010). All are valid and can be used depending on what conveys the information best.

Table 6.1 shows a simple profile chart of key flavour and texture aspects of a sweet cookie/biscuit product that was tested during various times and plant production shifts during one day. The intent of the project was to determine the variability in product 'typically' produced at that plant. Multiple days were tested, but only one day is shown for brevity.

It is clear from the table that the plant is reasonably consistent in producing this product, but the molasses character of the product could be better controlled.

Table 6.2 shows a more traditional profile where order of appearance is considered and attributes that are not scored for that product are not included. In this case, a 'gold standard' profile for a cheese-flavoured potato chip from several years before is compared to a current version of that same chip to understand drift in flavour over time. In this case, it is quite apparent that the flavour of the product has changed over time. The saltiness, brothy/umami and processed cheddar cheese flavour have increased over time, the dried onion flavour has disappeared, and multiple flavour notes are now in a different position. Thus the original chip that started off as salty, buttery and cheesy now begins with a salty, brothy, cheesy flavour and the heated oil note that was near the end of the chip has increased and moved forward in the flavour. Consumer complaints about the current chip may be based on flavour formulation or processing changes that have been incrementally implemented since the 'gold standard' was tested.

Table 6.2 Flavour profile of the same chip: gold standard versus current.

	Gold standard cheese-flavoured chip		Current cheese-flavoured chip
Flavour	**Intensity**	**Flavor**	**Intensity**
Salt	7.5	Salt	8.5
Butter-like	4.5	Brothy/umami	5.5
Cheddar, processed	3.0	Cheddar, processed	4.0
Brothy/umami	4.5	Garlic, dried	6.0
Onion, dried	6.0	Butter-like	4.5
Garlic, dried	3.0	Potato, processed	3.0
Toasted	7.0	Oil-heated	7.0
Potato	5.5	Toasted	7.0
Potato, processed	3.0	Sweet, overall	1.0
Sweet, overall	3.0	Sour	2.0
Oil-heated	5.5	Potato	4.0
Sour	2.0	Bitter	2.5
Bitter	2.5		

Up to 15 points where 15 is extremely high; note that attributes that are not scored at least 0.5 are not included.

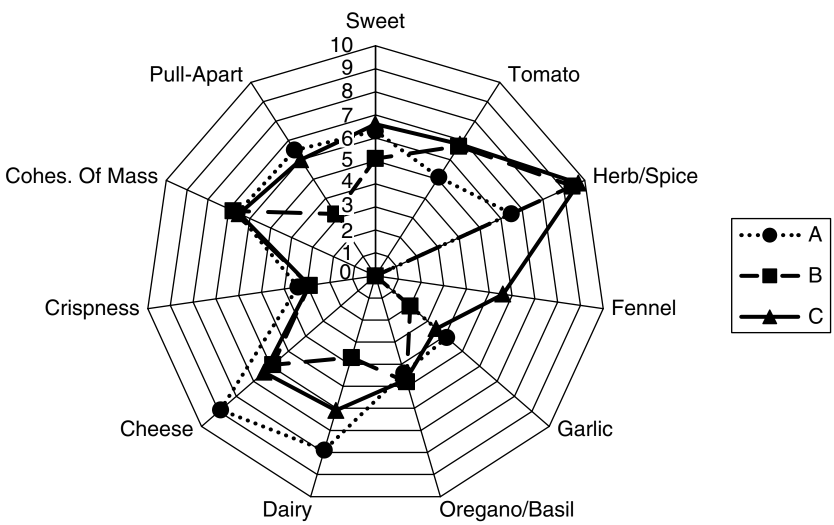


Figure 6.2 Radar chart of three pizza samples.

Consensus data can be used exactly the same as mean data in radar charts, 'trees' or mapping. For example, Figure 6.2 shows a radar chart comparing three pizzas for 11 key attributes determined using consensus methods.

The data clearly show that the pizzas in this test were similar for attributes such as crispness, cohesiveness of mass and oregano/basil, but were quite different in the perception of dairy and herb/spice characteristics.

Figure 6.3 is an example of a cluster tree that uses consensus data from over 100 cheeses to show sensory clustering. The tree diagram shows two major types of cheese (mould ripened versus non-mould ripened), two main subtypes within the non-mould-ripened cheeses (based mostly on milk type), and many sub- and sub-subtypes within those.

As with data from mean values, consensus scores can be used in multivariate mapping programmes to understand the relationship of attributes, products and/or consumer scores. In Figure 6.4, consensus scores for some key characteristics of pizza have been used to map the descriptive characteristics and products.

Variations of the consensus process are numerous, with differences in aspects such as initial training; the degree to which panellists have control over the attributes selected; scale values; multiple presentation, on a blind basis, of the

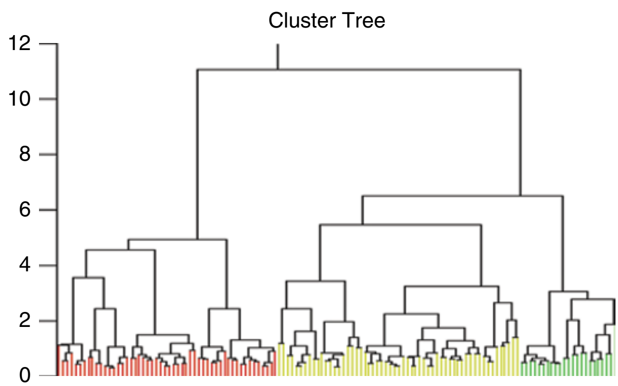


Figure 6.3 Tree diagram indicating similarities/differences among various types of cheeses.

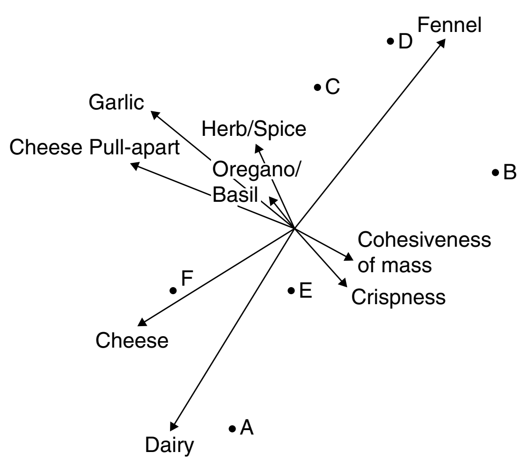


Figure 6.4 Two-dimensional map of consensus sensory scores and products.

same product to ensure consistency within the product or panel; differences in the number of panellists; and differences in the degree and manner in which panels reach consensus.

In many cases, the consensus method is adapted and used solely for the development of the lexicon or attributes to be tested. For example, in QDA (Stone 1992), 'panelists, as a group, meet with the panel leader and develop a common language that describes their perceptions of the products … The panel leader … does not participate in the actual development of the attributes …'. In the Spectrum Method (Muñoz & Civille 1992), a panel leader/administrator provides initial information to the panel to demonstrate attributes through a sample review and preliminary development of terminology and then works with the panel through subsequent reviews to develop and establish a final terminology that will be used for testing. Research commonly has been published using consensus methods for attribute development and then scoring samples individually.

Although not as common as using consensus methods for attribute development, there are situations, generally quality control, where consensus intensity data are used to make decisions. Muñoz et al. (1992) and Carpenter et al. (2000) both describe quality control scenarios where several individuals or panels are asked to determine if products are within a standardized range for key product attributes. These methods typically are portrayed as using a mean score from the panel as the deciding factor, but in a number of instances within industry, these decisions are determined by consensus to ensure that everyone is in agreement that a product is within or outside specifications.

6.3 Applications

The consensus method can be used whenever data of a highly descriptive nature are needed. It is particularly well suited to testing where attributes are not known fully in advance, such as in the case of product development or shelf-life studies. When adapted to include aspects such as order of appearance or wholistic characteristics, it is well suited to situations where the individual attributes or intensities may not change much but the overall character of the product will be different, such as with ingredient substitutions. Because of the extensive use of references, high degree of panel training and easy adaptability to various products, consensus methods are often used to develop descriptive sensory 'gold standard' profiles of products. These become historical reference 'fingerprints' for products that can be compared later to current products to better understand drift or deviation over time.

A unique use of the consensus method is in developing 'identity profiles'. These profiles are hypothetical 'typical' products that may not actually exist, but are typical for that particular product. In this situation, the panel develops consensus scores for a number of samples of the same product, such as 15 samples of the same grade and type of beef steak. Because the product varies from sample to sample, it is impossible to say that any particular sample is representative of that particular grade and type of beef steak. However, the panel can identify characteristics and intensities that are most 'typical' of the product under examination and can create an 'identity profile' for that product. This 'identity profile' then serves as the target for that product. This is fundamentally different from taking a mean value and is more similar to taking a median or mode of the scores and making human perceptual adjustment. This avoids determining what is 'typical' simply by averaging data and assuming that a mean value represents a typical product. For example, in the beef steak test, 10 of the 'roasted' flavour scores might range from 11 to 12 on a 15-point scale, two scores are 8, and three products score below 5. Thus, the typical score is 11–12 (maybe 11.5), but the average score might be 9–9.5. These data with five scores below 10 for roasted flavour do not contain outliers that simply can be 'thrown out', but those scores are not typical of most steak samples in the test. A 'typical' product (i.e. two-thirds of the products) would be expected to have a high 'roasted' score, which would not be shown using a non-consensus descriptive method that averages scores over panellists and replications.

Lastly, a common use for consensus methods is comparison of products that are different only in nuances that may not be obvious when focusing on specific attributes on a ballot. A criticism of individual-based descriptive systems is that panellists may be so focused on trying to differentiate products on a ballot that they miss nuances in products that may be important to consumers. For example, a cereal manufacturer used its typical, somewhat generic ballot to screen a new product introduction. The product performed well in the tests, meeting various pre-established criteria for appearance, flavour and texture. Unfortunately, a non-typical 'sandy' character was apparent as the product was chewed for swallowing. One panellist noted this on the ballot, but many did not and the analysts typically did not follow up when only one comment of a specific type was made. Weeks later, after a large batch of the product was made for consumer testing, several analysts were tasting the product before it was sent for consumer testing and noted the sandiness. The product was shown to the panel who agreed that a sandy texture was present, but had not been scored because they were so accustomed to the ballot that other aspects typically were ignored. A consensus approach may have alleviated this problem because the person who noted the problem in the original test would have provided that comment to the group and the group probably would have noted it in their evaluation.

6.4 Case Study

The Happy Cow Dairy Company has received complaints about its butter samples. Consumers are complaining about a 'bad taste' when they use the product.

- *Objective*: identify off-notes that appear to result in recent customer complaints of butter.
- *Samples*: samples from eights lots referenced in consumer complaints were obtained and served to the panel. A control gold standard product (labelled as the gold standard) with its flavour profile was served first during each session for use as a comparative reference. Two test samples were served during each session with a 10-minute break between samples.
- *Method*: flavour comparison to standard (consensus method, six panellists, 15-point scale with 0.5 increments, gold standard product and profile provided for reference). Complete profiles were not generated. Panellists were asked to taste the butter sample and compare it to the reference product/profile. Panellists evaluated the samples for differences in attributes present or gross differences in intensity of standard attributes.

Results are shown in Table 6.3.

In all samples, a moderately strong butyric acid note was found (see Table 6.3), suggesting that all the samples were deteriorated, most likely by a combination of oxidation and heat. In half the samples, additional attributes were found that suggested potential flavour transfer from packaging (cardboard) or contamination from microbes (musty, mouldy) or stale air (cardboard, musty, mouldy). Further investigation of the problem found that the temperature in the storage facility had exceeded recommendations for more than a week before the samples were shipped. Although an initial sensory test failed to find any problems before shipping, the higher than desired storage temperature apparently started a deterioration process that accelerated during in-market and consumer storage, resulting in a poor-quality product. A voluntary recall was initiated.

6.5 Advantages

6.5.1 Adaptability with Attributes

Because the method can be adapted as it is used, the ability to make changes during the experiment exists. For example, if a product attribute has not been determined prior to the start of the test, it can be added as needed by panellist agreement. This commonly occurs with off-flavours or odours in products where an unexpected attribute can appear as processes or time change. In consensus methods, panellists are encouraged to identify, reference and score new attributes during the test. Although many methods for trained panels may include a

Table 6.3 Attributes, with intensities, noted in test samples that were not found in the control sample.[1,2]

Lot number							
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Butyric - 8	Butyric - 6 Cardboard - 4	Musty - 6 Butyric - 5	Butyric - 7 Diacetyl - 2	Butyric - 6	Butryic - 8	Mouldy - 8 Butyric - 3	Butryic - 7 Cardboard - 4

[1] Intensity is based on a 0–15 scale with 0.5 increments.
[2] Panellists also were asked to note any major intensity differences (>2) in test samples compared to intensities of existing attributes of the gold standard. Several lower intensities were found in the test sample, but all were attributed to being overwhelmed by the butyric note found in the test samples.

'blank' space for adding attributes, consensus methods actually assume that such additions are not only possible but may be necessary.

The flexibility of the method allows panellists to add information to the profile or comments if they agree the information may be important to researchers. For example, the panellists used by Chang and Chambers (1992) used brackets to identify those attributes that occurred simultaneously. Such information usually cannot be obtained in other commonly used descriptive methods.

6.5.2 Group Think versus Individual Decision

When a group is working to make a team decision, a range of knowledge and experience is used and communication among the participants can stimulate thinking and bring new ideas (e.g. attributes that might not be noticed by everyone) to the group. The idea behind most group discussion and consensus methods, such as those used in decision making in other fields (e.g. Delphi and nominal group techniques and social judgement analysis), is that the cross-pollination of ideas makes the end result more robust than individual analysis that is then combined and averaged. The nominal group technique, widely used in psychology and medicine for consensus decision making, is quite similar to the process used in consensus sensory methods.

6.5.3 Time Sequence Information

One aspect included in the original flavor profile method was 'order of appearance'. This concept, which is now common in the sensory field, assumed that flavour did not occur all at one time, but that specific attributes appeared over time. Although it is possible to partly include such information in other general descriptive methods, consensus methods are best at this task. In many descriptive procedures, the time sequence can be indicated in a static way by listing the attributes in their general order of appearance on the ballot and in subsequent reports. However, with consensus methods, this order of appearance can be changed for each individual product as it is evaluated. This information is key in some industries such as liquid pharmaceuticals or beverages where the order of appearance of bitter and sweet tastes is essential to success. Various published studies have included such information and identified the key aspects (Caul & Vaden 1972; Chambers & Robel 1993).

6.5.4 Data Collection Time

It is possible with some panels that the data collection time for consensus methods could be lower than that for traditional testing. No published data were found comparing the time it takes to conduct studies in either mode, but experience with multiple types of studies indicates that time varies considerably within methods and occasions when data collection would be shorter for consensus are special circumstances. One case is where the panellists are particularly well trained and have worked together for a long period of time on particular products. In this case, the likelihood is that the evaluation will occur quickly because the products are well known to the panel and the group dynamics of the panel are such that they quickly come to agreement. A second scenario is the case where a broad range of products is under study and the ballot for a particular set of products is long and includes a number of attributes that are only found in a few products. Completing such long ballots may take more time for each individual panellist than coming to consensus on the attributes and intensities that are present in a particular sample.

6.6 Disadvantages

Amerine et al. (1965) listed a number of disadvantages of the method, many of which they determined could be overcome by including the use of a larger scale. However, they identified four criticisms that posed the greatest problems for the use of consensus methods:

- the time and cost involved with training and using consensus methods
- individual variation among panellists that cannot be accounted for
- discussion, which can lead to data bias
- the lack of statistical analysis.

6.6.1 Time and Cost

The first of these issues is critical to keep in mind. A well-trained panel using consensus methods does need extensive training. Hootman (1992) indicated that training panellists to use the flavor profile method, for example, generally would take approximately 6 months, with 1–3 sessions of testing needed for each product. However, Sjöström et al. (1957) indicated that some panels could be effective after 3 months of training. Other authors have described training of 120 hours over a 12-week period (Chambers et al. 1981; Lotong et al. 2002) and Chambers et al. (2004) showed that a panel being trained for potential use in consensus studies was better able to describe and differentiate small differences in flavour of pasta sauces with longer training (120 h). Drake et al. (2003) used a panel with 50 hours of training and an additional 40 hours of testing to develop a lexicon for dairy powders. Chambers et al. (1981) compared two panels with different levels of training and experience for three types of meat products. The semi-trained panellists, who had only three training sessions, were not as consistent as the more trained and experienced panel for most flavour attributes. However, both panels were equally consistent for most of the texture attributes in the study. The authors concluded that more training is required for complex flavour attributes and that less training can be used for attributes that are easily understood by panellists.

Of course, for some consensus aspects, training may take far less time. Stone et al. (1974) indicated that training usually took less than 20 hours for QDA which includes time for the panel to develop the lexicon of attributes, and studies using panels with little training proliferate in published literature. Bramesco and Setser (1990) used two panels, one with extensive prior training and one with no prior training, to develop terminology for texture of bakery products. Both panels had the same sensory analyst serving as panel leader. Although some differences existed in the terms and definitions that were developed, and the panel with no previous training took slightly longer to become proficient, the results were similar.

Clearly, the concern that panels for consensus studies need extensive training is part myth and part fact; common sense must be used. The types and numbers of different products that a panel is expected to test and the types of attributes to be tested influence the training a panel needs. This is true for most descriptive methods using trained panels. Difficult attributes in complex products require more training than the measurement of such attributes as sweetness intensity in model systems or urine/ammonia odour in cat litter, for example.

6.6.2 Individual Variation

Individual variation has been extensively studied in the sensory literature. Although it is widely understood that odours such as 'boar taint', sensitivity to chemical compounds such as 6-n-propylthiouracil (PROP) and 'colour blindness' are genetically controlled and result in huge variations in character or intensity, such differences generally are the exception rather than the rule and most companies do not develop consumer goods for such extremes. Certainly, smaller individual differences, such as those related to saliva secretion, do occur and can impact sensory perception. The use of static references, such as those described by most authors using full consensus or highly trained panels, tends to mitigate those differences. For example, Muñoz and Civille (1998) describe the impact of setting references at various intensities to change scale use. The potential disadvantage also serves as an advantage. If scores are simply averaged together, means may not represent either end of the spectrum of individual differences. One advantage of consensus methods is that they allow the panellists to comment on differences that cannot be agreed on and such commentary allows researchers to better understand issues they may face with customer variation.

6.6.3 Panel Discussion and Data Bias

Perhaps the most persistent criticism that continues to flourish is the idea that consensus descriptive data derived from profiling studies are easily influenced and, thus, give different information than data from methods that rely on individual quantitative assessments. That idea was discussed as early as 1958 by

Hall. Amerine et al. (1965) use that study as a primary argument against the use of consensus panels. Hall compared the results of an experienced panel with and without the use of two panellists who had been told to deliberately score high or low. When the two panellists gave scores that were intentionally high or low early in the discussion, other scores were affected. However, Hall also stated that 'To say that they (biases) can occur, however, does not mean that they will'. Meilgaard et al. (1991) indicated that bias can be caused 'with opinion dominated by … a senior member or a dominant personality and equal input from other members is not obtained'. Of course, this issue is true if the panel leader or panellists have been improperly trained and monitored. In fact, Hall (1958) noted that this potential hazard can be minimized by appropriate panel training, but indicated that such training takes time. Although she does not define appropriate training, the idea, as practised by sensory analysts, includes the concept of 'unconditional positive regard', a psychological concept that each individual's opinions and actions matter and that interaction should be kept positive. The ability of a group to work together long term and to come to consensus is much easier when everyone works together positively rather than being dominated by one or two individuals. This group dynamic has been shown to be understood and articulated even by fifth-grade children (Evans 2002).

It is critical in consensus descriptive methods that each and every panelist contributes to the discussion in a meaningful way. In many cases, the panellists are so highly trained that they provide the same data and need to do little discussion, but when they do not initially agree, it is up to the panel leader to ensure that true consensus is reached with input from the entire group. Hall (1958) is correct – training is necessary to overcome tendencies by a panellist or panel leader to dominate a group and tendencies by panellist to be swayed easily by other panel members. Clearly, if extensive training of panellists and panel leaders is not an option, consensus methods may not be the best choice for the group.

6.6.4 Lack of Statistical Treatment

Powers (1988) stated the 'major weakness the (consensus) method has is that judgment as to differences in profiles is almost always purely subjective' and Stone and Sidel (1985) said 'the lack of a true numerical system (and statistical procedure) for assessing product differences' was a major flaw of consensus methods. These criticisms have an element of truth, but also some inaccuracies. There is not an agreed-upon criterion for determining differences in consensus studies. From a purely hypothetical aspect, it could be said that because the data are consensus and, thus, represent the same score by each panellist, they are invariate and, by definition, significantly different whenever the numbers are different. That is not a completely satisfactory answer and most companies that have used consensus methods have made determinations other than simple difference to determine what is important. 'Is half a point or 1 point different enough to be meaningful' is a critical question that depends on the scale length, product category, objective and risk tolerance of the researcher. Of course, this is a problem for methods that use individual assessment and then conduct statistical determinations of difference. It is understood by most researchers that statistical differences do not always equate to meaningful differences.

Unfortunately, little research has compared mean scores with consensus methods. Syarief et al. (1985) used data collected in previous studies on a variety of products and compared consensus and mean scores that were obtained using profiling. Attribute scores that had been averaged over all products in a test from individual panellists' scores were virtually identical to consensus scores averaged over all products in a test. That is not surprising, because the mean scores were calculated from individual scores taken during the consensus process. The two sets of values were compared using the coefficients of variation (CV) calculated from analysis of variance of product means, and cumulative proportion of variance (CPV) accounted for using principal components analysis. Data from the mean scores procedure had a lower CV and higher CPV than consensus data. The authors interpreted that result as showing slight superiority for the means procedure. However, they failed to indicate if a greater range existed in final consensus scores than in the mean scores among specific products, thus showing greater differences among individual products for the consensus method. That would be a desirable effect and also would account for the higher CV and lower CPV found for the mean data.

Lotong et al. (2002) compared results for a category appraisal of orange juice of two independent panels, one that used a consensus approach and one that scored the products. Because the attributes and products were not identical, these authors did not compare actual scores for each product, but did compare the overall results to determine if data interpretation would be different depending on the panel used. Results showed that two independent highly trained panels from two different companies, testing the same general category, were able to essentially reproduce the data relationships and interpretation using either consensus or independent assessment.

The study by Lotong et al. (2002) highlights the misunderstanding of suggesting that consensus data cannot be statistically analysed. Although univariate procedures such as analysis of variance (ANOVA) cannot be applied to compare the individual data, the data can be used just as mean values are used in regression, correlation, principal components analysis and other multivariate procedures. In fact, the consensus data from individual products can still be used even in ANOVA procedures from studies such as factorial experiments to test and compare 'higher level' effects. For example, Sanchez and Chambers (2015) used the consensus scores from individual coffee samples in ANOVA tests to compare effects of brewing, regardless of coffee type, on sensory properties.

6.6.5 Panel Size

Recent arguments have stated that panel sizes for consensus panels, typically 4–6 panellists, are too small for reliable data. This is interesting because from the earliest writings on descriptive analysis (Sjöström et al. 1957; Brandt et al. 1963; Stone et al. 1974), authors have recommended that panellists be selected or used based on performance, that is, the ability to give reliable data. In none of those cases did the authors indicate that using a certain number of panellists was necessary to provide reliable data, although several authors did report the number of panellists that typically had been used. Powers (1988) discussed many methods of determining whether panels or panellists are performing properly and never stated the necessity for a specified number of panellists to achieve reliability. Yet, the idea that 4–6 panellists is too few appears to be firmly entrenched in the minds of many sensory analysts.

Most arguments for having a large number of descriptive panellists are similar to that mentioned in the ASTM *Manual on Sensory Testing Methods* (1968), indicating that the use of fewer than five panellists relies too much on one person's judgement. That handbook also states, however, that 'There is no "magic" number.' Basker (1977) concluded that the number of panellists needed for difference tests is a function of the degree of difference in the samples. In fact, he concluded that 'for gross differences, one observer may be enough'. Chambers et al. (1981) found that a panel of three highly trained, experienced individuals performed at least as reliably and discriminated among products as well as a group of eight less trained individuals. Those results suggested that the use of well-trained individuals could reduce the number of panellists necessary for sensory testing. More recently, Heymann et al. (2012) suggested 8–10 panellists should be used for descriptive studies, but did not determine that number with consensus panels.

Panel size should be a function of the size of the difference studied and the degree to which a panel has been trained to find that difference. Appropriate consideration must be given to the concern that all in a small group of people could be insensitive to a particular characteristic, but that concern presents no problem for many routine tests. The analogy that it takes only one, non-colour-blind person to determine whether an apple is red or yellow-green is simplistic but true.

From a practical standpoint, discussion and consensus building are far easier with fewer people. Muller and Wittenberg (2006), who studied large numbers of business teams containing various numbers of members, indicate that 5–6 may be an optimal team, where discussion and decision making are paramount.

6.6.6 Data Comparison

Consensus and individual panel data on two types of products, pizzas and bananas, were collected for this chapter. Panellists in the studies had been trained and were experienced in a variety of sensory methods, including profiling, and several different, individually measured, quantitative, sensory techniques. Consensus flavor and texture profiles were developed, and scores were compared to mean values for selected attributes using individual quantitative methods (QDA-type for pizza, Spectrum-type procedures for bananas) and similar 10- or 15-point scales. Data used for mean values were collected using individual booths independently from the profiling sessions; some panellists participated in both tests.

Data from pizzas (Table 6.4) indicate that the means of the individual scores, if rounded to the nearest whole number, were essentially identical to the consensus scores for almost all attributes. If the mean scores were rounded to the nearest whole number (as used in the profile method for pizzas) only

Table 6.4 Comparison of sensory scores obtained by consensus or by individual scoring for pizzas.[1]

Attribute		Consensus[2]	Individual[3]	Individual[4]
Body				
	Pullaway	7	6.8	6.9
	Oily residual	7	6.9	7.1
	Initial crispness	4	3.9	3.5
	Chewiness	7	6.7	7.0
	Cohesiveness of mass	6	6.2	6.9
Mid-section				
	Cheesy	3	3.1	3.0
	Spicy	7	6.7	6.9
	Tomato	5	5.2	5.3
	Oily	6	5.8	5.9
	Toasted cheese	6	5.6	5.1
	Yeasty	4	4.0	4.1
	Doughy	5	5.0	4.9
	Salty	5	5.2	5.6
Outer crust				
	Initial crispness	7	6.8	6.9
	Firmness	7	6.4	6.7
	Oily	8	7.6	7.8
	Toasted grain	7	6.1	6.5
	Grainy	3	3.1	3.2
	Yeasty	3	3.0	3.1
	Doughy	3	3.2	3.3

[1] Scale: consensus scores ranged from 0 = none to 10 = strong on a numerical scale with 1 point increments; individual scores ranged from 0 = none to 10 on a graphical line scale.
[2] Scores for consensus are from 7 panellists testing 10 pizzas before agreement was reached.
[3] Scores for individual assessment are means of 14 panellists for 16 replications.
[4] Scores for individual assessment are means of 6 panellists for 4 replications collected in a study of pizza variations approximately 2 months after the initial study.

Table 6.5 Comparison of sensory scores obtained by consensus or by individual scoring for bananas.[1]

	Sample A		Sample B	
Attribute	**Consensus[2]**	**Individual[3]**	**Consensus[2]**	**Individual[3]**
Sweet	5.0	5.0	4.5	4.5
Sour	3.5	3.4	4.0	4.2
Bitter	1.0	1.1	1.5	1.4
Amyl butyrate	6.0	5.8	5.5	5.0
Fruity	5.0	4.9	5.0	4.5
Viney	2.5	2.9	4.0	3.8
Astringency	3.0	3.1	2.5	2.5
Earthy	1.0	1.3	1.0	1.1
Firmness	4.0	3.9	4.0	4.0
Mealy	1.5	1.3	1.5	1.0
Cohesiveness of mass	7.5	7.5	7.0	7.2

[1] Scale: scores range from 0 = none to 15 = extreme on a numerical scale with half-point increments.
[2] Scores for consensus are from 5 panellists testing 5 bananas each before agreement was reached.
[3] Scores for individual assessment are means of 5 panellists for 3 replications.

firmness and toasted grain for sample 1 and toasted cheese and salty for sample 2 would have been different and, before rounding, those scores differed by less than 1 scale value on the 0–10-point scale. The values for each of the attributes for bananas did not differ by more than 0.5 point on a 0–15-point scale (Table 6.5) and, in most cases, they differed by less than 0.2.

These data suggest that once a panel is well trained, and if they use reference products for scale anchors, the same values can be obtained whether by consensus or by mean scores from a quantitative descriptive sensory procedure, which challenges the myth that consensus data are less 'correct' than mean scores. The implication is that the training and use of standards appear to be more important than the data collection method.

6.7 Future Developments

One development that is already being tested and implemented in some software programs is electronic collection and display of individual data used for discussion. Previously, attributes and scores were read and discussed (with constant re-reading to remember what was said) or the panel leader went through a laborious process of copying the individual data onto display boards for comparison, discussion and determining consensus data. The use of technology to capture handwritten data (or data input using other methods) and display it in real time shortens the process of evaluating, transcribing and discussing by eliminating the second step in the process. New ways of inputting data and communication also will impact the consensus process and potentially allow it to occur in remote locations simultaneously.

Other than the use of technology to enhance the process, consensus methods are relatively simple to conduct if the panel is well trained and understands the concept of consensus. Thus, future developments are likely to be in areas of use and analysis rather than technique development. The expanded use of consensus methods for non-food products and for such items as time intensity, profiling the eating experience over multiple bites and multiple foods within a meal is likely. In fact, such work is already occurring. Additionally, the continued development of statistical techniques that will improve the comparison of order of appearance, compare products holistically and expand our ability to examine and display data in multivariate spaces will enable consensus techniques to overcome the univariate data analysis shortcomings that sometimes hold back its use.

6.8 Summary

Consensus methods of descriptive analysis have a long history of use in consumer products and other industries. They range from full consensus methods to methods using panels that will only use consensus for either attributes or intensities. Such methods are widely used by industry, government and academic institutions.

Typically, for full consensus methods, a highly trained panel is needed. This can be seen as either an advantage in terms of consistency or reproducibility of results or a disadvantage in terms of the amount of time, money and effort it takes to train to that level. Consensus methods, especially those with poorly trained panels and panel leaders, can be susceptible to various potential group biases, but training and monitoring can be undertaken to avoid such problems.

Because the consensus methods are readily adaptable to many situations, they can be used widely to evaluate products and processes for many different types of objectives. Although the data may not lend themselves to traditional univariate statistics, they work well in applications that are aided by multivariate analyses.

References

Amerine, M.A., Pangborn, R.M. & Roessler, E.B. (1965) *Principles of Sensory Evaluation of Food*. New York: Academic Press, pp. 277–386.

Asch, S.E. (1951) Effects of group pressure on the modification and distortion of judgments. In: Guetzkow, H. (ed.) *Groups, Leadership and Men*. Pittsburgh: Carnegie Press, pp. 177–190.

ASTM (1968) *Manual on Sensory Testing Methods*. Philadelphia: American Society for Testing and Materials, p. 20.

Basker, D. (1977) The number of assessors required for taste panels. *Chemical Senses & Flavor*, **2**, 493–496.

Berry, B.W., Maga, J.A., Calkins, C.R., Wells, L.H., Carpenter, Z.L. & Cross, H.R. (1980) Flavor profile analysis of cooked beef loin steaks. *Journal of Food Science*, **45**, 1113–1115.

Bramesco, N.P. & Setser, C.S. (1990) Application of sensory texture profiling to baked products: some considerations for evaluation, definition of parameters and reference products. *Journal of Texture Studies*, **21**, 235–251.

Brandt, M.A., Skinner, E.Z. & Coleman, J.A. (1963) Texture profile method. *Journal of Food Science*, **28**, 404–409.

Brooks, B.J. & Koretsky, M.D. (2011) The influence of group discussion on students' responses and confidence during peer instruction. *Journal of Chemical Education*, **88**, 1477–1484.

Cairncross, S.E. & Sjöström, L.B. (1948) What glutamate does in food. *Food Industry*, **20**, 982–983, 1106–1107.

Cairncross, S.E. & Sjöström, L.B. (1950) Flavor Profiles – a new approach to flavor problems. *Food Technology*, **4**, 308–311.

Carpenter, R.P., Lyon, D.H. & Hasdell, T.A. (2000) *Guidelines for Sensory Analysis in Food Product Development and Quality Control*, 2nd edn. New York: Aspen Publishers.

Caul, J.F. (1957) The profile method of flavor analysis. In: Mrak, E.M. & Stewart, G.F. (eds) *Advances in Food Research*, vol. **7**. New York: Academic Press, pp. 1–40.

Caul, J.F. & Rockwood, E.L. (1953) A flavor panel study of the application of monosodium glutamate to the flavor of pharmaceutical products. *Journal of Pharmaceutical Science*, **42**, 682–685.

Caul, J.F. & Sjöström, L.B. (1951) Effects of replacing sucrose by other sweeteners in blackberry and raspberry packs. *Quick Frozen Foods*, **13**, 59–61.

Caul, J.F. & Vaden, A.G. (1972). Flavor of white bread as it ages. *Baker's Digest*, **46**(1), 39, 42–43, 60.

Chambers, D.H., Allison, A.A. & Chambers, E. IV (2004) Effects of training on performance of descriptive sensory panelists. *Journal of Sensory Studies*, **19**, 486–499.

Chambers, D.H., Esteve, E. & Retiveau, A. (2010) Effect of milk pasteurization on flavor properties of seven commercially available French cheese types. *Journal of Sensory Studies*, **25**, 494–511.

Chambers, E. IV & Robel, A. (1993) Sensory characteristics of selected species of freshwater fish in retail distribution. *Journal of Food Science*, **58**, 508–512, 561.

Chambers, E. IV, Bowers, J.A. & Dayton, A.D. (1981) Statistical designs and panel training/experience for sensory analysis. *Journal of Food Science*, **46**, 1902–1906.

Chang, C.Y. & Chambers, E. IV (1992) Flavor characterization of breads made from hard red winter wheat and hard white winter wheat. *Cereal Chemistry*, **69**, 556–559.

Drake, M.A., Karagul-Yuceer, Y., Cadwallader, K.R., Civille, G.V. & Tong, P.S. (2003) Determination of the sensory attributes of dried milk powders and dairy ingredients. *Journal of Sensory Studies*, **18**, 199–216.

Evans, K.S. (2002) Fifth-grade students' perceptions of how they experience literature discussion groups. *Reading Research Quarterly*, **37**, 46–69.

Festinger, L. (1954) A theory of social comparison processes. *Human Relations*, **7**, 117–140.

Fiero, G.W. (1958) Applications of odor evaluation to petroleum products. In: A. D. Little, Inc. (ed.) *Flavor Research and Food Acceptance*. New York: Reinhold, pp. 221–248.

Gabel, M.J., Foster, N.L., Heidebrink, J.L. et al. (2010) Validation of consensus panel diagnosis in dementia. *Archives of Neurology*, **67**, 1506–1512.

Hall, R.L. (1958) Flavor study approaches at McCormick & Company, Inc. In: A. D. Little, Inc. (ed.) *Flavor Research and Food Acceptance*. New York: Reinhold, pp. 224–240.

Henry, R.A. (1993) Group judgment accuracy: reliability and validity of postdiscussion confidence judgments. *Organizational Behavior & Human Decision Processes*, **56**(1), 11–27.

Heymann, H., Machado, B., Torri, L. & Robinson, A.L. (2012) How many judges should one use for sensory descriptive analysis? *Journal of Sensory Studies*, **27**, 111–122.

Hicks, J., Monty, R. & Myers, T. (1966) Group consensus and judgmental accuracy: extension of the Asch effect. *Psychonomic Science*, **4**(4), 159–160.

Hootman, R.C. (1992) *ASTM Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia: American Society for Testing and Materials, pp. 1–3.

Kurtzman, C.H. & Sjöström, L.B. (1964) The flavor-modifying properties of disodium inosinate. *Food Technology*, **18**(9), 67–69.

Lin, V.S., Goodwin, P. & Song, H. (2014) Accuracy and bias of experts' adjusted forecasts. *Annals of Tourism Research*, **48**, 156–174.

Lotong, V., Chambers, D.H., Dus, C., Chambers, E. IV & Civille, G.V. (2002) Matching results of two independent highly trained sensory panels using different descriptive analysis methods. *Journal of Sensory Studies*, **17**, 429–443.

Meilgaard, M., Civille, G.V. & Carr, B.T. (1991) *Sensory Evaluation Techniques*, 2nd edn. Boca Raton: CRC Press, pp. 119–141, 163–179.

Muller, J.S. & Wittenberg, E. (2006) *Is Your Team Too Big? Too Small? What is the Right Number?* Available at: http://knowledge.wharton.upenn.edu/article.cfm?articleid=1501 (accessed 2 June 2017).

Muñoz, A.M. & Civille, G.V. (1992) The Spectrum descriptive analysis Method. In: Hootman, R.C. (ed.) *ASTM Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia: American Society for Testing and Materials, pp. 22–34.

Muñoz, A.M. & Civille, G.V. (1998) Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, **13**, 57–75.

Muñoz, A.M., Civille, G.V. & Carr B.T. (1992) *Sensory Evaluation in Quality Control*. New York: Van Nostrand Reinhold

Murphy, M.K., Black, N.A., Lamping, D.L. et al. (1998) Consensus development methods and their use in clinical guideline development. *Health Technology Assessment*, **2**(3), 1–7.

Oupadissakoon, C., Chambers, E. IV, Kongpensook, V., Suwonsichon, S., Yenket, R. & Retiveau, A. (2010) Sensory properties and consumer acceptance of sweet tamarind varieties grown in Thailand. *Journal of the Science of Food & Agriculture*, **90**, 1081–1088.

Powers, J.J. (1988) Current practices and application of descriptive methods. In: Piggott, J.R. (ed.) *Sensory Analysis of Foods*, 2nd edn. London: Elsevier Applied Science, pp. 187–266.

Prell, P.A. & Sawyer, F.M. (1988) Flavor profiles of 17 species of North Atlantic fish. *Journal of Food Science*, **54**, 1036–1042.

Sanchez, K. & Chambers, E. IV (2015) How does product preparation affect sensory properties? An example with coffee. *Journal of Sensory Studies*, **30**, 499–511.

Sjöström, L.B., Cairncross, S.E. & Caul, J.F. (1957) Methodology of the flavor profile. *Food Technology*, **11**(9, Suppl), 20–25.

Song, F. (2009) Intergroup trust and reciprocity in strategic interactions: effects of group decision-making mechanisms. *Organizational Behavior & Human Decision Processes*, **108**(1), 164–173.

Stevens, A., Abrams, K., Brazier, J. & Fitzpatrick, R. (2001) *The Advanced Handbook of Methods in Evidence Based Healthcare*. London: Sage Publications, p. 439.

Stone, H. (1992) Quantitative descriptive analysis (QDA). In: Hootman, R.C. (ed.) *ASTM Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia: American Society for Testing and Materials, pp. 15–21.

Stone, H. & Sidel, J.L. (1985) *Sensory Evaluation Practices*. Orlando: Academic Press, pp. 194–226.

Stone, H., Sidel, J., Oliver, S., Woolsey, S. & Singleton, R.C. (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology*, **28**(11), 24–34.

Suwonsichon, S., Chambers, E. IV, Kongpensook, V. & Oupadissakoon, C. (2012) Sensory lexicon for mango as affected by cultivars and stages of ripeness. *Journal of Sensory Studies*, **27**, 148–160.

Syarief, H., Hamann, D.D., Giesbrecht, F.G., Young, C.T. & Monroe, R.J. (1985) Comparison of mean and consensus scores from flavor and texture profile analyses of selected food products. *Journal of Food Science*, **50**, 647–650, 660.

Vara-Ubol, S., Chambers, E. IV, Kongpensook, V., Oupadissakoon, C., Yenket, R. & Retiveau, A. (2006) Determination of the sensory characteristics of rose apples cultivated in Thailand. *Journal of Food Science*, **71**, 547–552.

CHAPTER 7

Original Flavor and Texture Profile and Modified/Derivative Profile Descriptive Methods

Alejandra M. Muñoz and Patricia A. Keane

7.1 Introduction/Historical Perspective

The original flavor* and texture profile methodology represents one of the two fundamental and original descriptive analysis philosophies and approaches in sensory science. Together with the quantitative descriptive analysis (QDA) descriptive method, the flavor and texture profile methods have established the methodological foundations for almost all current traditional descriptive practices (Muñoz & Bleibaum 2001).

Philosophically, and ultimately methodologically, practitioners selecting a traditional descriptive method have a choice between implementing a rigorous technical descriptive approach (profile-based methodology) or a consumer behaviour approach (QDA) (ASTM 2011). Traditional descriptive methods (such as QDA and the original and modified profile methods), as referred to by these authors, are methods that involve a training programme, the evaluation of product attributes and the rating of their intensities by a trained panel. Conversely, non-traditional descriptive methods (e.g. free choice, flash profile, etc.) might not include a training programme and have incorporated different/new techniques to characterize products (Dairou & Sieffermann 2002; Marshall & Kirby 1988; Thomson & MacFie 1983).

The consumer behaviour QDA approach is covered in detail in Chapter 8 of this book, as well as in other well-known publications (Stone 1992; Stone & Sidel 1998; Stone et al. 1974). This chapter focuses on the second fundamental traditional descriptive approach: the original profile methods. In addition, this chapter presents the general characteristics of modified profile methods and how the fundamental principles of the original profile methods were adapted and are being used in these modified methods.

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

7.1.1 Historical Perspective of the Original Profile Methodology

7.1.1.1 The Flavor Profile Method

This was the first formal and technical descriptive method developed. Prior to its development, product evaluations were conducted only by one individual or a few experts.

In the 1940s, Arthur D. Little, Inc. (ADL) scientists developed the flavor profile method from their practical work in diverse food and non-food products. As Caul (1957) indicated, the flavor profile method was an approach that, unlike difference tests, focused on the whole flavour of a product, thereby considering the individual attributes of a flavour in relation to each other. The early 1940s techniques were improved and systematized into a more exact technique by Cairncross and Sjöström, and the flavor profile method was formalized and officially introduced to the food field in 1949 (Cairncross & Sjöström 1950). These investigators demonstrated that it was possible to select and train a group of individuals capable of describing their perceptions of a product in some agreed sequence, leading to actionable results without dependence on the individual expert.

As McDowell (1963) stated, the development of the flavor profile method resulted from the need for a more precise means of measuring the aroma and flavour of food products than what was in use. Caul (1957) provided a more detailed description of the flavor profile method, discussing the selection of panel members, panel training, panel leader characteristics, prerequisites for the profile panel operation, test execution (i.e. orientation, presentation of samples, development of vocabulary, panel sessions), and even advanced concepts such as time intensity evaluations and the use of intensity/quantitative references.

The development of the flavor profile method truly represents a milestone in sensory science. This method attracted considerable interest, as well as controversy; however, there is no question as to its historical importance to the field of sensory science.

The main contributions of the original flavor profile method, as viewed by these authors, are:

- in essence, having established the foundations of descriptive analysis
- the development of an approach to establish, through a formal training programme, a technical descriptive profile panel capable of fully characterizing the product's flavour attributes. This approach represented a remarkable evolution over the previous practice of conducting evaluations with only one or a few experts
- the implementation of a screening process to select the most qualified assessors
- the careful design and completion of a rigorous training programme to establish a well-trained descriptive panel
- the practice and benefits of:
  - establishing and using a comprehensive, thorough and technical terminology to characterize the products' flavour

  - using qualitative references during training to demonstrate and clarify attributes and terms
  - using quantitative/intensity references/anchors to standardize scale usage across assessors
- the evaluation of complex attributes to fully understand a product's flavour, such as:
  - amplitude and fullness
  - order of appearance
- the use of consensus to reach panel agreement on product attributes, the final profiles and product differences and similarities.

7.1.1.2 The Texture Profile Method

This descriptive technique was developed in the 1960s by scientists working at the General Foods Co. The milestones in the development of the texture profile method are described below and were:

- the work by Szczesniak (1963) and Szczesniak et al. (1963) (see sections 7.1.1.2.1 and 7.1.1.2.2), and
- the development and publication of the texture profile descriptive principles by Brandt et al. (1963), based on the work by Szczesniak (see section 7.1.1.2.3).

7.1.1.2.1 Classification of Textural Characteristics and Foundations of Sensory Texture Evaluations

Szczesniak (1963) pointed out that prior to 1963, texture was the least well described sensory dimension. Prior work focused only on individual foods, and there was a lack of a scientific approach (i.e. an adequate bridge between theoretical rheology and practical applications). Thus her work aimed to establish a system of texture nomenclature that would apply to all foods. She proposed the fundamental classification of texture sensory characteristics, consisting of three groups: Mechanical, geometrical and moisture/fat content characteristics. Her ideas were key for the development of the texture profile method in foods (see section 7.1.1.2.3), and years later the basis for adapting this classification to other texture-related evaluation techniques such as skinfeel, handfeel evaluations, etc.

7.1.1.2.2 The Development of Rating Scales and Use of Intensity/Quantitative References

Szczesniak et al. (1963) developed standard rating scales based on the proposed classification of textural characteristics suggested by Szczesniak (1963). The objectives of Szczesniak and coworkers in developing the texture scales were to:

- provide a defined quantitative evaluation method of the mechanical parameters of texture. These authors' ideas and scales were intended to encompass the entire range of intensity of textural characteristics rather than a selected range
- select specific examples for each point on the scale so that proper intensity reference standards could be available. This would allow the intensity rating of a given textural parameter in an 'unknown' product relative to the

known intensities of these reference products. As stated by Szczesniak et al. (1963), this would also eliminate the problem of panel drifting
- use the developed standard scales in seeking a correlation between sensory and instrumental methods of texture evaluation.

Although the flavor profile method is to be credited for first using intensity reference scales, the texture profile method and the work of Szczesniak et al. (1963) more clearly delineated the practice of developing and using intensity scale anchors. This practice has been implemented by several modified/ profile methods (e.g. A^{5}daptive Profile Method®, Spectrum™ Method).

7.1.1.2.3 Development of the Texture Profile Method Applying the Research by Szczesniak (1963) and Szczesniak et al. (1963)

The texture profile method (Brandt et al. 1963) was developed based on two important milestones:
- the flavor profile methodology. As Brandt et al. (1963) indicated, the flavor profile method was a good model to use in developing the texture profile method
- the work by Szczesniak (1963) and Szczesniak et al. (1963) on the classification of textural characteristics and the development of rating/intensity scales. As indicated by Brandt et al. (1963), the work developed by Szczesniak and coauthors contributed a logical and well-defined basis upon which to build a comprehensive sensory method for evaluating the texture of foods.

The texture profile principles published by Brandt et al. (1963) established the basis for texture evaluations and the main characteristics of the texture profile method.
- The measurement of mechanical, geometrical and moisture/fat-related characteristics.
- The consideration of order of appearance of texture properties in proposing the evaluation of these characteristics in defined stages of evaluation, that is, initial, masticatory and residual stages.
- The objectivity of the method through the use of a rigidly defined texture nomenclature and points of reference (intensity reference points/anchors).

7.1.1.2.4 The Texture Profile Method's Main Contributions

Besides expanding the scope of the flavor profile method and encompassing the texture sensory dimension, the texture profile method provided new principles that contributed to further grow the field of descriptive analysis.

The main contributions of the texture profile method, as viewed by these authors, are as follows.
- The strategy of adapting fundamental descriptive analysis principles (i.e. the flavor profile method's philosophy and principles) to develop a flavor profile derivative method (i.e. the texture profile method). This contribution and approach have been used ever since by all modified/derivative profile methods, in adapting the fundamental principles of both the flavor and texture profile methods to structure their modified profile methodology.
- The adaptation of the order of appearance assessment from the flavor profile method to evaluate texture properties in the order they appear/are

perceived. This is accomplished by specifying the stages of mastication/manipulation in a texture/skinfeel/handfeel evaluation and scoring the attributes perceived in those stages.
- The technical and rigorous nature of texture descriptive analysis, that is, establishing and using very technical and well-defined evaluation procedures and attributes.
- The continued emphasis on and use of quantitative/intensity reference scales, and establishing the practice of developing these scales and anchors (a practice currently used by many derivative profile methods).

Based on the above, the original flavor and texture profile methods established the key elements of descriptive analysis and the basis for all current modified/derivative profile methods (profile attribute analysis (PAA), quantitative flavour profile (QFP), A^{5}daptive Profile Method®, skinfeel profile descriptive analysis, Dynamic Flavor Profile® (DFP®), Spectrum™ Method, etc.). All these methods have embraced and used many of the fundamental principles of the original flavor and texture profile methods.

7.1.2 Historical Perspective of Modified/Derivative Profile Methodology

7.1.2.1 Development of Modified/Derivative Profile Methods and Milestones

7.1.2.1.1 Profile Attribute Analysis (PAA)

This represents the first modified/derivative profile method and the model taken for the development of all derivative profile methods (quantitative flavour profile (QFP), Spectrum Method, etc.).

Profile attribute analysis was developed in the 1970s by ADL sensory scientists and was used exclusively for and by their clients for several years. It was later published and shared with the scientific community (Hanson et al. 1983; Neilson et al. 1988).

Profile attribute analysis adapted the original flavor profile method's philosophy and practices. It was structured to provide more focused evaluations (i.e. limiting the number of attributes) and a numerical scale with more scale points that could offer statistical advantages. Other modifications incorporated in the PAA method were the evaluation of fullness and balance as independent measures, and attention to and evaluation of other non-flavour assessments relevant to the product such as visual, tactile and auditory attributes (Neilson et al. 1988). PAA also differed from the flavor profile method, in that PAA developed a ballot, listing the panel's selected and defined attributes and their corresponding numerical scales. However, it maintained the flavor profile method's objective to develop and use highly technical descriptive language.

Modified/derivative profile methods have followed the model and approach of PAA in adapting/improving concepts of the original flavor and texture profile methods to structure their methodology. Thus, PAA represents a true milestone and model for current modified profile descriptive analysis practices.

7.1.2.1.2 Texture Profile Derivative Methodology and Milestones

Parallel to the contributions made by PAA, new developments and growth of the original texture profile method occurred. These modifications provided additional foundations for all derivative/modified profile methods.

Additions to the Texture Profile Method

Sensory analysts working for the General Foods Co., who used the original texture profile method (Civille & Liska 1975), suggested some additions to the method, such as:

- measurement of surface properties (e.g. wetness, greasiness)
- awareness of all relevant texture product attributes and their inclusion in the complete texture evaluation. Additional attributes mentioned were cohesiveness of the mass, bounce/elasticity, spreading over tongue, uniformity during disappearance, etc.

Improved Texture Rating Scales and Establishment of a Technique to Develop Intensity Reference Scales

As discussed in section 7.1.1.2.2, the development of rating scales by Szczesniak et al. (1963) represented a true milestone in descriptive analysis. These scientists established the philosophy of developing and using intensity scales and anchors to rate attribute intensities. This practice has been implemented by several modified profile methods (A^{5}daptive Profile Method, skinfeel profile descriptive analysis method, Spectrum Method, etc.).

The following important modifications to the original intensity scales were made by several researchers. This work not only advanced the original scales and concepts, but provided an approach to establish or improve intensity reference scales.

- The update and improvement of texture reference scales by Muñoz (1986). Her work represents a significant contribution to descriptive analysis from two perspectives:
  - a substantial update and modification of reference materials and scale values of the standard scales developed by General Foods scientists (Szczesniak et al. 1963)
  - the development and publication of new reference scales with expanded definitions and evaluation procedures for selected texture attributes.

  The new scales published by Muñoz (1986) included wetness, surface adhesiveness, self- adhesiveness, roughness, springiness, cohesiveness of mass, moisture absorption, adhesiveness to teeth and manual adhesiveness.
- The adaptation of texture scales for other countries and cultures (Bourne et al. 1975 in Colombia; Hough et al. 1994 in Argentina; Jenkins et al. 2008 in Asia). Table 7.1 shows, as an example and for one attribute/scale, the evolution and modification of the texture scales: the original (Szczesniak et al. 1963), the updated (Muñoz 1986) and the country adapted (Hough et al. 1994) brittleness/fracturability scale.

This pioneering work has set the basis for how scales are adapted and new scales developed.

Table 7.1 Comparison of original and modified texture scales (example: brittleness/fracturability).

Brittleness Original Texture Profile Scale (Szczesniak et al. 1963)*		Fracturability Modified Profile Scale (Muñoz 1986)*		Fracturability Country Adapted Scale (Hough et al. 1994)*	
Scale value	**Product**	**Scale Value**	**Product**	**Scale Value**	**Product**
1	Corn muffin	1	Corn muffin	1	Cake
2	Angel Puffs	2.5	Egg jumbo cookie	2.5	Biscuit layer
3	Graham crackers	4.5	Graham cracker plain	5	Cracker
4	Melba toast	7	Melba toast	7	Sweet toast biscuit
5	Jan Hazel cookies	8	Ginger snap	8	Biscuit
6	Ginger snaps	10	Thin bread wafer	10	Thin bread wafer
7	Peanut brittle	13	Peanut brittle	12	Peppermint drop
		14.5	Hard candy	14.5	Hard candy

* Selected information is presented from tables published in these articles. Consult original publications for complete product information (brands and preparation).

Application of the Foods Texture Profile Methodology in the Descriptive Evaluation of Other Sensory Dimensions and Products

Schwartz (1975) is recognized for her pioneering work in adapting the principles of the texture profile method to develop the skinfeel profile descriptive analysis practices. Schwartz (1975) provided new suggested terminology for skinfeel evaluations which, as she indicated, represented the application of modified texture profile techniques to the evaluation of skincare products.

In her paper, Schwartz suggested these evaluation stages for skinfeel assessments: pick-up, rub-out and afterfeel. The skincare product attributes and definitions she suggested included thickness, consistency, spreadability, absorbency and residue (film, coating, flaky or powdery particles, dryness/moistness, oiliness, irritated, etc.).

This work has been further advanced by members of ASTM in the development of a standard practice for descriptive skinfeel analysis of creams and lotions (ASTM 2003). These practices were expanded to cover the technical (profile-based) and consumer-based approaches for descriptive skinfeel evaluations (ASTM 2011).

7.1.2.1.3 Other Modified Profile Methods

Based on the approach followed by the PAA method and the texture profile derivative methods described above, modifications of the original flavor and texture profile have been implemented in:

- the methodology used by current modified profile methods such as the QFP, A^{5}daptive Profile Method, skinfeel profile descriptive analysis, DFP, Spectrum Method, etc. Specifics on several of these methods can be found in the corresponding chapters of this book
- the application of modified/derivative profile techniques in diverse projects (Adhikari et al. 2011; Hongsoongnern & Chambers 2008; Thompson et al. 2009).

7.1.2.2 Definition and Characteristics of Modified/Derivative Profile Methods

Modified/derivative profile methods are descriptive methods that use one or several of the fundamental principles of the original flavor and texture profile methods. Examples of modified profile methods are PAA, QFP, A^{5}daptive Profile Method, skinfeel profile descriptive analysis, DFP and the Spectrum Method.

Table 7.2 presents a comparison of the key characteristics of the original and modified/derivative profile methods, showing how several of the original profile methods' philosophical and methodological principles are applied in modified profile methods, as follows.

- Technical training programmes to achieve sound and technical lexicons (e.g. attributes defined and measured in the most technical and specific way). Similarly to the original profile methods, modified/derivative profile methods strive to establish and use technical attributes demonstrated to the panel through carefully chosen qualitative references, which often are chemical references and ingredients. Examples of profile technical terms include the following.
  - *Flavour attributes*: diacetyl, methyl-anthranilate, butyric acid, caramelized/ browned, etc. This means using technical terms instead of general descriptors such as buttery (for a diacetyl perception), grape (for a methyl-anthranilate perception), etc.
  - *Texture/skinfeel*: cohesiveness of mass, moisture absorption, slipperiness, oily and greasy residuals, etc.
  - *Appearance*: hue, chroma/brightness, shininess.
  - *Fragrance*: aldehydic, mossy, musky, etc.
- Use of quantitative references/anchors by some modified/derivative profile methods. Reference materials are used to train assessors on intensities and by the panel to rate intensities, as follows.
  - Flavour references as used and published by the developers of the flavor profile method (Cairncross & Sjöström 1950; Caul & Sjöström 1951; Caul et al. 1958) and also by General Foods practitioners.
  - Texture reference scales for attributes such as hardness, fracturability and others originally published by Szczesniak et al. (1963) and updated and expanded by Muñoz (1986).
  - Skinfeel references (ASTM 2003).
- Measurement of one or several of the flavor profile complex attributes. In a few modified profile methods, such as PAA and the A^{5}daptive Profile Method, panels are trained to evaluate several of these attributes: overall impression. fullness, balance, order of appearance, aftertaste, etc.

Table 7.2 Comparison of original and modified/derivative profile methods' main characteristics.

Characteristics	Original profile methods	Modified/derivative profile methods
OVERALL PHILOSOPHY		
Technical approach	√	√
Rigorous training	√	√
Qualitative references (as technical as possible)	√	√
Quantitative/intensity references	√	√
Consensus	√	Only in some modified methods (e.g. PAA and A^{5}daptive Profile Method)
Individual data and analysed statistically	X	√
Panel pool	4–6	Larger panels, since in general, individual data, and not consensus, are collected
FLAVOUR/FRAGRANCE		
Aroma	√	X (only done as a special interest or limited)
Order of appearance	√	X (only done as a special interest)
Amplitude	√	X
Blendedness, balance, fullness	X	Only in some modified methods (e.g. PAA and A^{5}daptive Profile Method)
Dimensions and products	Foods, pharmaceuticals, food ingredients, package	Foods, pharmaceuticals, food ingredients, package, fragrance, odours
TEXTURE/SKINFEEL/HANDFEEL		
Mechanical attributes	√	√
Geometrical	√	√
Fat/moisture	√	√
Order of appearance/stages of mastication/manipulations	√	√
Texture scales (used in training)	Original 9-point scales by Szczesniak (1963)	15-point scale and other scales (e.g. Muñoz 1986)
Dimensions and products	Texture (foods)	Texture (foods) Skinfeel (lotions, creams, etc.) Handfeel (paper and fabrics, etc.)
SCORING		
	Five-point scale or modified 7-point scale (flavour) Nine-point scale (texture)	Any numerical scale (7, 10, 15, 100 points) Semi-structured and unstructured scales

- Texture evaluations completed by modified/derivative profile methods that follow the original texture profile philosophy and classification published by Szczesniak (1963) to evaluate:
  - mechanical, geometrical and moisture related characteristics
  - attributes throughout stages of manipulation as defined by the texture profile method (e.g. first bite/chew, chew down, residual) (Szczesniak 1963). Examples of manipulation stages for non-food products (e.g. skinfeel evaluations) are: pick-up, rub-out and afterfeel characteristics (ASTM 2003, 2011; Schwartz 1975).
- Consensus measurements from the original flavor profile method as practised by PAA and the A^{5}daptive Profile Method.

Therefore, the original flavor and texture profile methods have provided the key foundations for technical profile-based descriptive evaluations. Their principles and techniques have been used and applied successfully in the modified profile methods listed above and by many sensory practitioners and organizations (Adhikari et al. 2011; Lotong et al. 2000; Meilgaard et al. 2007; Schwartz 1975; Talavera-Bianchi et al. 2009; Thompson et al. 2009).

7.2 Fundamental and Philosophical Principles of Profile/Technical Descriptive Methodology

All profile-based descriptive methods share the following philosophical perspectives and characteristics.

- Emphasis on a detailed and technical terminology/lexicon and extensive use of qualitative references. Extensive use of carefully chosen qualitative references to train on and achieve the most technical lexicon/list of attributes Cairncross and Sjöström (1950) succinctly explained this fundamental principle for the flavor profile method, as follows: 'A direct association with a definite chemical is attempted whenever possible: for example, phenylacetic acid has been a reference odor for a principal aromatic note in both beer and honey'.
- The use of quantitative/intensity references. In addition to the extensive sets of qualitative references, quantitative references are used to train a panel to score intensities accordingly. This profile method approach of using quantitative/intensity references was introduced by founders of the profile methodology and philosophy (Cairncross & Sjöström 1950; Caul & Sjöström 1951; Szczesniak 1963). As Cairncross and Sjöström (1950) explain, 'Measurement of the degree of intensity of any factor can be reduced to quantitative evaluation by the use of outside standards'.
- Complete product information. The profile method's philosophy is to provide an extensive and complete profile/ picture of the product's sensory characteristics. Thus, profile methods focus on all elements of a product, and not just on attributes and intensities.
  - In *flavour*, assessment of aroma, flavour by mouth, aftertaste, overall impression, amplitude and order of appearance (in the original flavor profile and selected modified profile methods).

- In *texture, skinfeel, handfeel*: assessment of all relevant attributes in the different stages of manipulation: surface, first bite/compression, manipulation, residual. Thus, often, the same attribute is evaluated at different stages of manipulation.

7.3 Methodology – The Profile Methods

The approach to select, train and maintain a descriptive panel is presented for the original profile methods followed by modifications made by modified/derivative profile methods.

7.3.1 Screening of Profile Assessors

Training programmes based on profile methods are lengthy and elaborate. Thus, the screening process followed in these descriptive programmes is also involved and meticulous to make sure the best candidates are selected.

When the original flavor and texture profile descriptive methods were developed, assessors were mainly employees. Currently, assessors are primarily recruited from the local community by most organizations establishing a descriptive capability. The process followed in both the prescreening and acuity screening by profile methods is the same for employees and non-employees, except for the step investigating availability. Differences between both are mentioned below, when applicable.

Table 7.3 shows a summary of the panel screening tests completed by the original and modified/derivative profile methods, which are explained below.

7.3.1.1 Prescreening

This step is the same for both original and modified/derivative profile methods.

A main component of screening candidates is acuity, which focuses on the perception of characteristics and scoring ability. However, prior to completing acuity exercises, candidates are prescreened to determine overall interest, availability, ability to work with others and major medical issues and other factors, which might affect product evaluations.

7.3.1.1.1 Overall Interest

Candidates' interest in the training programme and in participating as panel members is explored through a prescreening questionnaire and an interview.

7.3.1.1.2 Availability

The availability of assessors for both the training programme and routine project work must be carefully explored. It is important that candidates be available for all training sessions/orientations, the majority of the practice sessions and for routine project evaluations on a regular basis.

Table 7.3 Panel screening tests for original and modified/derivative profile methods.

Screening test	Original profile methods	Modified/derivative profile methods
PRESCREENING AND NON-ACUITY QUALIFICATIONS		
Overall interest	√	√
Availability	√	√
Limitations (medical, psychological, etc.)	√	√
Personality and group behaviour	√	√
FLAVOUR ACUITY		
Basic taste identification	√	√
Odour recognition test	√	√
Amplitude	√	**X** Only PAA
Identification/independent judgement test	√	**X** Only PAA
TEXTURE ACUITY		
Hardness	√	√ Plus other attributes relevant to product (only in some modified profile methods)
SKINFEEL/HANDFEEL ACUITY		
Selected skinfeel attributes		√
Selected handfeel attributes		√
FRAGRANCE ACUITY		
Fragrance recognition tests		√
SCORING	Ranking	Rating

Exploring availability for employee and non-employee panels differs slightly.

Non-employees

These candidates, who are hired and paid to be assessors, tend to be very reliable and readily available for training and panel sessions. Special attention should be given to:

- *needed availability for whole days*: this scenario must be emphasized, when applicable. Candidates should know when and for how long whole day participation is needed, conditions, etc.
- *projected residency*: personal information cannot always be collected or it might not be legal/appropriate to explore it in the selection process. Whenever possible, it is useful to explore information on type of jobs held, previous moves and length of local residency. Candidates who might potentially move within the first years after the completion of the training should not be recruited.

Employees

Employees' availability is affected by the type of job and work responsibilities. In general, employees who travel frequently, managers/ executives and individuals with heavy workloads should not be recruited. Overall, these individuals have limited availability for panel sessions.

7.3.1.1.3 Limitations (Medical, Psychological, etc.)

All candidates must be in good health and have no medical issues with or illnesses caused by products' ingredients or variables to be evaluated. Thus, individuals with any existing medical/health issues should not be selected. Some examples of elimination criteria include allergies, sensitivities, hypertension, chronic colds and intake of medications affecting discrimination, concentration, etc. In addition, any issues negatively affecting the perception of sensory characteristics to be evaluated must be explored and considered in selecting candidates, as follows.

- *For flavour and fragrance evaluations*: in addition to the general issues described above, for these evaluations, issues such as allergies, frequent colds and sensitivity to high intensities of chemical feeling factors (when applicable) should be investigated and considered as elimination factors.
- *For texture, skinfeel, handfeel evaluations*: for these evaluations and in addition to the general medical issues described above, factors affecting the perception of mechanical and somesthetic/surface properties should be investigated and considered in the screening. Some examples include dentures, orthodontia, implants (affecting oral texture), excessive calluses and carpal tunnel syndrome (affecting manual texture, skinfeel, handfeel), etc.

7.3.1.1.4 Personality and Group Behaviour

Interaction with the candidates to unveil any major personality issues and determine their ability to effectively work with a group is essential in selecting suitable profile assessors.

Profilists interact extensively with each other and work together throughout the training and subsequent project work. They must be flexible and have the ability to work as good team members to contribute to the success of the panel's activities. Thus, it is essential that extreme, dominant or submissive personalities be unveiled and assessed in the selection process.

Personality and group behaviour are even more critical when the original profile approaches are practised because assessors have to work together to reach consensus results.

Conducting personal interviews and group activities are the best techniques to investigate these factors. A one-on-one interview allows the panel leader to notice negative personalities and explore overall interest and motivation. Group activities allow the panel leader to observe the group and team interaction and each participant's ability to work as a team member. One of the best team activities is a mock panel session. The panel leader guides discussions to allow the candidates to interact with each other, exchange ideas and participate in discussions. Through this group interaction, the panel leader is able to gain insights into the panel dynamics and the individual participation in group activities and discussions.

7.3.1.1.5 Other Characteristics

In addition, it is important for profile candidates to possess the following positive characteristics, and thus important to explore (Keane 1992; Muñoz et al. 1992b):

- ability to effectively communicate opinions
- confidence to report perceptions
- sharing personal experiences that might enhance the focus on sensory characteristics
- positive attitude even in the face of adversities and failures.

Modified/derivative profile methods might include other prescreening exercises to aid in the selection of the best candidates, such as exercises to explore articulation, focus and knowledge of sensory characteristics, and scaling ability (ASTM 2006, 2011).

7.3.1.2 Acuity Screening

In profile methods, the acuity exercises are designed to explore physiological factors and ensure that selected candidates have a normal to above average acuity to qualify as profile assessors.

7.3.1.2.1 Original Flavor Profile Method

The process for screening candidates for the flavor profile method is described by Keane (1992), and briefly summarized herein.

The screening tests focus on perception of the main flavour dimensions, such as basic tastes and aroma/aromatics.

Basic Tastes Test

Candidates should be able to identify the four basic tastes: sweetness, sourness, saltiness and bitterness. Solutions are prepared at above-threshold levels (e.g. Sucrose 10%, citric acid 0.1%, salt 0.7%, caffeine 0.1%). Candidates are presented with the four solutions, plus one or two blank samples. Candidates should be able to identify all four basic tastes.

Odour Recognition Test

This test is critical, since odour encompasses most of the notes in flavour evaluations. The test is completed in two steps.

- *Step 1*: ten odorants are presented blind and candidates are asked to identify the characteristics by name or association. For example, a correct response for eugenol would be clove, dentist office or brown spice (fewer points are given for associations).
- *Step 2*: a second set of 10 odorants is provided to complete a multiple choice test for candidates to choose the word that best describes the odorant.

Qualified candidates for the panel are those who correctly identify (or provide a reasonable association to) at least 70% of all odorants.

Amplitude/Arrangement Test

In the original flavor profile method, amplitude is a critical characteristic. Amplitude is the initial overall impression of the aroma and flavour of a product.

It encompasses body and blend (the balance of the detectable and undetectable character notes into the flavour or aroma complex). Amplitude may be influenced by order of appearance (the arrangement of the character notes in the order perceived), by intensity (the strength of the character notes) and also by how rapidly character notes appear and disappear.

The screening includes a test to explore the ability of candidates to understand and perform this integrated measurement. Keane (1992) describes the test as the evaluation of five juices differing in amplitude. Candidates are asked to rank the juices based on a criterion chosen by the candidate (e.g. blend, fullness, sweetness, sourness, etc.). People who correctly rank the products using blend or fullness receive the maximum score.

Identification/Independent Judgement Test

A series of basic taste solutions are prepared at threshold levels in this test. The candidates are given correct and false suggestions for the samples tested. This test ensures that candidates are confident in providing independent judgements, regardless of suggestions.

7.3.1.2.2 Original Texture Profile Method

In the original texture profile method, only one acuity test is conducted (Muñoz et al. 1992b). Candidates are presented with selected samples of the original standard hardness scale in random order and are asked to rank them, that is, to arrange them in increasing order of hardness. Candidates are expected to arrange all items in the correct order to pass the test.

7.3.1.2.3 Modified/Derivative Profile Methods

The screening process used by the derivative profile methods is based on the process followed by the original profile methods. Table 7.3 shows the panel screening tests conducted in modified profile methods, which are briefly explained below.

Flavour – Modified/Derivative Profile Methods

The screening tests used by modified/derivative profile methods are very similar to those used in the original flavor profile method described above, with some variations.

- *Odour recognition test*: candidates receive fewer sets of odorants than in the original flavor profile. For example, in the A^{5}daptive Profile Method, only one set of odorants is used (8–15 odorants). However, in this method, another set of samples might be prepared with other more complex materials, such as finished products (beverages, foods, fragrances, lotions, household products) that are relevant to the training.
- *Amplitude*: an acuity test for this attribute is not conducted, since this is not an attribute measured by modified/derivative profile methods.
- *Identification test*: this acuity test is not conducted by modified/derivative profile methods.

Fragrance – Modified/Derivative Profile Methods

For fragrance profile training programmes, odour tests are an essential part of the screening, and one or more odour/fragrance recognition tests are completed.

The odorants in this case are mainly fragrances or fragrance ingredients/components, and not food components/ingredients. However, odorants common to foods and fragrances might be used, such as citrus (orange, lime, lemon, etc.), other fruity notes (tropical, apple, peach, etc.), sweet (vanillin), spices (clove, anise), etc.

In addition, other important odorants/ingredients in fragrances might be presented in these tests, such as floral (violet, jasmine), woody (sandalwood), etc.

Texture – Modified/Derivative Profile Methods

Similar to the original texture profile method, most modified/derivative profile methods only include one or two texture attribute (e.g. hardness) in the screening.

Skinfeel and Handfeel – Modified/Derivative Profile Methods

Similar to the screening process completed in food texture modified/ derivative profile panels, the screening for skinfeel or handfeel panels includes relevant attributes in those dimensions. Candidates are asked to rate the intensity of attributes for the samples specifically chosen to represent the intensity ranges of the attributes tested. For both skinfeel and handfeel, one attribute from each evaluation stage might be chosen.

- *For skinfeel evaluations*: appearance (e.g. integrity of shape), pick-up (e.g. firmness or stickiness), rub-out (e.g. ease of spreading or thickness) and afterfeel (e.g. greasiness or amount of residue) (ASTM 2011).
- *For handfeel evaluations*: mechanical characteristics (e.g. stiffness), geometric characteristics (e.g. roughness) and, if relevant, moistness.

7.3.1.3 Scoring Screening Exercises

7.3.1.3.1 Original Flavor and Texture Profile Methods

Both the original flavour and texture profile methods use ranking in screening tests (Keane 1992; Muñoz et al. 1992b).

7.3.1.3.2 Modified/Derivative Profile Methods

Because of the additional information provided by intensity versus rank data, most derivative profile methods use scales. Therefore, in the screening, candidates are asked to complete rating exercises. Some other scoring variations might be used in modified profile methods. For example, in the A^{5}daptive Profile Method, intensity references are presented in one exercise for assessors to rate intensities relative to those references.

7.3.2 Staff Requirements in Profile Programmes

There are three important staff members required in profile training programmes.

7.3.2.1 Profile Trainer

The trainer is a key professional in profile training programmes. This person might be part of the organization or a consultant, and can assume different responsibilities depending on how the training is set up.

In modified/derivative profile methods, this person might also be an experienced panel leader who has previously trained panels. In this case this person assumes the role of trainer and panel leader.

The main responsibilities of the trainer might include several or all of the following activities.

- Present the benefits and characteristics of the training programme to management or stakeholders.
- Design the training programme.
- Lead orientation/training sessions to teach the panel the fundamentals of the profile methodology, sensory dimensions (e.g. flavour, texture, skinfeel, fragrance, etc.) and scaling.
- Lead the development of the terminology or lexicon during training.
- Train the panel leader, if applicable.
- Design the training and practice exercises, and assist in or set up the data collection and analysis systems.
- Retrain the panel or specific assessors, as needed.

7.3.2.2 Profile Panel Leader

The panel leader is a key team member in all profile descriptive programmes. The role of the panel leader is slightly different in original profile versus modified/derivative profile programmes.

7.3.2.2.1 Panel Leader's Role in Original Profile Programmes

In the original flavor and texture profile methods, the main function of the panel leader is to facilitate the panel sessions. The panel leader is an integral part of the panel and provides:

- background on the project at hand
- products for orientation purposes
- sources for reference materials
- assistance in reaching consensus
- composite of results
- final profile and detailed report.

During the initial training of a profile panel, all trainees take turns playing the role of panel leader for at least one practice session. Following this strategy, all profilists experience and appreciate the work that is involved in that position. After training, one person is usually selected to continue in that role. This selection may be based on their position in a company (e.g. a sensory professional) or on particular abilities demonstrated during the training process.

A profile panel leader should possess good leadership skills to be able to:

- get a diverse group to work toward consensus, and resolve conflicts
- maintain a collegial atmosphere during discussions, encouraging participation by all assessors
- provide resource materials and products for references using creative approaches.

In addition, the profile panel leader should have good communication skills to be able to:

- composite the individual profiles into the final and detailed product profile
- interpret and communicate sensory information in a way that is understandable to the recipients.

7.3.2.2.2 Panel Leader's Role in Modified/Derivative Profile Programmes

In most modified profile methods, the experienced panel leader not only guides but teaches and leads the panel, particularly in lexicon/ ballot development, and in retraining, when issues arise.

The following qualifications are desirable in a successful modified profile panel leader.

- Technical knowledge in the relevant sensory dimensions (e.g. flavour, texture, skinfeel, etc.). Thus ideally, a derivative profile panel leader should be a profilist previously trained in the sensory dimension(s) of focus and the profile method.
- Leadership and communication skills. A modified profile panel leader should have good leadership and communication skills to successfully lead the discussions during training and panel sessions.
- Availability. The panel leader must be available to fully participate in all screening and training activities, and to give the panel attention and care.
- Knowledge of sensory methods and statistical analysis approaches. This is particularly important if the panel leader also acts as a project manager and is in charge of data analysis and reporting.

7.3.2.3 Technician

The technician's main role is to assist the panel leader in the acquisition and preparation of all materials, products and references needed during the training and practice exercises. The technician is also responsible for the distribution of samples and references in all profile panel sessions (training, practice and regular evaluations).

This staff member may also have other responsibilities, such as completing the data entry and some data analyses, scheduling assessors and sometimes leading selected panel sessions. In some organizations the technician's responsibilities are undertaken by the panel leader.

7.3.3 Profile Descriptive Analysis Training Approach

The training process in the original and modified/derivate profile methods is lengthy and involved. As described earlier, the fundamental philosophy of the profile methodology is to develop a technical panel that is qualitatively very strong. Thus, in order for a panel to be able to provide the most technical and detailed descriptive information, an involved and technical training is needed.

In general, the training approach followed in the original and derivative profile methods has the same fundamental principles and steps. Variations exist to cover specific topics unique to each method. For example, the original flavor profile method incorporates modules to train assessors in amplitude and order of appearance. These modules are not part of most derivative profile training programmes.

A description of the general process to train either an original or a derivative profile panel is provided below. Differences are highlighted when they exist. In addition, and for illustration purposes, Table 7.4 shows a comparison of the training approach followed by an original and a modified/derivative profile method (original texture profile method and skinfeel derivative profile method). This table allows the reader to:

- recognize the importance of the original profile methods' contributions. The flavor and texture profile methods provided to the sensory community the fundamental principles of the profile descriptive methodology, which are used by all modified/derivative profile methods
- understand the way modified/derivative profile methods apply the principles of the original flavor or texture profile methods in their modified versions
- realize the similarities in philosophies between the original and modified/derivate profile methods.

A well-designed profile descriptive training programme includes several orientation/training and practice sessions, which are skilfully designed to accomplish the training level desired.

The following is a description of the general process to train a profile descriptive panel. The reader is also encouraged to review the milestone profile publications describing the training process (Brandt et al. 1963; Caul 1957; Keane 1992; Muñoz et al. 1992b; Neilson et al. 1988).

7.3.3.1 Profile Training Sessions

Several orientation/training sessions are completed by the profile methods.

7.3.3.1.1 First Training Session

This is a very important training session, since the fundamentals of the profile methods and evaluations, terminology or lexicon development, scoring and overview of sensory methods are covered. This initial phase usually takes 3–4 full days.

In general, this initial training session covers:

- sensory evaluation principles (methods and applications)
- fundamentals of descriptive analysis and profile methodology

Table 7.4 Comparison of the original and modified/derivative profile panel training approaches. Example illustrates a texture training approach (original texture profile method) and a skinfeel training approach (modified/derivative profile method).

Step	Original texture profile method following recommendations by Brandt et al. (1963) and Muñoz et al. (1992b)	Derivative profile skinfeel profile method summarized from ASTM (2011)
Prescreening	Interest, availability, general good health, ability to describe differences, ability to apply abstract concepts, positive attitude and ability to work with others	Same as in the original texture profile method
Panel screening	One acuity texture exercise (hardness)	Several acuity texture/skinfeel exercises, usually one of each category/stage: appearance, pick-up, rub-out and afterfeel
Panel training (orientation/ training sessions)	TRAINING COMPONENTS • Orientation/training sessions • Practice sessions APPROACH • Discussion of controlled evaluation procedures • Introduction to texture attributes relevant in the training/products, perceived in the stages of evaluation (surface properties, first bite/chew, chew down, residual) • Definition of texture attributes • Illustration of texture attribute through intensity reference standards ENSUING TRAINING SESSIONS • Same approach as in the first training • Reinforce concepts and review intensity scales • Include specific products of interest and review of the complex texture attributes	TRAINING COMPONENTS Same as in the original texture profile method APPROACH Same as in the original texture profile method, adapting techniques to skinfeel attributes • Discussion of the strictly controlled procedures • Introduction to all skinfeel attributes in the stages of evaluation (delivery, pick-up and rub-out) • Definition of skinfeel attributes • Illustration of skinfeel attributes through intensity reference standards ENSUING TRAINING SESSIONS Same approach as in the first training • Reinforce concepts and review intensity scales • Include specific products of interest and review of the complex skinfeel attributes
Panel training (practice sessions)	Several sessions to review: • texture attributes and intensity reference scales • strict evaluation procedures • evaluation of food products	Same as original texture profile Several sessions to review: • skinfeel attributes and intensity reference scales • strict evaluation procedures • evaluation of skinfeel products

Table 7.4 (Continued)

Step	Original texture profile method following recommendations by Brandt et al. (1963) and Muñoz et al. (1992b)	Derivative profile skinfeel profile method summarized from ASTM (2011)
Examples, attribute definitions and procedures	FIRMNESS and DENSENESS (first chew) Procedure: Place product sample (e.g. 1 in^2) between molars and bite completely 2–3 times. Evaluate: Firmness: Force required to fully bite through product (no force/soft – high force/firm) Denseness: Degree to which product feels dense/compact when chewing it 2–3 times (low compactness/ airy – high compactness/denseness)	FIRMNESS (pick-up) Procedure: Scoop premeasured product from container and compress product slowly between index finger and thumb 3 times Evaluate: Firmness: Force required to fully compress product between thumb and forefinger (no force – high force) DENSENESS (rub-out) Procedure: Spread the premeasured amount over inscribed surface of forearm using gentle circular motions and the designated rate/strokes per second. Evaluate: Denseness: Amount of compactness perceived after rubbing product on skin (3 rubs) (no compactness/ airy – high compactness/denseness)

- principles of sensory dimensions being taught:
  - aroma/aromatics, basic tastes, trigeminal sensations (in flavour)
  - mechanical, somesthetic/geometrical and fat/moisture attributes (in texture, skinfeel, handfeel)
  - aroma/odour (in fragrance and odour evaluations)
- scoring principles, intensity/quantitative references and their use in scoring.

In addition, and representing one of the most important activities in this first orientation, the language/lexicon development and ballot generation are covered. Section 7.3.3.2 describes the process. The activities completed in this step differ in the original and derivative profile training programmes as follows.

- In the original flavor profile method:
  - development of terminology (see section 7.3.3.2)
  - introduction to consensus judgements.
- In the original texture profile method: review of reference scales to teach the panel texture attributes and their definitions, their evaluation procedures, and attribute intensities through quantitative/intensity references.

- In modified/derivative flavor or fragrance profile methods:
  - lexicon development process (see section 7.3.3.2)
  - ballot generation process based on the lexicon developed
  - simple evaluations to start familiarizing assessors with evaluation practices and use of ballots.
- In modified/derivative texture, skinfeel, handfeel profile methods:
  - review of reference scales to teach the panel texture/skinfeel/handfeel attributes and their definitions, their evaluation procedures and attribute intensities through intensity references
  - lexicon development process (see section 7.3.3.2) for one or a few products with simple textures, by defining the manipulation stages and the attributes to be evaluated in each stage
  - simple evaluations to start familiarizing assessors with evaluation practices and use of ballots.

7.3.3.1.2 Subsequent Training Sessions

Subsequent training sessions are completed after the first practice sessions have been conducted. The number of sessions needed depends on the programme design, product categories and philosophy of the method. In general, one to three subsequent sessions are conducted, for a total of 2–4 training sessions. These sessions are usually 2–4 full days.

Examples of the activities completed in these subsequent profile training sessions include:

- training on more complex products
- introduction and evaluation of complex sensory attributes: for example, in the original flavor profile: amplitude, balance and fullness; in the A^{5}daptive Profile Method: balance, blendedness, etc.
- further work on consensus (if applicable)
- panel work to address questions and issues encountered during practice, etc.

7.3.3.2 Key Training Step: Profile Terminology/Lexicon Development

This is a critical step in the training programme and project work of a profile panel. Given its importance, the topic is discussed in depth in this section.

The process and principles described below are taught to the panel during the first orientation. The approach is complex and is repeated during practice sessions and ensuing orientation/training sessions, until learned by the panel. In project work, the terminology/lexicon development represents the most important activity, and thus must be mastered by the panel prior to initiating projects.

There are three steps in the process of terminology/lexicon development. The first two steps are common to both the original flavor/texture profile and modified/derivative profile methods. The approach differs only in the third step and the differences are explained below. In addition, Table 7.5 shows an example of the terminology/lexicon development process and illustrates differences between the original flavor profile and a modified profile approach (A^{5}daptive Profile Method).

Table 7.5 Terminology/lexicon development – the original flavor profile and a modified profile method (A^{5}daptive Profile Method). Example: cheese/herb crackers.

Step	Panel activities	Example of references and products used
Review of samples/ product category for individual terminology development (both methods)	1) *Category Review Design* Careful selection or production of 4–8 diverse samples 2) *Category Review Evaluation* Assessors independently evaluate the set of samples and generate individual terminology 3) *Brief discussion* of individual terms. Preliminary elimination, grouping and identification of flavour families 4) *Planning of References* a) Panel leader refines the list of references planned for the next step based on group discussion (A^{5}daptive Profile Method®) b) In the original flavor profile, assessors themselves determine the needed references	Commercial and prototype samples displaying the range of flavour dimensions of interest (e.g., different cheese blends and herbs and spices)
Review of references and development of group/consensus terminology (both methods)	1) *Review of Qualitative References* References (carefully selected to show the main product category's flavour dimensions) are reviewed and discussed 2) *Development of Group Terminology/ Lexicon* a) Discussion of references vis-à-vis individual terms/lexicons and disagreements b) Clarification of terms with aid of references c) Generation and discussion of group terminology/lexicon	Examples of reference sets reviewed: • Samples varying in toast level • Cheeses and dairy references (e.g., cheeses in blend, diacetyl, butyric and isovaleric acids, etc.) • Herbs and spices (e.g., green herbs (like oregano), black pepper, etc.) • Other (HVP, oxidized oil)
Use of consensus terminology (a) Original flavor profile Evaluation of products, discussion of results and reaching consensus on terminology and product profile	1) *Product Evaluation* Working independently and with a blank sheet, assessors evaluate one product, based on the consensus terminology 2) *Discussion of Results and Review of Additional References* Panel discusses results and identifies areas of potential disagreement (usually 'order of appearance' or intensities). Previously reviewed and new references are examined and used in the discussions	• Products (two of the most different cheese crackers) • References and products previously evaluated, as needed • New references [e.g., more green herbs, more HVPs, chili powder, off notes (e.g., cardboard)]

(*Continued*)

Table 7.5 (Continued)

Step	Panel activities	Example of references and products used
	3) *Refinement of Terminology and Final Profile* Assessors again examine the same product along with the written composited profile and discuss and agree on the character notes/reach consensus 4) *Validation* A second product (most different from first) is profiled and the 'side by side' comparison of finalized profiles is used to confirm that the major differences between samples are described completely by the panel	
Use of consensus terminology (b) A[5]daptive Profile Method Evaluation of products, discussion of results and lexicon validation	1) *Product Evaluation* The panel evaluates two products using a ballot which incorporates the group/consensus lexicon 2) *Discussion of Results and Review of Additional References* Panel discusses results and identifies attributes requiring additional discussion. Previously reviewed and new references are examined and used in the discussion and refinement of terms 3) *Refinement of Lexicon* Based on the discussion and review of previously presented and new references, required lexicon/ballot changes are made 4) *Validation* The same or a different pair of products is evaluated. Results are discussed to ensure that the ballot is adequate for the evaluation of the product category	• Products (2–4 different cheese crackers) • References and products previously evaluated, as needed • New references [e.g., more green herbs, more HVPs, chili powder, off notes (e.g., cardboard)]

7.3.3.2.1 First Step: Generation of Individual Terminology

A carefully selected set of samples is used in this step, which should demonstrate the key sensory characteristics/variables of the category. These samples should be sufficiently different and span the sensory space of interest.

In this first step, the members of a profile panel review the samples to generate their individual and independent terminology on a blank sheet of paper. This process is difficult at the beginning, especially in flavour and odour/fragrance evaluations, because assessors lack the focus and sensory terminology to describe perceptions. With every exercise, and as the training progresses, the process becomes easier for assessors, since they learn new terms and sensory concepts and are more focused. In the original texture profile and modified texture, skinfeel and handfeel methods, this process occurs after the panel has learned the texture (or skinfeel, handfeel) attributes and scales. With this knowledge and to generate the lexicon for the product category, each panel member individually chooses the sensory characteristics that apply to the products reviewed from the attributes and terms they learned.

7.3.3.2.2 Second Step: Generation of Consensus Terminology

In this second step, the profile panel leader guides a discussion to review all individual terms. This step's objective is to arrive at a consensus/group terminology which is understood and accepted by all panel members, and there is agreement on the attribute names and definitions.

The success of this process is based on:

- the richness/completeness of the individual terminology generated by assessors
- the ability of the group to efficiently and co-operatively assess all individual terms and references, and arrive at a consensus terminology
- the presentation of appropriate products and references needed to clarify terms and to select the most technical and appropriate attribute names.

Several sessions might be required to complete this process. The length of the process depends on the level of expertise of the panel (initially, the process is longer because assessors lack expertise) and the complexity of the product category.

7.3.3.2.3 Third Step: Validation and Use of Consensus/Group Terminology

Original Flavor Profile Method

In this third step, the flavor profile panel validates the terminology by completing a simple evaluation and generating a product consensus profile. In this exercise, profile assessors trained in the original flavor profile method start again with a blank sheet of paper and generate their individual profiles based on the terms and discussions held in step 7.3.3.2.2 (generation of consensus terminology).

Individual profiles provide information on the product attributes in the perceived order of appearance and their corresponding intensities.

A group discussion follows to review the individual profiles and reach a consensus on the final/group product profile (Krasner et al. 1985). Also, any issues encountered in this validation exercise are discussed and any necessary modifications in the terms or definitions are incorporated. Upon completion of this step, the flavour terminology is considered final for the product category and to be used in future projects.

Original Texture Profile Method

In this third step, validation of the texture lexicon and evaluation procedures, the panel leader summarizes all individual information and leads a group discussion for the panel to develop the texture evaluation procedures for the product category, which include:

- the product manipulation stages to be included, such as first bite, first chew, chew down, residual
- the attributes evaluated under each manipulation stage
- the attribute definitions
- the evaluation procedures for the attributes to be evaluated.

Upon completion of this step, the texture terminology and procedures are considered final for the product category and to be used in future projects.

Modified/Derivative Profile Methods

Muñoz (in Moskowitz et al. 2003), in the chapter entitled 'Language development in descriptive analysis and formation of sensory concepts', lists the three steps followed in the lexicon development process for modified descriptive profile methods.

- Review of samples/product category for individual lexicon development (similar to the process described in section 7.3.3.2.1).
- Review of references and development of common/group terminology/lexicon (similar to the process described in section 7.3.3.2.2).
- Evaluation of products, discussion of results and lexicon validation.

This last step represents the validation and use of consensus/group terminology and is completed as follows.

1 *Ballot development*

In modified flavour and fragrance profile evaluations, the ballot includes the consensus/agreed upon lexicon and a scale for each of the attributes.

In modified texture, skinfeel, handfeel evaluations, the ballot consists of:

- the product manipulation stages to be included; for example, for textures: manual evaluations, first bite/chew, chew down, residual; for skinfeel: pick-up, rub-out, afterfeel
- the attributes evaluated under each manipulation stage
- the attribute definitions
- the evaluation procedures for the attributes to be evaluated.

2 *Validation of ballot and procedures*

Once the ballot is finalized, Muñoz (in Moskowitz et al. 2003) describes the following activities to validate the ballot and procedures.

- The evaluation of products using the generated ballot.
- The discussion of results and ballot adjustments. Based on the results and their discussion, the panel might choose to modify the ballot to address the issues unveiled by the data (e.g. change attributes and definitions, the ballot structure, etc.).
- The validation of the lexicon and evaluation procedures. Through the process above, the ballot is validated, that is, it is confirmed that the ballot is adequate for descriptive evaluation of the product category.

The above process followed by the original flavor and modified profile method is illustrated for a food product category (cheese crackers) in Table 7.5.

7.3.3.3 Profile Practice Sessions

The main objective of these sessions is to apply and practise all principles learned in the training sessions. A series of review exercises and evaluations is planned by the trainer or panel leader for these practice sessions. It is important that these exercises include the evaluation of products of increasing difficulty (Keane 1992).

The number of sessions needed depends on the programme design. There might be 10–20 practice sessions scheduled after each training session, and they may last 1–2 h/day. Fewer practice sessions are needed if the sessions are 3 h/day.

Results are summarized and, when applicable (in modified/derivative profile methods), statistically analysed. Feedback (either consensus data or data statistically analysed) is provided to the panel to show its progress and address any issues that need to be revisited (i.e. attribute definitions, qualitative and quantitative references, scaling issues, etc.).

7.3.3.4 End of the Profile Training and Validation

A profile training programme is completed with a validation study.

Samples to be evaluated by the panel are carefully selected in designing the validation study. Since the panel is now trained, samples should be sufficiently but not extremely different. One or several sets of samples can be included in the validation study.

Results are analysed/assessed to determine the readiness of the panel for project evaluations, or if more training is necessary.

When samples with known variables are selected, it is possible to assess the panel performance and determine if the panel is ready for routine projects. A profile panel is expected to show differences in the attributes known to be different.

Other comparisons that can be completed to validate the newly trained modified/derivative profile panel are:

- comparison of results from the recently trained modified/derivative profile panel with results from an experienced/trained panel. It is expected that the results show the same direction of difference in at least 80% of the attributes
- as per ASTM (2011), comparison of results from the recently trained panel with consumer attribute data for those attributes for which consumers have demonstrated their understanding and the ability to differentiate among products. If the panel data correlate well with the previously validated consumer responses for similar attributes, the recently trained panel is considered trained and reliable.

7.3.4 Profile Project Work

A brief description of project activities is provided by Keane (1992) and Muñoz et al. (1992b).

7.3.4.1 Profile Panel Pre-Project Activities

In general, prior to completing formal evaluations (or finalizing the profile), the following activities are conducted.

7.3.4.1.1 Development or Review of Terminology or Ballot

If the panel has previously developed the terminology (original profile methods) or the ballot, only the review of this information vis-à-vis a selected subset of products is sufficient prior to a project evaluation. Any necessary adjustments are incorporated prior to the evaluation (or generation of the final profile).

However, if the panel does not have previous experience with the test products, the terminology or ballot is developed following the same approach used in training and practice exercises (see section 7.3.3.2). In general, the process followed in project work is shorter and easier since the panel is already trained.

7.3.4.1.2 Review of References

Prior to project evaluations, the panel leader and panel determine which references need to be reviewed. When the panel has experience with the product category, few references are needed (in most cases, only the references for complex attributes, or for characteristics posing some difficulty to the panel).

A more extensive set of references should be reviewed when the panel must develop new terminology (original profile methods) or a new ballot, as done during the training (see section 7.3.3.2).

7.3.4.1.3 Initial Profile Development/Practice Evaluations

If deemed necessary, additional sessions might be conducted to either develop an initial profile (original profile methods) or complete a practice evaluation with a small subset of products. The panel discusses the profile generated or their results to make any necessary adjustments in the terminology or ballot.

Once these activities are completed and the panel leader considers that the panel is ready to initiate the project, the formal evaluation sessions can be scheduled.

7.3.4.2 Product Evaluations: Profile Descriptive Analysis

7.3.4.2.1 Original Flavor Profile Method

To complete the project's product evaluations, the preliminary group profile developed and discussed during the pre-project sessions or available from previous projects is used. This profile is refined until all assessors agree and a final composite judgement is reached, representing the final profile (Keane 1992).

7.3.4.2.2 Original Texture and Modified/Derivative Profile Methods

The execution of the project product evaluations is a simple process, once all pre-project activities have been completed (see section 7.3.4.1). Assessors evaluate the test products in the order presented, and using the final ballot developed in the pre-project sessions. In some modified profile methods, one to two replications are completed. Data are summarized and prepared for data analysis.

7.3.4.3 Analysis/Review, Interpretation and Report of Profile Results

7.3.4.3.1 Data Collection and Analysis

In the original profile methods, the panel meets for several sessions to discuss individual profiles, generate the consensus profile(s) and refine it/them. The process is complete when all assessors are in agreement with the profiles generated, the attributes and intensities, and when determined that the profiles reflect the differences and similarities among the products tested.

As previously indicated, consensus data are collected in the original profile methods. Thus, results are not statistically analysed.

In modified/derivative profile methods, assessors individually rate the products' attributes and thus the data are statistically analysed. The statistical analysis completed depends on the test design and the study's characteristics. In general, all modified profile methods use analysis of variance (ANOVA) to test for sample/product effects.

7.3.4.3.2 Report of Results

Reports in descriptive analysis are not method specific. Thus, no standard or special format is presented for profile descriptive analysis reports.

In these authors' opinion, a descriptive analysis report should have the following components.

- Executive summary
- Background
- Project objective
- Test objective
- Sample identification, and special preparation or presentation procedures
- Sensory test design
- Evaluation procedures (especially important in texture/skinfeel/handfeel)

- Terminology/lexicon
- Ballot (in modified profile methods)
- Qualitative and quantitative references used
- Results in tabular or graphical form
- Summary of data analysis (in modified/derivative profile reports)
- Discussion and interpretation of results
- Conclusions and recommendations
- Appendices

Some of the above information might be best presented in an appendix, such as the ballot, attribute definitions and references used.

7.4 Advantages and Disadvantages of Profile Methods

Figure 7.1 shows a summary of the advantages and disadvantages of the original and modified/derivative profile methods.

7.4.1 Advantages

The main advantages of the profile methods listed below are the result of the detailed and technical nature of the data generated. Profile descriptive analysis produces the most detailed and technical qualitative and quantitative characterization of products.

- Very technical, detailed and actionable product guidance provided to users due to the nature of the profile results. For example, a flavor profile panel evaluates and provides information on very specific flavour characteristics, and if needed, using chemical descriptors, such as benzaldehyde, butyric acid, etc. Therefore, the information on product differences and similarities described in those technical terms provides specific and actionable product guidance.
- In general, a well-trained and calibrated profile panel might be more discriminating and descriptive, compared to panels trained using other descriptive philosophies/approaches. This increased discriminative ability is the result of:
    - the comprehensive and detailed terminology. Other methods which do not invest the training time needed to develop a panel capable of providing detailed qualitative characterization might miss attributes that distinguish products
    - the use of quantitative references, which yield lower within panel variability
    - the extensive group interaction and discussions that often original profile and some modified profile panels conduct to compare products. Through consensus, products might be differentiated that may appear similar by ballot but have subtle differences.
- Since the original profile and some modified profile methods use quantitative references, the data are more stable across evaluations, offering the following main advantages.
    - The possibility of accurately comparing data from different panels and different time intervals.

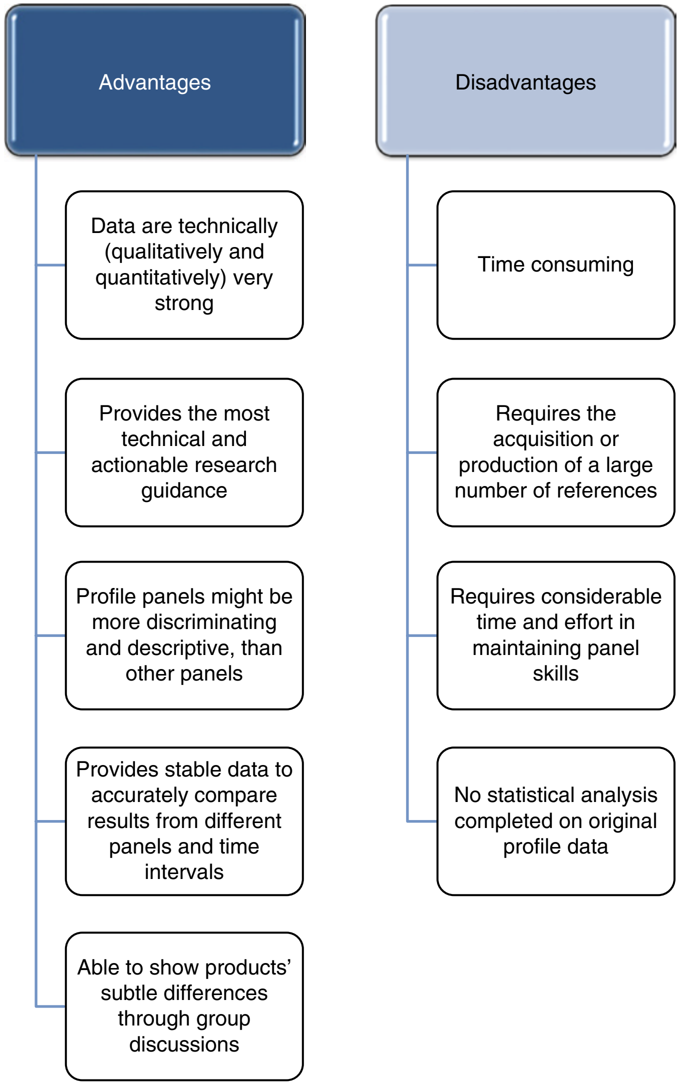


Figure 7.1 Advantages and disadvantages of the original and modified/derivative profile methods.

- The detailed characterization of a product at a given point in time. This information is extremely useful in representing a product that might no longer be available, or an ideal product for a target.

7.4.2 Disadvantages

The main disadvantages of profile programmes are the result of the characteristics that yield the main advantages, that is, the time and effort involved. The rigorous profile trainings are long and elaborate in order to establish a panel that can complete the detailed and technical characterization of the products' sensory properties as described in this chapter. Also, a large number of references are required to be able to (1) demonstrate nuances and details of characteristics (qualitative references), and (2) achieve uniform scale usage and calibrate the panel (quantitative references).

In addition, time and effort are needed to maintain the panel skills. While most traditional descriptive methods incorporate a maintenance programme, more work is required to maintain the skills of a profile panel because assessors need to periodically review quantitative references to stay calibrated.

A disadvantage of the original flavor and texture profile methods is that results are not statistically analysed. This is considered a disadvantage by those professionals who wish to support their findings only through data that are statistically analysed. Although original profile methods yield consensus and not individual data, they produce reliable, technically strong data that deliver the same conclusions as modified profile methods.

7.5 Applications: Profile Descriptive Analysis

Original and modified/derivative profile data have many applications. Since profile methods provide technical information, and base their ratings on quantitative/intensity references (which yield absolute instead of relative quantification of intensities), the most important applications of profile data are in projects where one or both of the following properties are important.

- A detailed and technical qualitative product characterization. Original and modified profile data are the most detailed and technical compared to other descriptive methods' data. Therefore, results from a well-trained and calibrated profile panel are able to show differences among similar products given the comprehensive and detailed evaluations. This feature is important in several applications such as product guidance for product development, consumer data interpretation, category appraisals, etc.
- Stable and absolute (not relative) quantitative product characterization. This applies to the original flavor and texture profile and those modified profile methods that use quantitative references. The attribute ratings are not relative to the samples in the session/set, because assessors rate intensities based on quantitative references.

Given the above characteristics, profile data have the following applications.

7.5.1 Research and Development (R&D)

The main application of profile data is to provide technical product information to guide researchers/product developers in their research work. Since profile data (from original and modified/derivative methods) provide the most technical qualitative and quantitative characterization of products, they are preferred by many research groups for the actionable guidance they provide. Within R&D, profile data are successfully used in the following projects.

- Product research guidance during formulation (product development) and reformulation (product improvement) of products
- Product research guidance in the optimization of products

- Research guidance for product matching or differentiation
- Shelf-life studies
- Product maintenance
- Approval of alternative ingredients, materials and processes
- Category appraisals and characterization of competitive and own products
- Understanding and interpretation of consumer data
- Product documentation for other objectives such as:
  - characterization of the product differences found by discrimination tests
  - characterization of standards, prototypes and competitors
- Assessment of instrumental data through data relationships

7.5.2 Quality Control/Assurance

The applications of descriptive data in this field are covered by Muñoz et al. (1992a). Descriptive analysis data are key in QC and QA/sensory projects, and are used at different stages of programme implementation and operation, such as in:

- product variability studies
- documentation of variable attributes and ranges
- documentation of standards and targets
- establishment of QC/sensory specifications
- selection of training samples
- panel monitoring
- trouble-shooting production problems.

7.5.3 Marketing and Market Research

Through the technical characterization of products obtained by profile panels, market researchers who use profile data in their projects are able to:

- design sound consumer questionnaires by using the sensory products' profile data, which accurately describe their sensory attributes
- select the most appropriate test products for consumer tests through assessment of the products' sensory profiles
- design the best consumer tests (e.g. based on the product's profiles, be able to make the best decisions on number of products, product presentation design, etc.)
- interpret consumer data
- interpret category appraisal results and better understand the product space and the sensory position/advantages of own and competitive products
- support advertisement claims using the descriptive characterization of products of interest in the claim.

7.6 General Recommendations: Profile Methods

Throughout this chapter, the needs and procedures for implementing a profile programme have been described. This section summarizes the most important recommendations to successfully design and establish a profile descriptive programme.

7.6.1 Management Support

The establishment of a profile descriptive programme represents a considerable investment for a company. Section 7.3 (methodology) described all the requirements to implement a programme, including resources and time commitment.

In seeking long-term support for the programme and to justify the investment, it is extremely important to present the information on requirements and needed resources to management vis-à-vis the advantages and applications of the profile programme (see sections 7.4 and 7.5). If appropriate, information on comparison of the profile methods with other descriptive methods (Chapter 20) should be provided, focusing on the advantages of the profile methods that are relevant to the company.

Periodically, sensory staff in charge of the descriptive programme should update management on the success of the profile programme and the impact/assistance that this tool has offered to different groups and in critical projects in the organization. Continued management support is required to successfully maintain the established profile programme.

7.6.2 Staff and Assessors

Management support is essential to be able to implement and maintain a profile descriptive programme. A second critical element in a successful profile programme is the people (staff and assessors) involved.

7.6.2.1 Qualified Trainer and Support Staff

The success of the profile programme greatly depends on the qualifications and ability of the sensory staff to effectively implement and maintain the programme, as described in section 7.3. A qualified trainer is essential to ensure that the technical concepts and procedures of the profile methods are effectively learned, and that adequate guidance is provided to the panel throughout the training. When internal resources or expertise are not available, an outside professional/consultant can provide the needed support.

In addition, the successful completion of the technical and involved profile training and practice sessions depends on the ability of the sensory support staff to carefully complete all the preparation and activities required. These activities include the scheduling of assessors, and the acquisition and thorough preparation of samples and references. Therefore, in addition to the trainer and panel leader, it is essential that the necessary support staff be available to complete these activities for a successful outcome. When support staff/technician(s) cannot be available, the above activities need to be completed by the panel leader. In this case, it is important that these additional panel leader responsibilities be communicated and planned for.

7.6.2.2 Qualified Assessors and Maintenance Programme

Due to the technical nature of a profile programme and the investment to train assessors, it is essential to select the most qualified assessors. The training process is easier and potentially shorter when qualified assessors are selected. Therefore, it is important to invest the time in the screening process to identify the individuals who are best qualified and have the greatest interest, by completing the activities described (see section 7.3.1) and others that can aid in the selection process.

The completion of the profile training does not mark the end of the programme and learning. A maintenance programme is required to ensure that the panel continues growing and providing quality data. The nature of the sessions completed for this purpose depends on the panel and the issues at hand. In general, in these sessions, product evaluations for practice or feedback purposes, review of existing ballots and references and feedback to the panel are completed. The necessary activities are designed to ensure that the profile panel remains calibrated and motivated, their needs are met and performance is maintained.

7.6.3 Training Materials/References

Emphasis has been given throughout this chapter to all the training materials needed to achieve the profile programme's desired technical level. Also it has been acknowledged that this represents a disadvantage of the method.

The sensory group implementing a profile programme should be aware of this very important programme need and ensure the availability of the required resources (i.e. staff, the support of other groups within the organization for the references' acquisition or preparation, and expenses incurred in purchasing materials).

7.6.4 Panel Attrition

Early on in the profile programme's planning and as part of its design, the sensory group should create a strategy to maintain an adequate panel size. In general, and disregarding those unique cases where assessors resign/stop participating at the onset, panel attrition may start occurring 2–3 years after completion of the training. A programme for the addition/training of new assessors should be designed in the preliminary stages in order to be implemented when the panel size decreases. Through the completion of individual/one-on-one training or training of a group, new assessors are added to the existing panel.

7.7 Modifications of Original Profile Methods

7.7.1 Modifications Driven by Philosophy or Need for Updated Methodology

The original flavor and texture profile methods, together with the QDA method, are the fundamental descriptive methods which provided the foundations of descriptive analysis. This chapter has described how modified profile methods use many fundamental practices developed by the original profile methods (e.g. technically strong terminology, extensive use of both qualitative and quantitative/ intensity references, etc.).

Sensory practitioners considering the profile methodology can implement the original profile methods or modify them. The desired modifications may only be a few or several, in which case one of the modified profile methods can be implemented (e.g. PAA, QFP, A^{5}daptive Profile Method, Spectrum Method, etc.).

To summarize concepts from this chapter, the list below shows several modifications of the original profile methodology that have been implemented by sensory practitioners or are intrinsic to the modified profile methods listed above. These changes are driven by philosophy or the need for updated methodology, and not by limited resources (described in section 7.7.2).

- Use of ballots (listing the attributes selected by the panel in a defined order)
- Evaluation of other sensory dimensions (e.g. appearance, sound, etc.)
- Evaluation of fewer attributes
- Use of other scales, and not the original seven-point expanded original profile scale
- Statistical analysis of the descriptive data
- Involvement of and guidance by a panel leader
- Larger panel size

7.7.2 Modifications Driven by Limited Resources

Modifications to the original flavor and texture profile methods are necessary with limitations, such as limited time, funds, staff, etc. Below is a list of recommendations to consider with such limitations. The reader is referred to other chapters in this book to learn how specific modified profile methods (QFP, A^{5}daptive Profile Method, Spectrum Method) address limited resources.

7.7.2.1 With Time and Budget/Funds Limitations

When limited funds or time preclude the completion of screening and involved profile panel training as described in this chapter, the following variations could be considered.

- Focus on the most relevant product categories to shorten the training period.
- Train a smaller panel/fewer assessors.
- Use consensus exclusively.
- Conduct fewer sessions to discuss profiles and product differences/nuances.
- Use fewer qualitative and quantitative references.

7.7.2.2 Without Internal Expertise in Original Profile Methodology/No Qualified Profile Trainer

Often, there might not be a sensory staff member with the knowledge and required skills to be the profile trainer. While there might be a panel leader, he/she might not have experience in original profile methodology to complete the training.

In this case, the company should seek external support in order for this expertise and knowledge to be developed internally. This external support/consultant can work with the panel leader/sensory group to initiate the profile programme.

If acquiring external expertise is not possible, the group should consider the implementation of another descriptive method based on the sensory staff's expertise. Alternatively, a descriptive method requiring neither a training programme nor a profile trainer should be considered.

7.7.2.3 With Limitations of Reference Materials

A profile training programme requires the use of extensive references (both qualitative and quantitative). Reference materials often have to be prepared by internal resources since they might involve product ingredients and chemical references. With time limitations or lack of internal support, some of the necessary references may not be available. In this case, modifications might include:

- the use of more simple training references that can be purchased and prepared by the sensory staff
- the development and use of simpler terms that can be demonstrated by more basic references.

7.8 Statistical Analysis

7.8.1 Original Flavor and Texture Profile Methods

Consensus data are collected in the original profile methods, as explained in previous sections. Thus, no individual data are collected and results are not statistically analysed.

7.8.2 Modified/Derivative Profile Methods

In modified profile methods, assessors complete individual evaluations and provide attribute ratings. These results are statistically analysed.

The statistical analysis used to analyse modified profile data depends on the test design and completion of the study. In general, all modified/derivative profile methods use ANOVA to study sample/product effects.

The reader is referred to relevant chapters in this book to learn how specific modified profile methods (QFP, A^{5}daptive Profile Method, Spectrum Method) analyse their data.

Chapter 5 presents a discussion of the statistical analyses recommended for descriptive data, as well as the ANOVA analyses and philosophies for this type of data.

7.9 Case Study: Use of Profile Descriptive Analysis in Optimization Research Guidance

As a result of new technologies and ingredients, the cheese blend of a company's leading cheese cracker had to be reformulated. The cheese cracker had maintained leadership in the marketplace for many years, so the manufacturer planned a series of research initiatives to ensure the development of an equally acceptable optimized cheese cracker.

The company's sensory and consumer insights group worked with the product development group in this effort. Three phases/steps were planned in this research study.

- *Phase 1: Optimization of Cheese Blends*
  - The completion of an optimization study to investigate the effect of several sources and levels of key cheese blend ingredients on the product's flavour.
  - Based on the optimization results, the identification and selection of 1–3 optimized cheese blends for Phase 2, which should meet the target criteria established by product development.
- *Phase 2: Texture Improvement Using Optimized Cheese Blends*
  - The study of the effect of a new wheat on the cracker's texture characteristics. One to three optimized cheese blends identified in Phase 1 were to be used in this step.
  - A comparison of the optimized prototypes with the current product/control to determine if improved flavour and texture were achieved.
- *Phase 3: Consumer Research Guidance Study*. The design and execution of a consumer research guidance study to confirm that at least one of the optimized prototypes achieves the same or higher consumer acceptance as the current/control product. (The details of this step are not shown since they fall beyond the scope of this case study.)

7.9.1 Phase 1: Optimization of Cheese Blends

7.9.1.1 Sensory and Consumer Insights Group's Resources and Data Comparison

7.9.1.1.1 Descriptive Panels and Resources

Profile descriptive analysis has been the descriptive methodology used by the company's sensory and consumer insights group.

An original flavor profile panel was successfully used in the company for many years. When the research centre was moved to another location, the company decided to train a modified profile panel using the A^{5}daptive Profile Method.

For all the company's key products, complete product profiles generated by the original flavor profile panel were available (historical data).

7.9.1.1.2 Comparison of Historical (Original Flavor Profile) and Current (A^{5}daptive Profile Method) Panel Results

Prior to initiating any reformulation efforts, management's goal was to compare the original flavor profile data (historical data) with the current A^{5}daptive Profile results. The objective was to ensure that the current profile results of the leading cheese cracker accurately reflect the sensory properties of the product produced for many years and documented through original flavor profile results.

Descriptive Studies (Descriptive Evaluation 1)

The original flavor profile and the A^{5}daptive Profile studies were conducted at different times and with different panels.

The methodology followed to establish the terminology/lexicon for both panels is described in section 7.3.3.2 and summarized in Table 7.5. In addition, the project activities were completed as described in section 7.3.4. Both panels had been trained and had ample experience with the terminology (flavor profile panel) and the ballot (A^5daptive Profile panel) developed for cheese crackers.

Results

Tables 7.6a and 7.6b show the original flavor profile (historical data) and the A^5daptive Profile flavour results of the current cheese cracker. The methodology and results/conclusions were compared and showed the following.

Table 7.6a Case study. Original flavor profile results of the current cheese cracker.

Flavor Profile – Current Cheese Cracker*
Code: XYZ 321
Date: Oct 10 2005

Sensory Characteristic	Intensity (consensus)
AROMA	
Amplitude 1½	
Cheesy/buttery (isovaleric + butyric)	2
Toasted wheat (slightly cooked)	1
Caramelized sweet	½
Heated oil (slightly oxidized)	½
Black pepper, spices	½
Onion/Garlic	½
FLAVOUR	
Amplitude 1½	
Cheesy (butyric + slight isovaleric)	2
Salty	2
Sweet	1
Heated oil (slightly oxidized)	½
Sour	½
Wheat (mainly toasted)	1½
Caramelized sweet	½
Oily mouthfeel	½
Spice burn	1
Black pepper/paprika	1
HVP	1
Others: buttery, onion/garlic, bitter, NFDM/casein	
AFTERTASTE	
Cheesy, salty, oily + oily mouthfeel	

* Scale – 7-point scale (0–3).
HVP, hydrolysed vegetable protein; NFDM, non-fat dry milk.

Table 7.6b Case study. The A^{5}daptive Profile Method flavour results of the current cheese cracker.

Sensory Characteristic	Average intensity*
AROMATICS	
Blendedness aromatics	6.2
Total cheese/dairy	8.3
Isovaleric	2.2
Diacetyl	1.5
Butyric acid	3.1
NFDM/casein	1.2
Nutty	1.1
Total wheat	6.7
Cooked	3.1
Toasted	4.9
Caramelized/sweet aromatic	2.8
Black pepper	4.1
Paprika/woody	1.8
Onion/garlic	1.6
Heated oil	2.0
Cardboard	1.7
Painty	0.8
HVP	2.9
BASIC TASTES AND TRIGEMINAL SENSATIONS	
Saltiness	7.7
Sweetness	2.8
Sourness	2.2
Bitterness	1.1
Heat/burn	3.4
AFTERTASTES (2 min)	
Total aftertaste aromatics	3.7
Sweetness	1.3
Saltiness	1.6

* Scale – 10-point scale, adaptable to a 15-point scale for higher intensities.
HVP, hydrolysed vegetable protein; NFDM, non-fat dry milk.

Comparison of Methodology. The key differences in methodology between the two descriptive approaches used were as follows.

- *Amplitude versus blendedness*. Amplitude was measured in the original flavor profile method, while in the A^{5}daptive Profile Method, only a component of amplitude (blendedness) was measured.
- *Order of appearance*. The flavor profile method provides information on the attributes and intensities in the order they are perceived. In the A^{5}daptive Profile Method, the panel develops a ballot that lists the attributes in a set order.
- *Aroma*. Aroma is always part of a flavor profile evaluation. The A^{5}daptive Profile panel did not consider it necessary to measure aroma attributes in this project.

- *Scale*. The original flavor profile uses a seven-point scale (measured in half points from 0–3), while the A^{5}daptive Profile Method uses a 10-point scale (which is expanded/adapted to a 15- point scale for attributes with high intensities).

Comparison of the Products' Characterization. The original flavor profile and the A^{5}daptive Profile panels yielded similar product findings.

- The current cheese cracker has a medium amplitude (original flavor profile) or blendedness (A^{5}daptive Profile Method), indicating that the flavours are not fully blended and that the product has some off-notes.
- Overall, the main flavour signature/profile of the current cheese cracker documented by the flavor profile method and the A^{5}daptive Profile Method is similar – specifically:
  - relatively high cheese intensity (mainly butyric, with low intensities of isovaleric, diacetyl and non-fat dry milk (NFDM)/casein notes)
  - the total cheese intensity is higher than the total wheat intensity (cracker component)
  - saltiness and total cheese are the attributes perceived at the highest intensity. With slightly lower intensity, wheat (mainly toasted) is the other flavour note perceived at medium intensity
  - characteristics perceived at low-medium intensities are black pepper (black pepper/paprika), heat/burn, sweetness, and hydrolysed vegetable protein (HVP).

In addition, this evaluation showed the unique characteristics of each method.

- The flavor profile panel captured aroma information and order of appearance of the flavour notes.
- The A^{5}daptive Profile Method offered more information/details on the 'total cheese' character (total cheese assessment), and more quantitative/intensity discrimination and details because of the scale used.

This study demonstrated that the A^{5}daptive Profile panel showed a similar flavour characterization to the one provided by the original flavor profile panel. Therefore, management felt confident in using the A^{5}daptive Profile descriptive panel in the optimization research guidance efforts.

7.9.1.2 Optimization Study (Descriptive Evaluation 2)

Different sources and levels of the cheese blend's key ingredients were investigated in a fractional factorial design. The A^{5}daptive Profile panel was used to evaluate all prototypes.

The objective of the optimization study was to identify the cheese blends that had the highest intensities in cheese flavour and blendedness, and the lowest intensities in off-notes and HVP.

Only flavour characteristics were evaluated by the panel (descriptive evaluation 2) since in this phase the focus was on flavour. Therefore, the ballot was adapted to complete a more simple/focused evaluation (i.e. a product screening).

These A^{5}daptive Profile flavour results identified two prototypes, B67TX42 (ID in this discussion as prototype B) and F95AR71 (ID in this discussion as prototype F), that met the objectives: the highest cheese and blendedness

Table 7.7 Case study. The A^{5}daptive Profile Method focused evaluation results of two prototypes identified in the optimization of cheese blends.

Sensory Characteristic	Average intensities*	
	Prototype F	Prototype B
Blendedness aromatics	8.4	7.2
Total cheese/dairy	8.7	8.2
Total wheat	7.4	7.2
Caramelized/sweet aromatic	2.3	1.2
Black pepper	3.1	2.3
Paprika/woody	1.0	0.5
Onion/garlic	2.2	1.8
Cardboard	0.0	0.7
HVP	1.6	2.2
Saltiness	9.1	8.9
Sweetness	3.2	2.8
Sourness	2.1	3.0
Bitterness	0.5	0.5

* Scale – 10-point scale, adaptable to a 15-point scale for higher intensities.
HVP, hydrolysed vegetable protein.

intensities and relatively low intensities of off-notes and HVP (Table 7.7). These two blends were chosen for Phase 2.

7.9.2 Phase 2: Texture Improvement Using Optimized Cheese Blends

7.9.2.1 Study Design (Descriptive Evaluation 3)

In this phase, a different wheat was studied to improve the current cracker's texture, especially the crispness and moistness. A test was conducted (descriptive evaluation 3) for the A^{5}daptive Profile panel to complete the flavour and texture characterization of the three products (current product and two prototypes).

The table below shows the characteristics of the prototypes produced for this phase's descriptive study. Both prototypes (F and B) were produced with the improved wheat, and they each had one of the two optimized cheese blends identified in Phase 1.

Variable	Current product	Prototype F	Prototype B
Flavour	Current cheese blend	Improved cheese blend F95AR71	Improved cheese blend B67TX42
Texture	Current wheat	Improved wheat	Improved wheat

7.9.2.2 Results

Tables 7.8a and 7.8b show the A^{5}daptive Profile's flavour and texture characterization of the three products (current and prototypes F and B).

Table 7.8a Case study. The A[5]daptive Profile Method flavour evaluation results of the current product and two prototypes (F and B).

Attribute	Average intensities*		
	Current	Prototype F	Prototype B
AROMATICS			
Blendedness aromatics	6.3 *c*	8.1 *a*	7.1 *b*
Total cheese/dairy	7.9 *b*	8.9 *a*	8.0 *b*
Isovaleric	2.0 *c*	3.2 *b*	4.2 *a*
Diacetyl	1.9 *a*	1.3 *a*	0.5 *b*
Butyric acid	3.4 *a*	3.3 *a*	2.6 *b*
Soured dairy	0	0	1.1
NFDM/casein	1.3 *a*	0.8 *ab*	0.6 *b*
Nutty	1.1 *b*	2.0 *a*	0.2 *c*
Total wheat	6.4 *b*	7.2 *a*	6.8 *ab*
Cooked	2.9 *a*	2.2 *b*	2.8 *a*
Toasted	4.3 *b*	5.9 *a*	5.5 *a*
Caramelized/sweet aromatic	3.2 *a*	2.8 *a*	1.5 *b*
Black pepper	3.7 *a*	2.5 *b*	2.0 *b*
Paprika/woody	2.0	1.4	1.6
Onion/garlic	1.2 *b*	2.9 *a*	2.6 *a*
Heated oil	1.5 *b*	2.2 *a*	1.8 *ab*
Cardboard	1.3	0	0.5
Painty	0.4	0	0
Musty	0	0	0.4
HVP	3.3 *a*	2.2 *b*	2.7 *b*
BASIC TASTES AND TRIGEMINAL SENSATIONS			
Saltiness	7.9 *b*	8.5 *a*	8.7 *a*
Sweetness	2.6 *b*	3.0 *a*	2.5 *b*
Sourness	2.3 *a*	1.6 *b*	2.5 *a*
Bitterness	1.5	1.1	0.8
Heat/burn	2.9 *a*	2.3 *ab*	1.8 *b*
AFTERTASTES (2 min)			
Total aftertaste aromatics	4.1 *a*	3.3 *b*	2.1 *c*
Sweetness	1.2	1.8	0.7
Saltiness	1.8 *b*	3.1 *a*	3.0 *a*

* Scale – 10-point scale adaptable to a 15-point scale for higher intensities.
Means that are followed by the same letter are not significantly different at the 0.05 significance level (Fisher's LSD).
HVP, hydrolysed vegetable protein; NFDM, non-fat dry milk.

7.9.2.2.1 Key Flavour Findings

In flavour (see Table 7.8a), prototype F appeared to be a better candidate compared to prototype B, because it met all the desired properties of an improved/optimized cheese cracker as delineated by product developers.

- Highest blendedness and significantly higher than the current product's blendedness

- Highest cheese intensity and significantly higher than the current product's cheese intensity
- No off-notes
- Low HVP and significantly lower than the current product's HVP intensity

7.9.2.2.2 Other Flavour Findings

In addition, prototype F was considered a better candidate compared to prototype B, because:

- prototype F more closely resembled the cheese profile of the current product. More differences were found between the control/current product and prototype B. Compared to the control, prototype B had significantly lower butyric, diacetyl and nutty, and higher isovaleric intensities. Prototype B also had a slight soured dairy note
- prototype B differed from the control in other flavour characteristics, such as significantly lower sweet aromatic and heat/burn intensities.

7.9.2.2.3 Texture and Wheat Type Effect (see Table 7.8b)

The descriptive results showed significant effects on texture due to wheat type, since both prototypes F and B (produced with the improved wheat) differed from the control in some texture characteristics. Specifically, prototypes F and B had significant higher intensities in characteristics targeted for improvement compared to the control: crispness, persistence of crispness and moistness.

In addition, prototypes F and B showed other significant texture effects with the improved wheat, including lower surface roughness and higher hardness.

7.9.3 Conclusions

The profile descriptive analysis panel provided invaluable research guidance to the product developers throughout the optimization process of the company's cheese cracker.

The A^{5}daptive Profile Method provided the technical flavour characterization of many prototypes produced in the optimization study to aid in the selection of cheese blends that met the target criteria of product development: highest intensities in cheese flavour and blendedness, and lowest intensities in off-notes and HVP. In addition, in the research study of the new wheat, the A^{5}daptive Profile Method clearly identified the texture characteristics affected by this new ingredient.

Based on the flavour and texture descriptive characterization of the current product and the two prototypes, prototype F was preliminarily chosen by the product development group because of its improved flavour and texture characteristics. Phase 3 was completed to confirm that prototype F had the same or higher consumer acceptance than both the current product and prototype B. The details of the consumer study fall beyond the scope of this case study and are not discussed. However, the consumer study confirmed that prototype F was significantly better liked and preferred over prototype B and the current product.

Table 7.8b Case study. The A^{5}daptive Profile Method texture evaluation results of the current product and two prototypes (F and B).

Attribute	Average intensities*		
	Current	Prototype F	Prototype B
SURFACE PROPERTIES			
Roughness	4.5 a	3.1 b	3.5 b
Loose particles	2.2	2.4	2.5
FIRST CHEW (1–2 CHEWS)			
Hardness	5.8 b	6.7 a	6.9 a
Denseness	6.1	6.2	6.5
Crispness	6.1 b	7.9 a	7.8 a
MASTICATION (8–12 CHEWS)			
Persistence of crisp	5.4 b	7.1 a	6.8 a
Moisture absorption	7.4	7.6	7.5
Cohesiveness of mass	6.4	6.4	6.5
Roughness of mass	5.0	5.2	5.4
Moistness of mass	5.2 b	7.1 a	7.3 a
RESIDUAL			
Loose particles	4.7	4.6	4.5
Toothpack/toothstick	5.0	4.9	5.0
Oily mouthcoat	2.1	1.5	1.6
Chalky mouthcoat	1.5	0.9	0.8

* Scale:10-point scale adaptable to a 15-point scale for higher intensities.
Means that are followed by the same letter are not significantly different at the 0.05 significance level (Fisher's LSD).

Based on this information, the company's decision was to incorporate all the product improvements researched in this study, and produce and market prototype F as the company's improved cheese cracker.

7.10 Summary

The original flavor and texture profile methods were the first descriptive analysis methods developed and used in sensory science. Therefore, as described in this chapter, the profile methodology established the technical foundations that have been used by all other traditional descriptive methods.

Furthermore, and considering that 'training of a panel' was not even a practice prior to the development of the flavor profile method (1940s), its conception was truly ground breaking at that time. Thus, it would be appropriate to state that any descriptive method (traditional or rapid) conceptually bases its core nature (having a group of people/panel evaluate products) on the first descriptive methods that develop this concept: the profile methods.

This chapter's main objectives were to summarize for both the original flavor and texture profile methods, (1) their contributions, key characteristics, needs,and steps to train profile panels, and (2) the way many of the philosophical concepts and approaches developed by the original profile methods were adopted and are being used by all profile derivative/modified profile methods.

In this chapter, updated profile descriptive approaches were presented and discussed. The authors described the practices that have been modified and adapted by profile-based traditional descriptive methods. Currently, the profile methods as originally developed are still being used by several organizations. In addition, and more importantly, the key contributions and core practices of the original profile descriptive methods continue to be used, in a modified form, by all profile-derivative methods (e.g. PAA, Spectrum Method, A^{5}daptive Profile Method, etc.).

In summary, the main contribution of the original flavor and texture profile methods is having conceptualized the use of panels for the sensory evaluation of products, compared to the practice of having one expert provide product information. Their other key and extremely valuable contributions include having established the foundations of descriptive analysis to be followed by all methods, establishing and using highly technical and comprehensive terminology to fully characterize the products' sensory attributes, the practice and benefits of using both qualitative references and quantitative/intensity references/anchors, the evaluation of complex attributes (e.g. amplitude, blendedness, etc.) and the use of consensus (with or without individual evaluations and the corresponding statistical analysis) to reach panel agreement on product attributes and product differences/similarities.

7.11 Future Directions

Since their inception, the flavor and texture profile methods have provided direction to the sensory community by laying the necessary foundations for the development of all descriptive methods.

Notably, the profile methodology still has direction and contributions to offer to sensory practitioners. This involves promoting the learning and use of very valuable profile practices. Regrettably, young sensory professionals have not learned or explored these practices, and an effort should be made through this publication and other sources to share the value and main characteristics of these founding descriptive methods. Furthermore, the sensory community has been remiss in failing to acknowledge the original flavor and texture profile methods' contributions and how directly or indirectly all professionals involved in descriptive analysis use the profile method's techniques as developed or adapted. These main contributions and key original core descriptive concepts should be known by all professionals involved in descriptive analysis. Therefore, the current future direction of the flavor and texture profile methods is to guide sensory professionals to:

- learn their basic characteristics, contributions and core valuable practices
- use/integrate these methods' useful practices in current descriptive practices, as explained below.

The sometimes overlooked, very valuable practices developed by the original flavor and texture profile methods that practitioners should learn about and consider using, regardless of the method/philosophy followed, are as follows.

- *The use of consensus*: this approach should not be viewed as a data collection technique when no statistical analysis is available or is not completed. Instead, it should be used for the purpose and benefits originally proposed by the flavor and texture profile methods; specifically to provide a set-up to conduct and foster panel discussions to meticulously evaluate the products' behaviour and sensory properties. Often, as explained in this chapter, product differences are missed when collecting and statistically analysing individual data. Parallel to the collection of individual data and the data analysis, sensory practitioners should conduct, when appropriate, panel discussions as done by the original profile methods to discuss product characteristics or a product issue and reach a consensus.
- *The use/evaluation of complex/integrated attributes*: this chapter explained the value of evaluating, when appropriate, selected complex/integrated attributes proposed by the profile methods in addition to the individual characteristics. Often, products differ in their overall perception properties, or in these integrated attributes, such as blendedness, fullness, etc.
- *The use of intensity/quantitative references*: except in rapid methods or approaches which philosophically are opposed to this practice, the use of intensity/quantitative references as envisioned and originally practised by the flavor and texture profile methods should be considered for their advantages. Practitioners who currently do not use intensity/quantitative references are encouraged to further assess their advantages and disadvantages. If for their particular case the advantages outweigh the drawbacks, sensory practitioners are encouraged to use this very valuable technique developed by the profile methods.

References

Adhikari, K., Chambers, E.IV, Miller, R., Vázquez-Araújo, L., Bhumiratana, N. & Philip, C. (2011) Development of a lexicon for beef flavor in intact muscle. *Journal of Sensory Studies*, 26, 413–420.

ASTM (American Society for Testing and Materials) (2003) *E 1490–03. Standard Practice for Descriptive Analysis of Creams and Lotions*. West Conshohocken: American Society for Testing and Materials.

ASTM (American Society for Testing and Materials) (2006) *E 2082–06. Standard Guide for Descriptive Analysis of Shampoo Performance*. West Conshohocken: American Society for Testing and Materials.

ASTM (American Society for Testing and Materials) (2011) *E 1490–11. Standard Guide. Two Sensory Descriptive Analysis Methods for Skin Creams and Lotions*. West Conshohocken: American Society for Testing and Materials.

Bourne, M.C., Sandoval, A.M.R., Villalobos, M.C. & Buckle, T.S. (1975) Training a Texture Profile panel and development of standard rating scales in Colombia. *Journal of Texture Studies*, 6, 43–52.

Brandt, M.A., Skinner, E.Z. & Coleman, J.A. (1963) The Texture Profile Method. *Journal of Food Science*, 28, 404–409.

Cairncross, S.E. & Sjöström, L.B. (1950) Flavor profiles. A new approach to flavor problems. *Food Technology*, 4, 308–311.

Caul, J.F. (1957) The Profile method of flavor analysis. *Advances in Food Research*, 7(1), 1– 40.

Caul, J.F. & Sjöström, L.B. (1951) Effects of replacing sucrose by other sweeteners in blackberry and raspberry packs. *Quick Frozen Foods*, 13, 59.

Caul, J.F., Cairncross, S.E. & Sjöström, L.B. (1958) The flavor profile in review. In: *Flavor Research and Food Acceptance*. New York: Reinhold.

Civille, G.V. & Liska, I.H. (1975) Modifications and applications to foods of the general foods sensory texture profile technique. *Journal of Texture Studies*, 6, 19–31.

Dairou, V. & Sieffermann, J.M. (2002) A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, 67, 826–834.

Hanson, J.E., Kendall, D.A., Smith, N.F. & Hess, A.P. (1983) The missing link. *Beverage World*, 102–108.

Hongsoongnern, P. & Chambers, E. IV (2008) A lexicon for texture and flavor characteristics of fresh and processed tomatoes. *Journal of Sensory Studies*, 23, 583–599.

Hough, G., Contarini, A. & Muñoz, A.M. (1994) Training a texture profile panel and constructing standard rating scales in Argentina. *Journal of Texture Studies*, 25, 45–57.

Jenkins, A., Chambers, D.H., Chambers, E.IV & Oupadissakoon, C. (2008) Adjusting Texture Scale References for Asian Markets. Presented at Society of Sensory Professionals Conference, November 5–7, Covington, Kentucky.

Keane, P. (1992) The flavor profile. In: *ASTM Manual on Descriptive Analysis Testing*. West Conshohocken: American Society for Testing and Materials.

Krasner, S.W., McGuire, M.J. & Ferguson, V.B. (1985) Taste and odors: the flavor profile method. *Journal of the American Water Works Association*, 35–39.

Lotong, V., Chambers, E.IV & Chambers, D.H. (2000) Determination of the sensory attributes of wheat sourdough bread. *Journal of Sensory Studies*, 15, 309–326.

Marshall, R.J. & Kirby, S.P. (1988) Sensory measurement of food texture by Free Choice Profiling. *Journal of Sensory Studies*, 3, 63–80.

McDowell, I.T. (1963) ADL's panel training program. In: Arthur D. Little, Inc. (ed.) *The Flavor Profile*. Cambridge: Arthur D. Little Publications.

Meilgaard, M., Civille, G.V. & Carr, B.T. (2007) *Sensory Evaluation Techniques*, 4th edn. Boca Raton: CRC Press.

Moskowitz, H.R., Muñoz, A.M. & Gacula, M. (2003) *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*. Connecticut: Food and Nutrition Press.

Muñoz, A.M. (1986) Development and application of texture reference scales. *Journal of Sensory Studies*, 1, 55–83.

Muñoz, A.M. & Bleibaum, R.N. (2001) Fundamental Descriptive Analysis Techniques. Exploration of their Origins, Differences and Controversies. Workshop presented at 2001: A Sense Odyssey', Fourth Pangborn Sensory Science Symposium, Dijon, France.

Muñoz, A.M., Civille, G.V. & Carr, B.T. (1992a) *Sensory Evaluation in Quality Control*. London: Chapman & Hall.

Muñoz, A.M., Szczesniak, A.S., Einstein, M.A. & Schwartz, N.O. (1992b) The texture profile. In: *ASTM Manual on Descriptive Analysis Testing*. West Conshohocken: American Society for Testing and Materials.

Neilson, A.J., Ferguson, V.B. & Kendall, D.A. (1988) Profile methods: flavor profile and profile attribute analysis. In: Moskowitz, H. (ed.) *Applied Sensory Analysis of Foods*. Boca Raton: CRC Press.

Schwartz, N.O. (1975) Adaptation of the sensory texture profile method to skin care products. *Journal of Texture Studies*, 6, 33–42.

Stone, H. (1992) Quantitative descriptive analysis (QDA). In: *ASTM Manual on Descriptive Analysis Testing*. West Conshohocken: American Society for Testing and Materials.

Stone, H. & Sidel, J.L. (1998) Quantitative descriptive analysis: developments, applications and the future. *Food Technology*, 52(8), 48–52.

Stone, H., Sidel, J., Oliver, S., Woolsey, A. & Singleton, R.C. (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology*, 28(11), 24, 26, 29, 32, 34.

Szczesniak, A.S. (1963) Classification of textural characteristics. *Journal of Food Science*, 28(4), 385–389.

Szczesniak, A.S., Brandt, M.A. & Friedman, H.H. (1963) Development of standard rating scales for mechanical parameters of texture and correlation between the objective and sensory methods of texture evaluation. *Journal of Food Science*, 28(4), 397–403.

Talavera-Bianchi, M., Chambers, E.IV & Chambers, D.H. (2009) Lexicon to describe flavor of fresh leafy vegetables. *Journal of Sensory Studies*, 25, 163–183.

Thompson, K.R., Chambers, D.H. & Chambers, E.IV. (2009) Sensory characteristics of ice cream produced in the USA and Italy. *Journal of Sensory Studies*, 24, 396–414.

Thomson, D.M.H. & MacFie, H.J.H. (1983) Is there an alternative to descriptive sensory assessment? In: Williams, A.A. & Atkin, R.K. (eds) *Sensory Quality in Foods and Beverages: Its Definition, Measurement and Control*. Chichester: Horwood.

CHAPTER 8
Quantitative Descriptive Analysis

Joel L. Sidel, Rebecca N. Bleibaum and K.W. Clara Tao

8.1 Introduction

Can the output from a trained descriptive panel be statistically analysed? That question was directed to a speaker in the question and answer session following a presentation at the Institute of Food Technology (IFT) about results from flavor profile* panelling (late 1960s). The answer offered was 'no', that this kind of panel and result was not quantifiable. I recall my feeling of disbelief at that response. Psychologists had been quantifying perception since the early beginnings of psychophysics. Subjects, stimuli and an unbiased measurement and analysis system were all that should be necessary to quantify sensory descriptive perceptions.

The development of quantitative descriptive analysis (QDA) in the early 1970s (Stone et al. 1974) clearly demonstrated that 'output from a trained descriptive panel is quantifiable and statistically analysable'. The following chapter describes the early development of QDA to the more recent introduction of Tragon QDA®.

8.2 Background

As its name implies, the QDA method relies on measurement and quantification. The approach reflects the science backgrounds of its developers, Herbert Stone and Joel L. Sidel, their commitment to the scientific method and the renowned research environment where it was developed – Stanford Research Institute, now known as SRI International.

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

QDA has been described in several publications and books and, of course, in the writings of Stone, Sidel and their associates (Stone & Bleibaum 2009; Stone et al. 2012; Stone et al. 1974). From QDA's inception, its developers compiled the best available sensory practices and knowledge of human sensory measurement, and expected the method to evolve as practices and knowledge base grew. The evolution of the original QDA method continues today as Tragon QDA.

This chapter describes the Tragon QDA method and provides the rationale behind each of its components. As the physician's rule is to 'first, do no harm', the primary rule for the QDA developers and those contributing to its evolution is to eliminate or minimize bias in the descriptive process. Bias contaminates tests, rendering them useless or, worse, misleading. Bias can occur as a result of the subjects participating in the research, based on how they are recruited, screened and trained. Bias is also possible in the test design, procedures and how subjects' responses are examined, analysed and reported. Descriptive analysis management is not for the faint-hearted!

The objectives of the QDA method include:

- providing a scientifically rigorous procedure with quantifiable results
- producing output directly generalizable to consumers
- employing a panel leader who is a discussion facilitator, not a teacher
- involving all sensory modalities and their interaction
- being cost-effective
- screening and training subjects in less than 2 weeks
- providing multiple applications in product development, quality control and marketing.

The current chapter discusses these objectives, why they are important to descriptive analysis and how they are specifically addressed by QDA.

8.2.1 Method Development

8.2.1.1 Pre-1971

With his colleagues at the Stanford Research Institute (now SRI International) in Menlo Park, CA, Dr Herbert Stone developed a rudimentary series of procedural steps and data analysis for descriptive testing. The procedure included the use of a semi-structured line scale for scoring attribute intensities. The line scale, based on research conducted by Norman Anderson at the University of California in San Diego (Anderson 1970), provides a basis for quantification and statistical analysis using powerful parametric techniques such as the analysis of variance (AOV). Stone and his team developed a quantitative descriptive procedure rather than relying on qualitative descriptive panels popular at that time, for example the flavor profile (Caul 1957).

The SRI approach is not viewed so much as a method but as a series of progressive steps founded on sound scientific principles for data collection and analysis. One of the earliest descriptive studies conducted at SRI was for bottled water, involving 12 subjects, 12 replications and 12 products. Among the primary components of this new approach to descriptive testing were panel number, line scales, replication, evaluation of large numbers of products, a democratic procedure for language development, the non-autocratic role of the panel leader, individual testing booths, individual data (rather than consensus) and use of discrimination tests screening for subject sensory acuity.

8.2.1.2 1971–1974

In 1971, Joel Sidel joined Stone's food department team to lead the sensory group. Sidel, a psychologist, previously headed a sensory group for the US Army Natick Research Laboratories and for Hunt-Wesson Foods. He had little confidence in qualitative descriptive procedures, considering them fraught with bias that rendered them misleading to useless. The elements of Stone's work toward a quantitative descriptive procedure were consistent with Sidel's interest in developing scientifically defensible sensory tests. From 1971 to 1974, the team made several refinements and additions that enhanced the original procedure.

During a capabilities presentation to an SRI client in New York, Sidel was introduced to describe a 'new testing method'. Asked for the name of the method he hesitated a moment, then said 'quantitative descriptive analysis'. The audience wanted a name for the technique, so the procedure now had a name. The eventual intention to create a name with more pizzazz never materialized; interest in the method itself left little time for the name game.

Clients immediately accepted that subject screening and training do not require months and years, as do behaviour modification procedures, and can be accomplished in a couple of weeks. Further refinements were added to the subject selection procedure, including a survey about a variety of common and unique products and flavours. Subjects unable to score products successfully on a scale or those demonstrating unusual scoring patterns are eliminated from further participation. Subjects liking a product or flavour are candidates for further screening. The latter must also be a user, or potential user (products not yet available in the marketplace), of the product or product category.

Language development procedures were also further refined and the panel leader role was influenced by principles learned from psychology (e.g. learning, cognition, group dynamics, social psychology, educational psychology, etc.). The approach of psychologist Carl Rogers (1959), which focused on personality and interpersonal relationships, was very influential. His work demonstrated the value of a group solving problems actively, rather than being passively taught answers. At SRI, this translated to requiring subjects to develop the score sheet, definitions, etc. without the panel leader imposing his/her will.

Additional statistical tests were added to the analyses for rank order testing, including Kendall's *W* coefficient of concordance (Siegel & Castellan 1988) and Duncan's range test (1955). A unique graphing technique, affectionately referred to as a 'spider web', was developed for QDA and first published in 1974 (Mecredy et al. 1974; Stone et al. 1974). QDA graphs are easy to read and useful for comparing products simultaneously across many sensory attributes. Spider graphs are now used for a broad range of product categories.

The QDA procedure extended to include home evaluation of products such as paper goods, diapers and various personal use items. Screened subjects take the product home, use it, make written observations and return the next day for language development. The sequence proceeds until the completion of training, usually in 1–2 weeks.

SRI's food science department began offering workshops in sensory evaluation. Stone and Sidel presented most of the material, generously assisted by guest speakers Rose Marie Pangborn and Dr Howard G. Schutz. Consumer product companies sent employees representing marketing, quality control, product development and sensory evaluation to the SRI workshops. Workshop presentations describing the new quantitative descriptive method, QDA, received a strong positive response. Attendees representing the Pillsbury Company asked if SRI would be willing to teach them the QDA method and develop a panel specifically for them. A method first used in our sensory laboratory now extended outside SRI, and other companies soon counted themselves among those trained by us in the method.

8.2.1.3 1974–2008

Sidel left SRI in July 1974, and for a short time provided consulting services to the organization. Stone also left SRI shortly after Sidel and together they formed Tragon Corporation, offering consulting services in the field of food science and sensory evaluation. The company soon offered workshops and a testing service to augment its consulting activities. Tragon grew in reputation and size. Workshops were offered in the US and abroad, as was QDA training and general sensory consulting.

The value of QDA panels and sensory evaluation convinced many companies to develop and expand their internal sensory capabilities. As demand for sensory scientists grew, universities expanded their testing and course offerings. Tragon's QDA training programme was continuously refined, broadened to include more products and used by an ever increasing number of domestic and international companies.

QDA moved into a more central role when paired with consumer acceptance data for developing predictive models for optimizing consumer acceptability of products (Sidel & Stone 1983; Sidel et al. 1975; Stone & Sidel 1981). QDA is quantitative, relies on a rigorous scientific procedure and uses attribute language and definitions that are consumer driven. Marriage of the two sets of data, one from the trained panel and the other from untrained consumers, is a natural evolution for QDA. Dr Howard G. Schutz was instrumental in developing the optimization modelling techniques that used various multiple regression and multivariate analyses (Schutz 1983).

Dr Brian Yandell, Professor at the Department of Statistics from the University of Wisconsin-Madison (Yandell 1997), updated the QDA statistical software programme. The update used the S-Plus platform and incorporated many of the decision norms developed when analysing QDA data over the years. It was designed to run on PCs, provided many tables and charts for examining subject performance and incorporated QDA graphing tools into the output. Jean Bloomquist, Heather Thomas and Joel Sidel of Tragon worked closely with Dr Yandell over a 3-year period on the software advancements.

8.2.1.4 2008–Present

Although Sidel semi-retired from Tragon in 2007 and Stone in 2008, they still regularly consult on sensory and QDA staff education programmes. Tragon came under new management in 2007; when pursuing a trademark for the term 'QDA', it was discovered that QDA was in such widespread use in the sensory research community that a trademark was rejected. It was therefore necessary to use the term 'Tragon QDA®' to protect the method's integrity and advancements.

In 2009, a major revision and update of the Tragon QDA software programme was undertaken. The objective was to build on the advancements from the 1997 version but to incorporate state-of-the-art tools to provide decision makers with readily available information. The software focuses on key decisions of significance, interactions, correlations and principal component analysis (PCA) as well as resulting maps, panel performance, attribute performance and product similarities and differences. Results are easily used in conjunction with other data for advanced multivariate analysis techniques. The significant software rewrite was published in 2011. It offers many new features and enhancements that make the Tragon QDA data more user friendly, including report-ready plots and tables.

8.2.2 Central Thesis

The central thesis of Tragon QDA is that human sensory perception can be – and should be – measured, quantified and statistically analysed. With appropriate controls, the measurements can be reliable, valid and generalized to larger populations and consumers. The science of sensory measurement requires adhering to basic scientific principles for hypothesis formation and hypothesis testing, without introducing avoidable sources of bias into the research.

The quantitative measurement process in QDA is founded in psychology. For example, 'successive approximation to a goal' is adopted from skills training and learning experiments for use in developing and training panels. The panel leader's role is traceable to the teachings of Carl Rogers. Social psychology and group dynamics teach about the risk of the 'emerging leader'. Physiological psychology and sensory processes provide a knowledge base for how the sensory system is structured and how it functions – from receptor through the peripheral nervous system and the central nervous system. The structure and function of those systems can be quite different between individuals, highlighting a genetic source for individual differences in perception and supporting the need to quantify and statistically analyse descriptive data.

Are observed scoring differences to be expected by chance or do they demonstrate a statistically significant difference? QDA uses unstructured or semi-structured line scales for measuring and scaling perceived differences and intensities. These equal-interval scales are described in psychophysics literature (Anderson 1970). The best information from experimental psychology, experimental design and behavioural statistics is incorporated into Tragon QDA testing. Abundant examples are available; however, by now the reader will recognize that the method takes an eclectic approach integrating the best principles from psychology, while avoiding the potential errors and biases that behavioural research cautions us to avoid.

8.3 Method Description

Tragon QDA method involves measuring perceptions derived from products as the source of the stimulation. The method embraces the uniqueness of our measuring instrument (humans) based on physiological, psychological and behavioural differences, along with experiential differences with a specific product or related products. The responses reflect what is perceived in its entirety – how that product looks, smells, feels and tastes or, for non-food items, the product before, during and after use.

This matter of 'perceived in its entirety' is an aspect of the sensory process that is often disregarded or not fully appreciated by all descriptive analysis methods. When planning a test and considering the evaluation process, there is a tendency to think in narrow terms of 'flavour or texture'; however, this is not the entire picture of what is perceived. One must be careful about the extent to which a modality, such as appearance or before usage, is excluded from a questionnaire because the experimenter does not consider it an important property of the product or relevant to the objective. However, the inter-relationships among modalities cannot be ignored. These sensory inter-relationships must be considered, regardless of the type of test that is planned. If not accounted for, useful information will be lost and the overall value of the sensory resource will be compromised.

8.3.1 Subjects

A high priority in Tragon QDA is the choice of subjects. All individuals interested in participating must be voluntary and qualified to participate, based primarily on their demonstrated sensory skills in the product category of interest. The following guidelines are a reminder that subjects must be qualified and provide directions as to how best to proceed with the qualifying process. However, they should not be read as a rigid formula, because each product category and each test type will have some unique requirements. Therefore, we believe that a basic set of guidelines is far more useful for the sensory professional when developing a pool of qualified subjects.

Subject guidelines to consider include the following.

- Participation should be voluntary and can be stopped at any time.
- All subjects' personal information should be kept confidential.
- Any allergic or medical conditions that may impact testing should be determined.
- Subjects should have no technical knowledge of the products being tested.
- Subjects should be likers and average or above-average users of product categories.
- Subjects should have demonstrated sensory skill within the product categories of interest.

This list is unique in that it does not identify 'standardized' screening products, it does not assign 'standardized' numerical sensory values to products and it does not identify a 'standardized' test method for sensory acuity screening. Listing a specific commercially available product along with a 'standardized' sensory value can be misleading for a variety of reasons, beyond genetic, perceptual and behavioural differences. For example, companies reformulate products on a regular basis; each country may have a unique formulation; storage conditions vary greatly; normal production variability creates a moving target; the product may not be available in a specific market; or it may be culturally inappropriate.

Stating a specific method is equally risky because it, too, may not be relevant to the nature of the problem. Qualifying subjects, as noted above, should be based on sensory skill with the products that will be tested, that is, the product(s) used should be the actual product or products from the category being tested. For example, if a company is in the energy bar business, then using energy bars is appropriate but using cream cheese, hard candy or grape juice is not. The second requirement is that the subjects must demonstrate they can discriminate differences among those products at better than chance. The decision about qualifying will be made by the experimenter based on his/her knowledge about the individual's past performance and the objective(s) for the test.

8.3.2 Subject Screening and Qualification Procedures

The primary purposes of screening are to identify and eliminate those individuals who cannot follow instructions; are insensitive to differences among products being evaluated; exhibit little interest in the activity; or lack verbal fluency. Perhaps most important, the screening should identify those individuals who can detect differences at better than chance among the products of interest. These are the ones identified as qualified based on sensory acuity. This can easily be achieved in about 5–6 hours (over several days and sessions) starting with individuals with no prior experience ('naive subjects').

Potential subjects are pre-recruited following a specific script (usually via telephone or online) that begins with appropriate category usage qualification. The interview includes assessment of availability, verbal fluency, interest and comfort with participating in a group activity. Once qualified in the preinterview, about 30 consumers report to a central testing facility and are given a series of up to 20 discrimination tests over the course of several days. These tests are designed to cover the range of products in the category(s) of interest and cover each modality that will be evaluated by the panel including differences before, during and after usage. The discrimination tests range from easy to moderate to difficult and are based on the product set(s) of interest. Subjects scoring significantly above and beyond chance in the series of screening tests (for example, ≥70% correct duo-trio testing) qualify as having satisfactory sensory acuity.

To establish an array of discrimination tests, the sensory professional must bench-screen the category of products to be tested, identify observable differences and prepare a series of product pairs that include differences before, during and after usage. The number of pairs should be up to 20; with replication, this provides 30–40 trials/judgements, sufficient for subjects to demonstrate their abilities and for the sensory professional to be able to classify individuals based on their sensitivity and reliability. The effectiveness of this approach is based on the subsequent test performance of those individuals. The degree of difficulty of the product pairs should be considered. If more than a few product pairs are easy to differentiate, the screening process will not be effective, as nearly everyone will qualify. Only after this procedure has been used once or twice will the sensory professional know whether pairings represent the desired easy, moderate and difficult options.

After screening, about 12 subjects will be identified as potential candidates for language development based on their sensory acuity and their availability to complete the project.

8.3.3 Language Development

The Tragon QDA language development focuses on using a common, everyday consumer-based language to describe the products along with developing an evaluation protocol that is as close as possible to what the consumer may experience, given the constraints of the research. A trained moderator is the group discussion facilitator. The moderator provides the schedule of activities and works with the panel members to help them develop the common vocabulary that will describe the products of interest along with an evaluation protocol. For most product categories, about 40–50 attributes are usually sufficient to cover all modalities. Generally, 8–12 hours of group discussions over about five sessions (following an iterative process where each session builds on previous sessions) is sufficient to develop a comprehensive language that covers before, during and after usage. Some activities in the language development process may be done at home or through an extended-usage situation with the products of interest.

To begin the language development process, an individual and group orientation is provided for the subjects. During the orientation, the panel leader facilitates introductions and introduces the general concepts of language development, describing their sensations and perceptions of the product category. This introduction and orientation should take no longer than 20 minutes.

Immediately following the orientation, the language development process begins as a group activity. Panel members are provided with an appropriate amount of product for evaluation – more than they will actually use. The first product given to each subject should be the one most 'typical' for the category (for example, a 'gold standard,' control or one that defines the category), along with a category sheet. Each panel member is asked to divide his/her perceptions into categories/modalities such as 'before usage' (visual, aroma, etc.), 'during usage' (application, flavour, mouthfeel, handfeel, etc.) and 'after usage'. Once each panel member has written down his/her individual perceptions, the panel leader will call on each assessor to describe what they have written, tracking each response on the board. This process is repeated for three or four products that best represent the range of products in the research, when typically 90% of the words needed to describe the product category will have been generated.

Products are thoughtfully selected by the panel leader to ensure the range of differences within the product array of interest has been provided to the panel. Each panel session is about 90 minutes in length, and it may require more than one session to describe three or four products because of the physical nature of the category. The language development process is iterative; that is, the words generated in the first few sessions to cover the product category are reviewed, discussed and defined. The consumers then practise scoring products using an unstructured graphic rating scale. They develop a comprehensive list of words to describe the product array and the specific procedures for their evaluation that is most typical for the category of interest. In addition, the subjects decide upon appropriate anchor words for each scale (such as 'weak' to 'strong' or 'slightly' to 'very').

The panel leader creates definitions for each attribute scored, based on input from the panel so that the final definitions represent a true group consensus. The definitions are always present in the data collection sessions so that assessors can be reminded of the meaning of each attribute. This panel method is designed to provide direct consumer feedback to the technical developers and marketing teams on how these products are similar and different based on their sensory properties, absent of brand and imagery. Measures from this consumer-based descriptive analysis method correlate well with consumer affective measures.

8.3.4 Scale Usage

During the language development process, the subjects practise their scoring and discuss their perception on the scale. Scale context is established through exposing subjects to the actual range represented by the stimuli (i.e. products) in the category or project. This approach is observed in Anderson's (1970) research with graphic line scales. Subjects establish their own scale location for samples included in the range of, and practise with, the training stimuli. It is not expected, or necessary, that all subjects perceive or score the low-intensity product at the same exact scale location. Individual subject differences due to scale location are accounted for by, and extracted with, the AOV model used to analyse the data.

In training sessions where subjects scores are distributed over an unusually large scale range, those scores are shown (e.g. blackboard) and the panel discusses whether that range is consistent with their perception. The subjects decide for themselves whether their score represents what they perceive. As indicated in the previous paragraph, differences in scale location are accounted for. All that is required is that subjects remain reasonably consistent (i.e. normal expected variation) in applying their internalized frame of reference and scale. The initial reaction from most individuals is to look for 'the correct location' on the scale or to see if the person adjacent to them is making a mark in the same place.

A highly variable scale more typically reflects product variation than panel inconsistency.

8.3.5 Pilot Test and Validation

Once the sensory language has been developed for the product category of interest, a pilot test is initiated. This test is designed to evaluate individual subject performance and determine if the panel, as a whole, is using the sensory attributes in the same or similar ways. Before the pilot test, the panel leader reviews the procedures, definitions and rating scales for each attribute. Generally, four products may be selected from the category that represents a relatively broad sensory range. Next, these products are evaluated by each subject, on an unbranded basis, in the individual testing booths and, if necessary, in an extended usage situation.

For the pilot test, four replications are recommended. In this case only, the products are evaluated in the same order for each subject. The reasoning behind this is that the pilot test data are analysed to determine individual panel performance and attribute agreement, not product differences. The data should be analysed thoroughly with one- and two-way AOV for each sensory attribute to determine whether the panel, as a whole, scored products differently from one another. Multiple comparison procedures (such as Duncan's new multiple range test) are calculated after the AOV to identify statistically significant differences among products for each sensory attribute. Calculations include product means, standard deviations and ranking of each product by attribute. Data are best presented in spider plots of product means created for each modality (and before, during and after usage). More will be discussed on the specific statistical analysis programme in a later section.

The pilot test data are only for the sensory department to make decisions on areas for remedial panel training. Based on results of the pilot test, the panel is reconvened to discuss attributes and products in which the panel leader seeks clarification of the definitions or evaluation methods or both.

After initial training and pilot testing, remedial training sessions may be scheduled if necessary, or the panel may begin their data collection sessions.

8.3.6 A Note About the Use of References

Tragon QDA uses qualitative references only when needed by the panellists to further clarify an attribute. The references may consist of complex finished products, less complex products or individual ingredients such as cocoa powder, pure vanilla extract, baby oil, petroleum gel, etc. When necessary, references are introduced to aid panel discussions, provide a common sensory experience and promote a common language among the panel members. Panellists are not limited to predescribed reference standards when describing perceived sensory characteristics. In other descriptive methods, quantitative reference standards are used to educate and calibrate panel members. In fact, panellists can be taught to agree with the panel leader on the assigned value. That subjects can be taught to associate a specific stimulus with a specific response demonstrates nothing more than a learned activity. The panel leader provides the reference and the answer for what to call the stimulus (e.g. a 5 sweet). The Tragon QDA panel leader is a facilitator requiring the panel members to arrive at their own conclusions and agreements about whether a reference clarifies a specific sensory experience. The subject is actively engaged in the solution, rather than being a passive receiver.

The variable nature of standards and the efforts of the subjects to adjust their perceptions to the assigned numerical values make it clear why the required training time is so long when quantitative references are used. We view quantitative reference standards and calibration as extensive behaviour modification. It may be an effective way to demonstrate repeatability of a behaviour; however, it tells us nothing about what is perceived or the validity of the activity for evaluating products. With the Tragon QDA method, the goal is to understand, not teach, perception. Therefore, the use of qualitative references is recommended to provide clarity on an attribute when necessary. However, the use of quantitative references with assigned intensity values is not recommended.

8.4 Practical Considerations

The product developer often needs to make changes quickly to accommodate project objectives. In general, changes can be readily incorporated into the Tragon QDA process. Several practical considerations and helpful suggestions are discussed below. Managing sensory attributes involves clarifying any confused words and reducing any redundancy of words used in describing the product. Most projects evolve and descriptive analysis may be useful at various stages of development. To best accommodate the research objectives, amendments to the original test design or additional testing may be required. A flexible research system is important for a successful product launch and the sensory information may be used to help identify a strategic sensory advantage.

8.4.1 Situational Adaptations

The Tragon QDA methodology is readily adapted to any consumer product as the overarching objective is to describe and measure the product's sensory experience. Each product comes with its own set of challenges, and the sensory professional must decide on testing protocols and methodology based on the research objective. The method has been adapted for a wide variety of international products including, but not limited to, apparel, footwear, sporting goods, office supplies, writing instruments, personal care, household care, art supplies, automotive and pet foods, as well as a vast array of foods and beverages.

To adapt the Tragon QDA methodology, the sensory professional must understand how product preparation and testing protocols can be designed to best reflect the end use of that product. Research design choices may impact selection of test location and tasks associated with language development and data collection. For some categories, it may be more appropriate to conduct an in-home evaluation or combine a laboratory procedure and an away-from-the-lab or in-home procedure. Generally, the overall objective is to work with the constraints of the product category and develop a consumer-based language that best reflects the product's sensory experience before, during and after usage. Many consumer usage situations, especially for non-food products, cannot fully be replicated in a laboratory environment. The sensory professional must determine how best to get as close as possible to typical behaviour associated with consumer usage. In this way, the Tragon QDA results can be readily correlated with consumer affective behaviour to determine how the products' sensory experience impacts consumer perception.

For example, with running shoes, subjects could put on the shoes in the laboratory or a central location, then walk around a room or building and describe their initial product sensory impressions. However, the running experience could not be fully understood without a variety of surfaces, distances and running times, along with considerations for indoor/outdoor and various weather conditions. The sensory professional must make key decisions during the recruitment of subjects/runners, the distance, frequency, duration and training level necessary to meet the project objectives. These types of sensory procedural decisions must be discussed prior to initiating the panel to ensure that they meet the business objectives and that the results reflect the intended usability of the data. This is just one example of how the Tragon QDA methodology may be adapted for any product of interest.

With widespread usage of direct data entry and web-based systems, both laboratory and out-of-laboratory techniques will increase in usage and application. Tragon QDA is a flexible methodology and is easily adapted to a wide variety of products for which the laboratory may not be the best location for data collection.

8.4.2 Managing Sensory Attributes

Subjects often ask to include one or more 'overall intensity' terms on a scoresheet. We can speculate about the reasons why they request these words; maybe they find comfort in them. Beyond speculation, we observe some recurring patterns with data from 'overall' terms. For example, several subjects demonstrate a high correlation between a sensory attribute they scored high and their 'overall' score. Where individuals score different attributes as most intense, the panel 'overall' score becomes a multifaceted conglomeration of scores with no single apparent source. In other situations, the overall score simply correlates with the most intense attribute, providing little new information. Our position is to include such terms when the panel requests them. The statistical analysis reveals how those terms are used and whether any new information is provided.

Subjects will commonly use different sensory attributes to describe the same sensations and the same attributes to describe different sensations. These differences are based on their product usage, motivation and genetics. Some subjects may use the term 'sour' for yoghurt while others may describe it as 'tart'. During the panel discussion, subjects will agree on the term to describe the sourness in yoghurt and include the other descriptive words that the panel discussed in the definition of the term. Managing sensory attributes is the panel leader's responsibility.

8.4.3 Adding Products to the Test

Last-minute additions of products within a research project are not uncommon. When this occurs, it is important to ensure that the panel has an opportunity to view these products prior to data collection. Some products create unique sensory experiences and it is important that the language is sufficiently broad to encompass the full sensory experience of the products of interest.

It may be necessary to encompass an array of products and objectives that were not part of the original plan. For example, studies planned to test only two products may be augmented to incorporate samples including control, shelf-life products, multiple competitors and different levels of variants.

By including competitive products, the development team will better understand the sensory landscape of the category. An incremental amount of ingredient such as flavour, acidulant or other compounds may be included to determine whether, or how much, the addition translates to perception differences. Using twice as much of an ingredient does not necessarily mean that subjects will perceive double the intensity. Selecting larger ratios or higher levels of ingredients can help expand sensory differences.

Additions of products are not expected to increase the number of language generation and training sessions, and may only add one or two extra days of data collection. However, if the new array of products represents a much wider sensory range than included in the panel training, it may be necessary to reconvene the panel to discuss and experience intensities beyond the scope of their initial training.

8.4.4 Multiple Category Testing

The Tragon QDA panel is to be used for multiple category testing and there are different approaches to managing various product categories. One method is to qualify subjects, *a priori*, for the different categories when first developing the panel. For example, to establish a chewing gum panel, the subjects should be screened and qualified not only for chewing gum but also for hard and chewy candies. This allows the new panel to participate in all three product categories. Qualifying an existing panel for other categories is accomplished by first screening them for their sensory acuity to the category with a brief set of discrimination tests. Five to six duo-trio tests should be sufficient; of course, the screening tests will include replicates. In a case where the category for initial panel training does not provide sufficient activity to occupy the panel, the panel will benefit by maintaining its overall sensory acuity by actively participating in new categories.

8.4.5 Over-Recruiting for Panels

An active Tragon QDA panel typically consists of 12 subjects, and over-recruiting for a panel provides several benefits. We find it helpful to include two additional subjects in the language development process. Many other subjects may have qualified on the screening tests but for reasons of crowd control, training is limited to 12–14 subjects.

During or soon after training, a subject may be absent or otherwise excused from the panel due to personal reasons or an unforeseen change in their health or lifestyle. Surplus subjects provide flexibility to select replacements. Surplus subjects who gain the advantage of generating language and are familiar with the set of definitions can easily become alternative or substitute panel members.

Surplus subjects who have performed well in their discrimination tests but were unable to participate on the descriptive panel can facilitate training for multiple categories. They require only a couple of warm-up sessions (two or three 90-minute sessions) before joining an existing panel. The purpose of warm-up sessions is to introduce subjects to products, attributes, definitions and scorecards already developed by the panel.

8.4.6 Options for Resource Constraints

Given constraints of resources, there are some options for more rapid language development and data collection. Within the Tragon QDA procedure, this process is referred to as diagnostic descriptive analysis (DDA). In general, to minimize risks, it is strongly recommended to conduct a full Tragon QDA approach first and then, if necessary, to reduce the scorecard down by using data reduction techniques such as PCA or by an in-depth understanding of the key sensory attributes that have the greatest impact on consumer behaviour. Within a research environment, attributes can be selected with a knowledgeable and experienced sensory team based on previous results. However, the panellists should always have direct input to ensure that the reduced set of attributes is well defined and reflects differences observed in the products of interest.

The use of a full panel and replication is strongly recommended with DDA. Diagnostic descriptive analysis is a shortcut procedure for unique situations where training a standard QDA panel is not possible, such as when there is no panel trained to evaluate the products and there is a need for a shortened scorecard, testing and reporting timeline. There are several variations of DDA, each depending on the specific products and requirements of the project. One example involves screening previously trained subjects qualified as likers and users (or potential likers and users) for a product new to the marketplace. As few as 3–6 discrimination tests is reasonable to qualify experienced subjects, and 2–3 language sessions to develop a one- or two-page scoresheet with 12–15 attributes. With this scenario, as few as two replications of the test products may suffice.

It is the responsibility of the sensory professional to make research design decisions to ensure the credibility of the resulting data. Historically, it has been challenging for sensory work to gain credibility as a science and any procedures that increase risks in decision making should be made with full awareness of the associated implications. One poorly executed research experiment or a report of sensory results that don't validate in the marketplace can quickly erode confidence in the sensory programme. Therefore, the sensory professional must take great care to ensure valid and reliable results, even with resource constraints.

These types of DDA programmes are often more common with second- and third-tier products where the risks associated with decision making are reduced.

8.4.7 Collaboration with Cross-Functional Team

During language development, it is always good practice to have a routine discussion with the various teams (e.g. consumer insights, product development and/or brand marketing) about the language and its progress. These discussions can provide feedback to the panel leader, such as where to probe subjects about specific attributes and definitions of interest. Updating the product developer or principal teams on the genesis of the language allows them to follow the evolution of the language. During the early stages of language development, some attributes or attribute terms will be eliminated by the subjects and new terms may be added. The definitions are refined further by the panel prior to initiating data collection. Team updates help prevent surprises for the product developer.

Multiple product testing, multiple category considerations and over-recruitment for a descriptive panel maximize the Tragon QDA process and diminish the need for screening and developing a completely new panel. With many projects that have time constraints and quick turnover, the ability to use the same panel for additional categories keeps the study on the required timeline and minimizes the time and cost for a new panel.

8.5 Statistical Analysis of Tragon QDA Data

As the name suggests, quantitative descriptive analysis (QDA and its successor, Tragon QDA) is a procedure for quantifying the descriptive analysis of sensory stimuli. Valid quantification requires that the measurement process adhere to specified rules and procedures for collecting and analysing data. Measuring and scaling sensory intensities with Tragon QDA are described elsewhere in this chapter; topics will include line scales, equal interval data and replication.

The primary statistical analyses for Tragon QDA, as with the original QDA, include analysis of variance (one-way and two-way), multiple range test (Duncan, Tukey, LSD) (Duncan 1955; Keuls 1952), correlation and multivariate analyses. QDA analysis also includes several propriety analyses and rank-order conversions to provide further insights into the responses of the trained descriptive panel. Where results and conclusions from parametric (e.g. AOV) analysis and non-parametric data (Siegel & Castellan 1988) are different, rank data (e.g. the number of subjects scoring one product higher than another) often will drive analyst conclusions about product differences.

Each subject's individual data are analysed to determine whether that subject is satisfactorily contributing to product discrimination. A one-way AOV for each subject is computed for each attribute. This computation takes into account the subject's ratings for replications of each product. The subject's one-way F-test is based on the mean square for the between variable (i.e. products) divided by the mean square for error based on the replication scores for each product. Subjects contributing worse than chance (in this case defined as 0.50–0.99) are judged to exhibit poor discrimination of product differences for that sensory attribute.

A two-way AOV is used to determine whether the overall panel result for products on an attribute is significant. The source of variance for this model includes product, subject, product × subject and error. The main effects and interactions for replications are typically collapsed into the error. Experience has shown that this has little statistical impact on the error term, as Tragon QDA panels seldom exhibit significant error associated with replication.

A proprietary analysis included in the Tragon QDA statistical programme computes a probability for each subject's contribution to the two-way product × subject interaction. We examine data from subjects with significant scores for this interaction to determine whether the significance is one of magnitude scale differences or cross-over scale differences compared to the panel. Magnitude differences are remedied by converting scores to ranks and performing a rank analysis. Cross-over differences cannot be remedied in the same way.

Cross-over may relate to genetic or experiential sensory differences, or may reflect a need for additional training. The experienced analyst examines the data to determine whether a subject's interaction has seriously skewed the data for an attribute. If the data are seriously skewed, that subject's data may be omitted from the analysis for that attribute. This action will reduce the N (number of subjects and observations for the data), which may consequently result in a loss of product significance. The Tragon QDA® statistical programme has a proprietary calculation measuring the degree of individual subject cross-over for attributes when the two-way AOV is significant. That calculation is valuable for the final decision about attribute, subject and product differences.

When the two-way AOV is significant, there is a variety of statistical tests to determine which products differ for the attribute. If the panel only evaluates two products, a significant AOV is all that is required to declare products different. The Tragon QDA is seldom used for only two product evaluations, and additional products are recommended to obtain full value from the test. When there are more than two products, we use one of the available multiple range tests. After meticulous evaluation of hundreds of studies over many years, we found that the Duncan's new multiple range test for product differences agrees most with rational decisions based on the combination of parametric and rank results. The data examination strategy requires first that the two-way AOV is completed and is significant before Duncan results are considered. This strategy helps protect against finding Duncan differences when the AOV is not significant.

The Tragon QDA statistical programme automatically converts scaled scores to ranks and provides rank tables. We examine each subject's converted rank order data for products and attributes, and for the number of subjects ranking one product higher in intensity than another. We use this information to see how well the rank data agree with the AOV and multiple range tests. When disagreements occur between those tests, rank data are best reported.

Over many years of analysing QDA data, we developed several proprietary norms to guide our examination. We have norms for mean scale difference and mean rank difference where we expect to see a significant difference. The same is true for the number of subjects needed to rank one product more intense than another product before we should expect significant differences. Several non-parametric tests for rank data and proportions are available for significance testing. Chi-square (χ^2) and binomial tests serve this purpose quite well. We encourage other researchers to develop norms based on their own results. Such norms are guidelines or expectations to assist in the statistical examination of QDA data.

Correlations and multivariate techniques identify similarity in attribute and product groupings – information that is useful for reducing attribute redundancy or understanding which attributes group together.

One benefit of Tragon QDA is that the number of subjects participating on a panel, the training and the scoring all contribute to quantification and use of parametric and non-parametric statistical testing.

8.6 Advantages

Descriptive analysis methods are generally classified into two fundamental approaches upon which all other methods are based. There are the profile or technical expert approaches (e.g. flavor profile, texture profile and Spectrum™ Method) that teach subjects using a technical language. Then there is the QDA or consumer behavioural approach, which is designed to understand words consumers actually use to describe a product category, to understand their sensitivity to perceived differences, and to determine how product changes may influence perception.

There are a variety of advantages to the Tragon QDA methodology.

8.6.1 Uses Common Everyday Language

The language used by the panel is a common, everyday language that is easily understood by the product development team as well as by marketing, sales and consumers. Since the method uses consumers who are likers and users of the product category, and selects those with demonstrated sensory acuity, the resulting language and data are easily combined with consumer-affective results to better understand how the products' sensory experience influences consumer behaviour.

8.6.2 Provides Rapid Results

Quantitative descriptive analysis is a relatively rapid descriptive analysis method, requiring only 8–10 hours of group discussion for a naive panel to create the comprehensive language and definitions of terms. This approach requires about 4 weeks from start to finish to recruit, screen, train and evaluate an array of products. Subsequent panel and language development time is reduced once the evaluation techniques are clearly understood and an initial language has been developed.

8.6.3 Requires No Standardized Reference Library or Subject Calibration

There is no need to maintain a vast standardized reference library or to calibrate panellists with subjective numerical values. Maintaining standardized references and doing continual panel calibration has proven to be a time-consuming and costly task for sensory professionals, and there are no data to scientifically support the assigned numerical values or the benefits of this process. In addition, standardized references are moving targets; they vary batch to batch and time to time, as all products do, and are not globally available.

8.6.4 Designed to Approximate Typical Consumer Usage Behaviour

Products are evaluated following typical usage behaviour expected for that product. For example, in personal care, hand lotions would be applied to the hands, body lotions applied more broadly, face creams on the face, and so forth, and this may be conducted in the laboratory or at home – not on a prescribed site on the forearm with prespecified evaluation protocols that create extensive behaviour modification (ASTM 2011). QDA protocols and evaluation procedures are developed by the panel in conjunction with the qualified panel facilitator and are designed to maintain as close a relationship to actual usage behaviour as possible.

8.6.5 Provides Extensive Data Analysis

Replications are an important part of any scientific procedure. They not only allow for more product sampling, but also provide data to evaluate subject performance for repeatability, reliability and validity of results.

8.6.6 Criticism of Tragon QDA

It is unlikely that any descriptive method is without its critics and limitations; QDA and now Tragon QDA are no exception. The name 'quantitative descriptive analysis' rightfully asserts that statistics and statistical software (or access to it) is necessary, as is having practitioners knowledgeable and comfortable with statistical output. Inaccessibility to those resources would be a limitation.

Replications are limiting because they take time and resources; however, they are one price a scientific endeavour needs to be willing to accept to utilize the measurement and analysis benefits they afford.

Some developers and product researchers consider a technical language superior to a consumer language. However, we choose to encourage those product developers to understand the consumer better by viewing product sensory perceptions through the eyes of the consumer.

8.7 Case Studies

In this section, the goal is to share three case studies that benefited from using Tragon QDA to meet different business objectives. The first case study on hand lotion demonstrates the ability to use QDA in a laboratory or natural environment (i.e. home) testing. The advantage of testing in a natural environment is that it allows the subjects to evaluate products in the way they would normally use them. Research that allows products to be evaluated in a more natural environment is well suited for certain products such as personal care items (shampoo and shaving cream), household supplies (laundry detergent and cleaning solution) and sports apparel (track, tennis court, swimming, etc.). The second case study uses the Tragon QDA method to monitor the shelf-life of a mint candy and time intensity of the sensory attributes. The third case study uses Tragon QDA to help understand the key sensory attributes of olive oils that appeal to specific consumer segments.

8.7.1 Hand Lotion

Most descriptive analysis tests are conducted in a laboratory environment to maintain control over the data collection process. Most sensory booths are typically equipped to test food products, but less ideal for certain products such as laundry detergent and shampoo. Some products may be more suitable for a home use test, for several reasons.

- A longer or extended usage time is required to evaluate the product (i.e. chewing gum for 60 minutes or overnight facial cream).
- Privacy is preferred during product evaluation (i.e. shaving, rinsing and washing).
- Some products may require longer breaks before testing the second product because of strong sensory characteristics (i.e. strong flavour or aftertaste).

In this case study, subjects were screened to ensure they were frequent users of hand lotion and had the ability to discriminate differences among products (Bleibaum et al. 2009). Qualified subjects were invited to participate in the language development process in the test facility. The hand lotion research was conducted to validate the QDA data collection in the laboratory and home environment. The same hand lotion panel evaluated the same set of lotions in the laboratory and at home. The same set of language generated by the hand lotion panel was used to evaluate the lotion in the laboratory and at home. The sensory attributes were statistically analysed and no significant differences were found between the data collected from the sensory booth and home environment (Figure 8.1). This study was successful in capturing reliable data from the panel evaluating the product at home without compromising laboratory testing protocols. Conducting the QDA data collection in its natural environment, such as at home, offers several advantages.

- Reduce testing cost.
- Maximize the test schedule without expanding facility resources.
- Minimize subject fatigue.

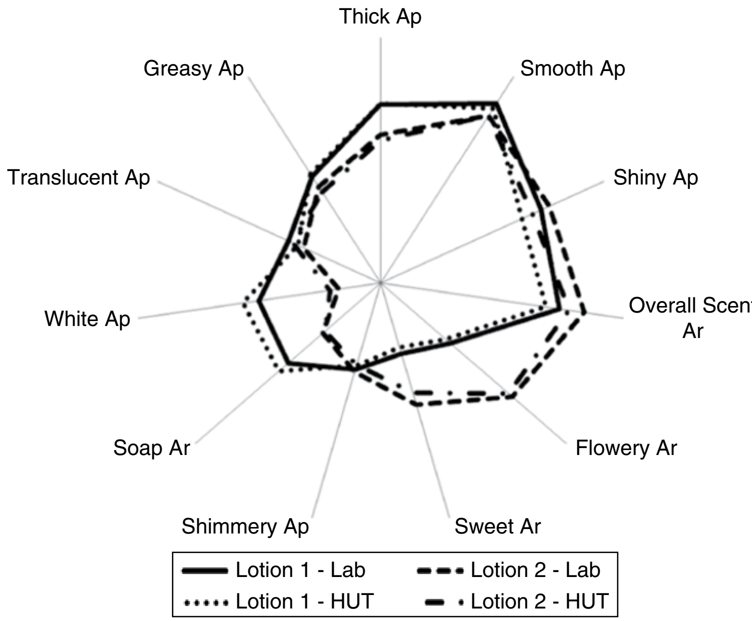


Figure 8.1 Appearances (Ap) and aromas (Ar): laboratory and home testing (scale 0–45).

A reasonable timeline completion should be established and discussed in advance with the panel; otherwise, subjects may forget or procrastinate on finishing the evaluation process.

8.7.2 Mint Candy

Product performance can shift over time due to multiple factors including reformulation, use of new or alternative ingredients, changes in processing or different production shifts. Although chemical and microbiological parameters may be key factors influencing product quality, sensory perception should not be overlooked. Tragon QDA can help validate the shelf-life of a product and understand which sensory attributes may be associated with a quality issue from consumer complaints. Some products, such as air freshener, chewing gum and breath mints, are also sensitive to time usage. Use of Tragon QDA is beneficial in monitoring the sensory changes over time, including shelf-life and the duration of consuming the product. Does the aroma, flavour, texture or aftertaste change between initial and postconsumption?

In this second case study on mint candy, the Tragon QDA method helped the project team understand the changes to the sensory characteristics to determine the optimum shelf-life and to recommend usage over time. Tragon QDA provided a comprehensive consumer-based language to describe the sensory characteristics of the mint candy. The results indicated that the candy sustained at least 15 months shelf-life (Figure 8.2). All three shelf-life products were comparable to control for up to 30 minutes evaluation for aftertaste (Figure 8.3).

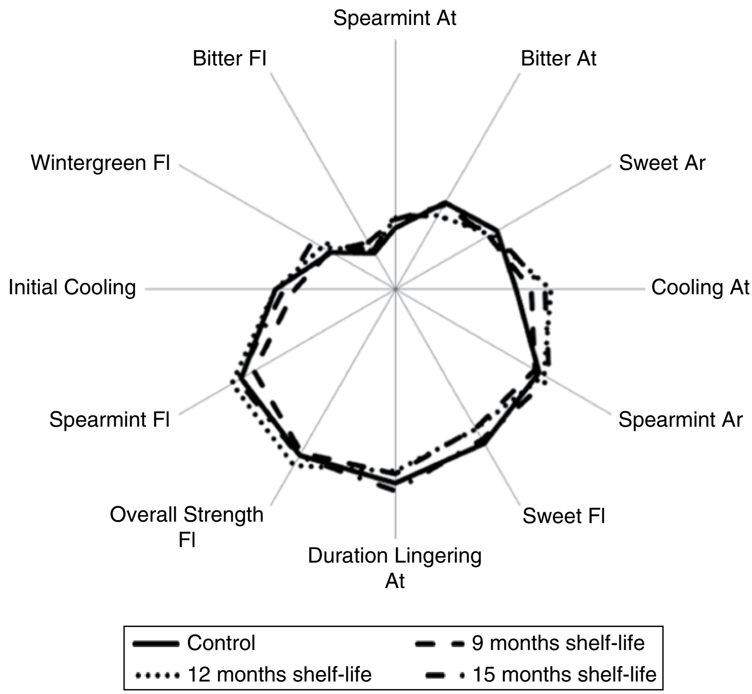


Figure 8.2 Mint candy shelf-life (Fl = flavour, Mf = mouthfeel, At = aftertaste) (scale 0–60).

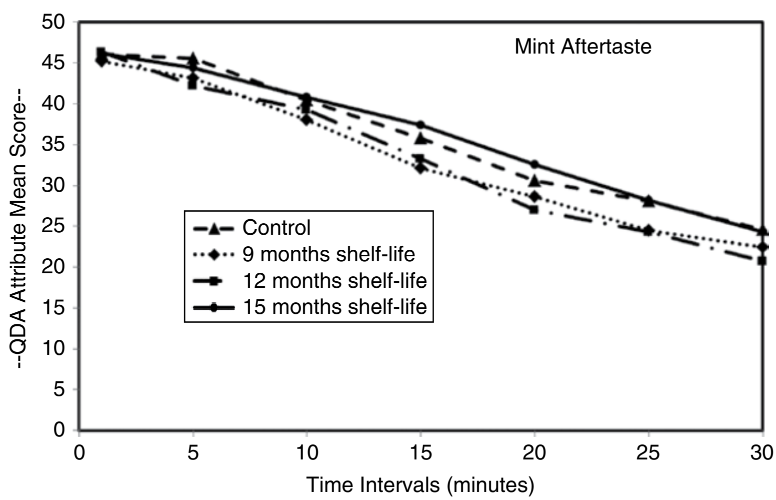


Figure 8.3 Mint candy shelf-life of spearmint aftertaste.

Compared to the control product, the shelf-life mint product achieved no significant differences, including the time intervals across key sensory attributes. The manufacturer was able to extend the shelf-life for an additional 3 months, as the mint candy satisfied the chemical, microbiological and sensory parameters without sacrificing product quality. The QDA data of the shelf-life and usage over time allowed the quality assurance team to update the product specification. The extended shelf-life allowed for validating the product specification as well as for reducing cost and waste.

8.7.3 Marketing Extra Virgin Olive Oils

Like wine, extra virgin olive oil (EVOO) comes in different styles and flavours. Many of the EVOO sold in the US market are from Italy, Spain, Greece and Turkey and used in salad dressings, cooking and bread dipping. In the third case study, marketing and product development were interested in different objectives.

- *Marketing team*: wanted to determine what sensory characteristics of EVOO best satisfy different consumer segments.
- *Product development team*: wanted to identify which sensory characteristics would be used for product optimization.

Tragon QDA is helpful in understanding the sensory basis for why consumers like or dislike certain products in a consumer test. Integrating the consumer test results with QDA data and multivariate analysis can identify consumer preference segments and sensory attributes associated with best-liked products.

Which EVOO should be included in the test? How many products should be considered? An array of EVOO was selected to include a range of sensory experience (Figure 8.4). A small team (3–4 members) screened a set of products to determine the range that best represented the EVOO categories. Team members can include product development, marketing, consumer insights and sensory staff. For this study, team members evaluated over 50 styles and brands of EVOO to select 14–15 products for the QDA test. The selected products represented several country origins.

Figure 8.4 Extra virgin olive oils.

Figure 8.5 Selected set of extra virgin olive oils.

A total of 14 EVOO was incorporated in the QDA and central location test (Figure 8.5). A qualified trained consumer panel developed a set of consumer-based language that consisted of 40 sensory attributes including appearance, aroma, flavour, mouthfeel (texture) and aftertaste to fully describe the differences of the EVOO. Using analysis of variance, key significant sensory attributes were identified, including the following descriptors (Figure 8.6):

- yellow appearance
- green appearance
- overall aroma, grassy aroma/flavour
- olive flavour/aftertaste
- bitter flavour/aftertaste
- burning sensation in the mouth
- lingering aftertaste.

Three preference segments were identified from the consumer acceptance group, based on their unique patterns of liking (Figure 8.7). Although total population preferred the EVOO Spain-3, the other three consumer segments liked different styles of EVOO. In addition, single correlations (Table 8.1) related to each segment were also examined along with developing a model for an ‘optimum product’ (optimization model not shown). The positive attributes describe the styles of EVOO that consumers preferred in each of the segments. For example, Segment 1 liked EVOO with green tint and yellow colour but disliked clear EVOO. Segment 2 favoured a simple EVOO with cooling mouthfeel and sweet aftertaste but they did not like the nutty flavour, bitterness and burning sensation mouthfeel. Segment 3 preferred EVOO with strong

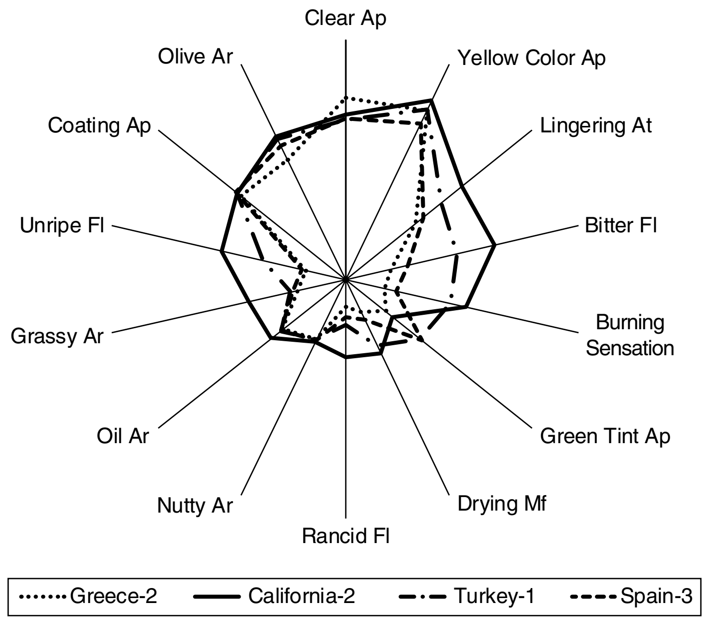


Figure 8.6 Key sensory attributes of EVOOs (scale 0–45).

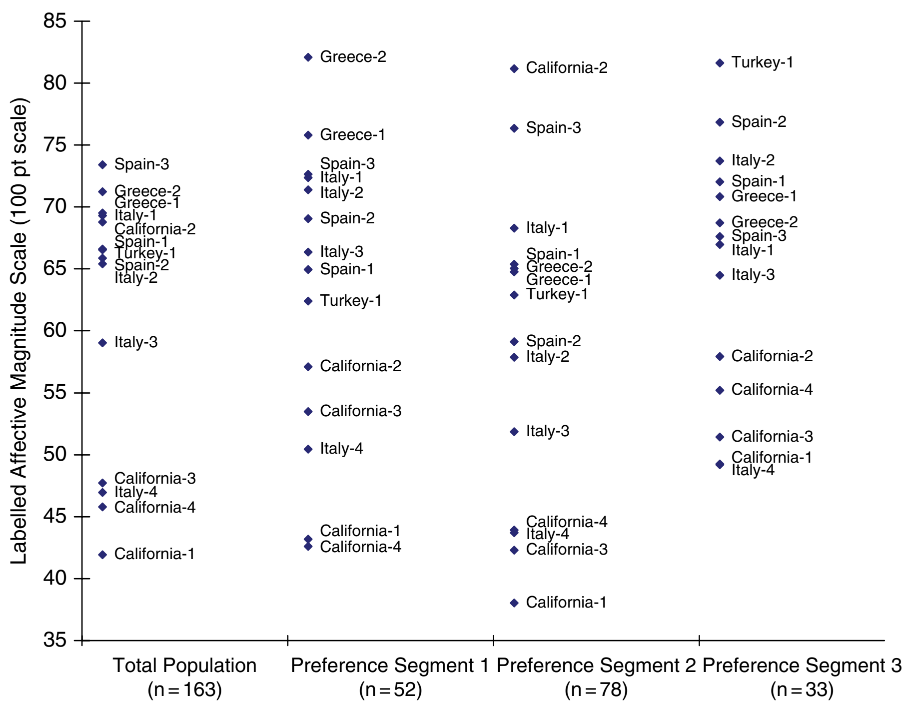


Figure 8.7 Preference segments of EVOOs. The preference segments liked different styles of EVOO than the total population. Preference segment 1 liked Greece-2, segment 2 liked California-2 and segment 3 preferred Turkey-1.

Table 8.1 Segment relationships with sensory attributes. The positive sensory attributes describe the styles of EVOO that consumers preferred in each of the segments. The negative sensory attributes are disliked by the segments.

Total population (n = 163)		Preference segment 1 (n = 52)		Preference segment 2 (n = 78)		Preference segment 3 (n = 33)	
Sweet At	0.68	Green tint Ap	0.74	Artificial Ar	0.67	Grass Ar	0.78
Coating At	−0.51	Coating Ap	0.73	Cooling Mf	0.55	Oil Ar	0.76
Grassy Fl	−0.51	Yellow color Ap	0.67	Sweet Af	0.53	Nutty Ar	0.74
Lingering At	−0.60	Olive Fl	0.55	Nutty Ar	−0.60	Nutty At	0.72
Drying Mf	−0.62	Herb Ar	0.55	Drying Mf	−0.69	Unripe Ar	0.67
Unripe Ar	−0.62	Nutty Fl	0.51	Grass Ar	−0.70	Olive Ar	0.67
Bitter Fl	−0.64	Clear Ap	−0.80	Yellow color Ap	−0.70	Grass Fl	0.66
Bitter At	−0.65			Olive At	−0.71	Herb Ar	0.64
Puckering Mf	−0.65			Lingering At	−0.73	Olive Fl	0.59
Unripe Fl	−0.68			Olive Fl	−0.73	Olive At	0.57
Burning sensation At	−0.72			Herb Ar	−0.75	Musky Ar	0.57
Burning sensation Mf	−0.74			Burning sensation Mf	−0.75	Yellow color Ap	0.57
				Puckering Mf	−0.78	Artificial Ar	−0.56
				Burning sensation At	−0.79		
				Bitter Fl	−0.79		
				Bitter At	−0.81		
				Grassy Fl	−0.83		
				Unripe Ar	−0.83		
				Olive Ar	−0.84		
				Oil Ar	−0.87		
				Unripe Fl	−0.88		

Ap, appearance; Ar, aroma; At, aftertaste; Fl, flavour; Mf, mouthfeel.

aromas including grassy and nutty. The QDA results of the EVOO provided formulation guidance to the product development team, and consumer segment information to the marketing team. The information allowed both the product development and marketing teams to collaborate and create a style of EVOO with a broad consumer appeal.

8.8 Additional Applications

Traditionally, Tragon QDA data are used by product developers to guide formulations and understand the sensory characteristics of current product and market leader products. The sensory attributes used in a Tragon QDA readily translate consumer language into technical requirement. Tragon QDA is applied in manufacturing, brand management, purchasing and advertising.

8.8.1 Product Development

At some point in the life cycle of a product, an ingredient may be discontinued and no longer available in the market. Tragon QDA studies provide strong direction for exploring new and alternative ingredients by demonstrating how closely prototypes match the current supplier. Another practical use is to incorporate Tragon QDA in the design of experiments (DOE) to determine the sensory results of the different variables. DOE studies help reveal ingredient interactions and how different levels of the ingredient affect key sensory attributes.

Analysing the Tragon QDA data from a DOE and PCA mapping result is a practical procedure in selecting products for consumer testing that are sufficiently different and consistent with the project or experimental design objectives.

8.8.2 Quality Control

Tragon QDA is a sensory tool for the manufacturing team to incorporate as part of its quality control and auditing programme. Cost improvement is an important objective for most consumer product companies. Plant managers strive for production efficiency by maximizing the production line and minimizing waste. Tragon QDA provides insights into how well the operation is performing by comparing control product to production samples. Are there sensory differences between the beginning, middle and end of production runs? Did these products deviate from the control and if so, on which attributes and by how much? Tragon QDA is used to establish sensory intensity ranges for the product specification. The product development and quality control teams develop a sensory specification as a means to deliver a well-liked and consistent product to consumers. Production shift and processing variability are to be expected, and correlating the preference-tolerance range is critical for reducing the number of products rejected by consumers. Source of ingredients can also contribute significantly to deviation from a product specification, and Tragon QDA data identify the size of these deviations. As mentioned earlier, the manufacturing team cannot ignore cost improvement. It constantly challenges less expensive, alternative ingredients. To understand the performance of these ingredients, Tragon QDA fingerprints the product's sensory characteristics to demonstrate product performance and establish sensory limits.

Plant managers are responsible for addressing consumer complaints, and Tragon QDA can be used to link sensory data to consumer complaints. Because the sensory attributes are derived from a screened and trained consumer panel, consumer complaints are easily compared to Tragon QDA data.

Manufacturers often seek maximum product shelf-life, and Tragon QDA is an effective way to capture sensory changes in product over time. It is not a surprise to find production variation between and within plants – some manufacturing plants have process equipment that can lead to product sensory differences. The corporate quality assurance team is typically responsible for investigating the degree to which these sensory differences will compromise consumer acceptance. The sensory specification developed using Tragon QDA is utilized in qualifying new equipment, as well as a new manufacturing plant. As part of a quality control audit, monitoring across all plants provides insights on regional product differences and consumer complaints that may be specific to a region.

8.8.3 Brand Marketing

Brand marketing seeks to learn what consumers like; Tragon QDA provides a quantitative sensory basis for why consumers like a product. Comparing Tragon QDA data to consumer data aids in understanding how different variables affect consumer acceptance. What are the key significant attributes correlating to likes and dislikes of the product? What sensory changes will optimize the product to improve consumer appeal? Determining the sensory direction for product launch, validating a new formulation and monitoring competitive products are all part of a strong business strategy to identify key attributes related to sensory preference. Brand marketing builds sales communication information and marketing messages, which often include important sensory attributes derived from Tragon QDA. Because the method uses a consumer-based language, the information is readily communicated to consumers and a sales team. Incorporating Tragon QDA results in an annual category review provides insights to brand marketing about reformulation requirements, and for monitoring competitive changes and product sensory advantages in the marketplace. Leveraging the Tragon QDA data, brand marketing may make changes to reduce or intensify the sensory characteristics to avoid compromising the product in the market (reduce sale, consumer alienation and other competitive-edge issues). Marketing uses the category review to explore whether a single formula will satisfy consumers or whether multiple formulas are necessary for specific consumer segments or regional markets.

8.8.4 Sensory Claims Substantiation for Advertising

When considering ad claim testing in support of brand marketing, Tragon QDA is an important research tool that can be used to substantiate sensory-based claims. Claims such as 'smoother shave', 'softer feel', 'more X (e.g. chocolate, vanilla, butter, etc.) flavour', 'longer lasting', 'crunchier and crispier' and other sensory claims may require sensory substantiation to establish product or ingredient performance. ASTM International has published a *Standard Guide for Sensory Claims Substantiation* (ASTM 2007) which is updated on a regular basis, and guidelines for using sensory research are outlined in that document.

8.8.5 Procurement Team

The purchasing team collaborates with product development and manufacturing to qualify ingredient suppliers by establishing ingredient identity and sourcing alternative ingredients that offer economic and/or environmental benefits. Tragon QDA data allow comparison of a gold standard product to suppliers' submissions of prototypes. Often a lower-cost ingredient has a limited shelf-life, lower potency and impurities that can compromise product quality and marketplace performance. Understanding these key similarities and differences, the purchasing team has the evidence to challenge suppliers to deliver a physical and sensory match. It is imperative to capture all sensory modalities. Frequently, new or alternative ingredients can affect different sensory attributes, not only in flavour, aroma or texture. Investigating all the sensory characteristics enables every phase of the operation – purchasing, product development, manufacturing and brand marketing – to make clear decisions without impacting product performance in the market.

Scientific substantiation is required when making a sensory performance claim about an ingredient or product, such as 'more chocolate flavor', 'long-lasting mint flavor', 'strong fruit aroma', etc. Tragon QDA data are quantitative, validating and confirming whether the sensory attributes are significantly more or less intense. As mentioned previously, direct comparison to the consumer is possible because the attribute languages are generated by the trained consumer panel and can be easily transcribed to support ad claims. The panel data can be applied not only to food and beverage products, but also to personal care, household products, office supplies, electronic gadgets, clothing and shoes.

Tragon QDA quantitative sensory data protect and monitor a product's sensory advantages. They answer such questions as whether the competitor reformulated to match the market leader or made significant changes to the formula. Maintaining a sensory advantage over the competitor is an important criterion for success and Tragon QDA helps to capture and quantify these sensory differences.

Regardless of the various stages of product development and maintenance, Tragon QDA is a useful sensory tool with multiple applications. The various teams involved in developing a new product, reformulating a new product or substantiating the sensory characteristics of the product in the market can easily use Tragon QDA data to provide better understanding and more insight into consumer perceptions.

8.9 Summary

Use of Tragon QDA has helped a wide range of companies to make successful product development decisions and grow their categories. Tragon QDA is a dynamic system in which the panel leader must make numerous decisions when organizing a panel, whether one is recruiting and screening a new panel, reviewing performance records for the next panel, deciding on the appropriate design, when and where products will be evaluated, and, last but not least, reporting results. Without sufficient knowledge about human behaviour and the objectives for a test, inadequate or incorrect product decisions could be reached to the detriment of the science and the product category.

However, when conducted following the principles outlined in this chapter, QDA continues to be effectively applied in many phases of product development including reformulation, quality control, optimization and new product development. The method offers valuable information by revealing sensory characteristics of the product that are readily related to consumers.

8.10 Future Development

Tragon QDA results provide powerful insights into the mind of the consumer. Since this method uses a consumer-based language and can easily be adapted to a wide array of products across the globe, the future is bright. There are several emerging trends for future development.

Direct data entry and portable systems, along with more sophisticated and rapid data analysis and reporting, will continue to be further refined. Having a panel evaluate products at home or in a typical product usage environment allows for considerably more detail as to how a product performs. Data collection out of the sensory laboratory/booth location is currently used globally and will continue to expand and be enhanced in the future.

Some comments also are warranted about the increasing use of automated data analysis systems. Tragon QDA tests yield a large sensory database, as much as 10 000 or more data points, enabling a wide range of statistical analyses. One of the main features of the QDA methodology (Stone et al. 1974) was the use of statistical analysis of the data, which represented a significant development for sensory evaluation at that time. With the wide availability of statistical packages and of PCs, sensory professionals have unlimited and low-cost resources, providing almost instant capabilities for obtaining means, variance measures, analysis of variance and a variety of multivariate analyses techniques. Panel performance is a fundamental process within Tragon QDA and we now have more sophisticated data analysis software that builds on the original QDA programmes. Enhanced features include more summary charts, tables and figures, to more rapidly guide moderators on panel performance as well as quickly determine significant differences between products with report-ready spider plots.

However, as with any readily available resource, statistics often are misunderstood and/or misused, particularly when the panel leader uses multivariate procedures without determination and sufficient study of the quality of the basic data. It is one thing to generate panel performance measures but quite another to understand the information and make the recommendations necessary to provide for a robust descriptive analysis programme.

In the future, Tragon QDA data will align more closely with the corporate business strategy. Beyond guiding product development, results will increasingly be used by marketing and innovation teams to manage or create brands that target specific consumer preference segments. The method will continue to expand into new categories with more focus on the total product experience, including emotions associated with packaging, consumer communication and advertising.

The development and use of predictive models of consumer product behaviour have had, and will continue to have, a strategic impact. This is because of the specific and immediate value it has to a company and its effectiveness at demonstrating a higher degree of sophistication than was realized for descriptive analysis.

To enhance further the success of this consumer-based descriptive analysis method, sensory professionals must move from a passive, service-oriented programme to an active, results-oriented resource. This transformation requires a concerted effort from the sensory professional who must anticipate needs and respond in a timely manner. Sensory professionals must support brand strategies and continue to educate others about the strategic use of sensory information. Tragon QDA programmes have proven beneficial for many companies and we expect this experience will continue. The methodology is well defined and it is the strategic application that will gain companies the most benefit.

For these reasons, we are optimistic about the future for sensory evaluation in business and as a science.

References

Anderson, N.H. (1970) Functional measurement and psychological judgment. *Psychological Review*, **77**, 153–170.

ASTM (2007) *E1958-07 Standard Guide for Sensory Claims Substantiation*. West Conshohocken: ASTM International.

ASTM (2011) *E1490-11 Standard Practice for Descriptive Skinfeel Analysis of Creams and Lotions*. West Conshohocken: ASTM International.

Bleibaum, R.N., Robichaud, J. & Tao, K.W.C. (2009) Exploring potential to increase descriptive analysis capacity by creating alternate data collection environments. Oral Presentation, Pangborn Symposium, Florence, Italy.

Caul, J.F. (1957) The profile method of flavor analysis. *Advances in Food Research*, **7**, 1–40.

Duncan, D.B. (1955) Multiple range and multiple F tests. *Biometrics*, **11**, 1–42.

Keuls, M. (1952) The use of the 'studentized range' in connection with an analysis of variance. *Euphytica*, **2**, 112–122.

Mecredy, J.M., Sonnemann, J.C. & Lehman, S.J. (1974) Sensory profiling of beer by a modified QDA method. *Food Technology*, **28**, 36–41.

Rogers, C. (1959) A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In: Koch, S.(ed.) *Psychology: A Study of a Science. Volume 3. Formulations of the Person and the Social Context*. New York: McGraw-Hill.

Schutz, H. (1983) Multiple Regression Approach to optimization. *Food Technology*, **37**, 46–48, 62.

Sidel, J.L. & Stone, H. (1983) Introduction to optimization research – definitions and objectives. *Food Technology*, **37**, 36–38.

Sidel, J.L., Woolsey, A.L. & Stone, H. (1975) Sensory analysis: theory, methodology, and evaluation. In: Inglett, G.E. (ed.) *Fabricated Foods*. Westport: Avi Publications, pp. 109–126.

Siegel, S. & Castellan, N.J. (1988) *Nonparametric Statistics for the Behavioral Sciences*, 2nd edn. New York: McGraw-Hill, p. 266.

Stone, H. & Bleibaum, R.N. (2009) Sensory evaluation. In: Campbell-Platt, G. (ed.) *Food Science and Technology*. Chichester: Wiley-Blackwell, pp. 323–331.

Stone, H. & Sidel, J.L. (1981) Quantitative descriptive analysis in optimization of consumer acceptance. Presented at Eastern Food Science and Technology Conference on 'Strategies of Food Product Development,' Lancaster, Pennsylvania.

Stone, H., Bleibaum, R.N. & Thomas, H.A. (2012) *Sensory Evaluation Practices*, Fourth Edition. Oxford: Elsevier, pp. 233–289.

Stone, H,. Sidel, J.L., Oliver, S., Woolsey, A. & Singleton, R.C. (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology*, **28**, 11, 24, 26, 28, 29, 32, 34.

Yandell, B.S. (1997) *Practical Data Analysis for Designed Experiments*. Boca Raton: Chapman & Hall/CRC.

CHAPTER 9
Spectrum™ Method

Clare Dus, Lee Stapleton, Amy Trail, Annlyse Retiveau Krogmann
and Gail Vance Civille

9.1 Introduction

Descriptive analysis methods began with the flavor profile* method (Caul 1957) developed by the Arthur D. Little Company in the early 1950s (see Chapter 7). Several food and pharmaceutical companies developed flavor profile panels over the next several decades. General Foods (now a part of Kraft) maintained several flavor profile panels for research and development and for quality purposes and began to adapt the flavor profile method to the various products and applications that needed detailed flavour descriptions. These adaptations included the following: expanding the flavor profile seven-point scale to 14 points; using a control sample as a calibration sample in each tasting session; using intensity references as well as references for the terms or attributes used in the description; training panellists extensively and validating them for use on specific product testing; providing rigorous treatment of product procurement and preparation to insure product sampling integrity.

In addition, the texture centre at General Foods was developing the texture profile method, based on the flavor profile method and underpinned by the texture technology developed by Alina Szcsesniak and her team (Civille & Szczesniak 1973; Szczesniak 1963; Szczesniak et al. 1963). With this foundation in flavour and texture profile understanding and application to documenting products in development and operations, Gail Vance Civille developed the Spectrum™ descriptive analysis method during the 1970s and presented the method at the 3 IFT Sensory Evaluation Courses in 1979. The Spectrum Method incorporates the rigor of the training and structure of the flavor and texture profile methods and then adds a more refined scale (over 150 points of discrimination); application of statistical methods to the descriptive data; and expansion to products outside

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

food and beverage, such as personal care products (both skinfeel and fragrance) and paper and fabrics (both fragrance and tactile feel).

In the last 20 years, the Spectrum Method has been used in government labs to document the sensory properties of commodities and products that use commodities as ingredients. Academia uses the Spectrum Method to monitor and track basic and applied research in foods, personal care and paper products. However, the bulk of the application of the method is in R&D departments of consumer product companies globally to document sensory properties for product development and quality control with accuracy and reliability.

The Spectrum Method supports the interest of product developers who need clear documentation of a product's attributes and their intensities that allows product profiles to be compared across products and across time with the same level of confidence that the researcher expects from instrumental data. Descriptive analysis data should provide a product profile that allows the product developer to see the direction that process, ingredients or time have had on the product's sensory characteristics.

9.2 Theory: History Based on Flavor and Texture Profile Methods

The Spectrum Method used the foundation of both the flavor profile method and the texture profile method in its base design. This foundation was then expanded and refined to produce what it is today.

9.2.1 Understanding the Flavor Profile Method

The flavor profile method had its origins in the late 1940s, developed by Arthur D. Little, Cambridge, MA (Caul 1957), and involved the careful selection of panellists who were then trained using attribute references and intensity scale understanding. The resulting trained panel was able to describe food, beverage and pharmaceutical samples in distinct aroma and flavour terms, as well as rate corresponding intensity, order of appearance and any aftertaste characteristics. Panellists were prescreened using physiological acuity exercises to test for basic tastes, olfaction and intensity discrimination. There were also general screening questions involving ability to describe foods/flavours, interest in the project and availability for participation.

The training involved showing the panellists the attribute reference samples within the product array needing evaluation. Discussions were aided by a panel leader who would make sure that references were clearly defined and understood by each panellist. Intensity of these referenced attributes was rated using a five-point flavour profile intensity scale. Samples were evaluated one at a time with final profiles derived from panel leader-led consensus discussions.

Only consensus profiles were outcomes from this descriptive method. This was considered a limitation by some since the trained group was small (usually 4–6 panellists) and subsequent statistical analysis was not possible. Other limitations included the panel's ability to discriminate at a detailed level due to a low degree of discrimination from a short five-point scale (eventually changed to seven points) (Keane 1992).

9.2.2 Understanding the Texture Profile Method

In the early 1960s, the product evaluation and texture technology groups from General Foods researched a way to refine texture evaluations of food. Using the foundations from the flavor profile method, this group developed the texture profile method to define textural parameters of foods (Skinner 1988). This texture profile method used underlying rheological properties to define and differentiate texture among solid foods. All original texture scales were correlated with instrumental measures on the GF Texturometer (Brandt et al. 1963; Szczesniak 1963; Szczesniak et al. 1963). Panellists were screened and selected based on their ability to discriminate the texture attributes in the target application.

During the training phase, the panellists were shown a broad range of attribute and intensity references across the texture terms before applying the resulting understanding to the product category. Panellists were also involved in defining the terms and specific evaluation procedures which helped limit variability that historically was part of texture evaluations. The texture references were used to set intensity scale values for future product evaluations. The texture profile method introduced an expanded 14-point intensity scale from the original seven-point flavor profile scale. Unlike the flavor profile method, the texture profile method used either group consensus evaluations or individual evaluations, which could be analysed statistically.

This method has been updated, expanded and incorporated into the Spectrum Method for texture evaluations of not only solid foods (Muñoz & Civille 1998) but also semi-solid foods (Civille & Liska 1975; Civille & Szczesniak 1973), beverages, skinfeel (ASTM 1997; Civille & Dus 1991; Schwartz 1975), fabric and paper goods (Civille & Dus 1990).

9.3 Traditional Methodology for the Spectrum Method

Descriptive analysis is designed to document the appearance, aroma/odour, flavour, texture, sound and texture/tactile properties of a product in terms of the attributes, perceived properties and the intensity (the strength of each attribute). This information or profile provides product development, manufacturing and marketing with clear information about products, prototypes and production samples of interest. The comparison of product profiles allows the professional to make decisions about whether the research or production has created products that are similar in attributes or different and, if different, in which attributes and to what extent they are different. With these results, product developers can see what changes need to be made to develop the target product: the same product (despite the need to change ingredients, process or package) or a product with new or improved characteristics.

The Spectrum Method was developed during the 1970s with several panels across several food and personal care industries trained and validated in the method. The method was first presented during the 1979 IFT sponsored short course on sensory evaluation. Gail Vance Civille was invited to present the section on descriptive analysis. The method was not yet called the Spectrum Descriptive Analysis Method as Sensory Spectrum had not yet been incorporated. Initially, panellists rated intensity on a ballot with 15 cm lines. Lines were measured with a digitizer or by hand. Because panellists had precise numerical references, they often complained that the line scales did not allow the precision they needed to see small product/sample differences. Thus, the scale was converted to a straight number rating system (0–15 measured in tenths = 151 points of discrimination) in the early 1980s.

Upon the incorporation of Sensory Spectrum in 1986, this evolving descriptive method was officially called the Spectrum Descriptive Analysis Method. It was different from methods such as the QDA method since, like the texture and flavor profile methods, the panels had references for both attributes (e.g. shown a controlled roasted peanut for 'roasted peanut' flavour) and intensity (e.g. shown intensity references aromatics or tastes that cover the rating scale).

9.3.1 Why the Universal Scale? Scale Assumptions

The Spectrum Method employs a universal scale, in which the scale boundaries are set considering the full space or 'universe or world' of finished products and their attribute intensities. The highest intensity point therefore is used as an absolute intensity and a reference of that point, that is, a 15 (or 151) on the Spectrum intensity scale. This intensity scale is based on an absolute zero intensity and ratio magnitude estimation relationships between intensity points. The other reference points along the scale are chosen to represent proportionate points (Muñoz & Civille 1998) along the scale, thus creating an absolute intensity scale for the panellists to use in determining attribute intensity. These references have been chosen and validated over 40 years by dozens of panels. Because the panellists are trained to rate on the scale proportionately (a 10 rating on the scale is understood to be twice a 5 rating on the scale), the scale can be extended past 15 as needed to describe very strong chemical feeling factors, tastes, flavours, fragrances and textures that are beyond the normal range (for example, sourness of pure lemon juice or aroma intensity of pure mint oil).

This scaling approach allows for the assessment of real product attribute similarities and differences as attribute intensities are rated relative to the absolute intensity scale and not relative to each other or relative within a specific sample category. The primary advantage is that it allows the researcher to compare attributes not only within a product category or sample set but also across product categories whether the samples are from one source or many. These intercomparisons are possible because all attributes are rated relative to the same absolute intensity references. Thus, product developers have a clear understanding of the attribute intensity ratings: a 5 rating for sweet = a 5 rating for bitter; a 7 rating for orange aroma = a 7 rating for cinnamon. The use of standard references for attributes and intensity creates an analytical tool that product developers can depend on for the same data from the same samples time after time.

This magnitude equivalence is applicable across modalities with the exception of texture, where the rate, type and direction of force application can change perceived attributes. The accuracy (based on realistic attribute definitions and intensity references) and the reproducibility (based on calibrated intensity scales with references) permit business decisions similar to those made as a result of other trusted laboratory methods. In addition, if a company produces several product categories, then one panel can evaluate any and all of the categories using a single universal construct.

9.3.1.1 Qualitative References and Lexicon Development

A qualitative reference has three features: a name that is specific to the attribute, a physical reference that demonstrates the attribute and a definition to explain the term (Civille & Lawless 1986). Developing a lexicon using the Spectrum training process ideally takes a minimum of five steps (Civille & Lyon 1996). Figure 9.1 provides an overview of the steps to develop a lexicon. These training/development steps with the panel and panel leader help ensure a clearly defined lexicon for use in future sample evaluations.

The first step is to collect 6–12 'frame of reference' products (commercial) and samples (prototypes) that have a broad representation of the flavour, fragrance or texture attributes of the category being developed. The second step is to present and review these 'frame of reference' samples with the panel. The panellists individually taste and generate a list of descriptors for the frame of reference. Together with the panel leader, the panellists each report their terms which the panel leader records for all to see. The panel leader, with the cooperation of the panellists, consolidates generated terms into groups (e.g. dairy group includes milky, buttery, cheesy). These groupings are generated by the panel leader and panel following the Spectrum guidelines on grouping as presented during training and based on the qualitative references, and comprise the initial lexicon for the product category.

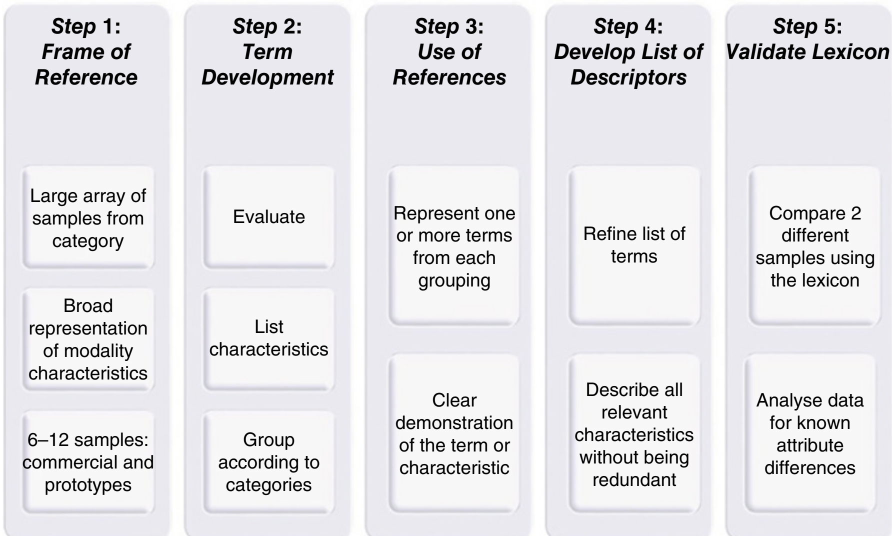


Figure 9.1 Overview of the steps to develop a lexicon.

The third step is where references are introduced to help define the generated attributes. Within each of the groupings, an external reference for one or more of the terms within should be shown to the panel. The objectives of the references are to both clarify an individual's perception of the term and to help the group as a whole reach consensus on how the term should be defined. This step can really help with small differences between terms such as all the nuances underneath a term 'sweet aromatics' (caramelized, honey, molasses, brown sugar, etc.). In step 4 the panel leader, with collaboration from the panellists, identifies an 'example' for each of the generated terms. An 'example' is a product or sample that has a predominant characteristic for discussion and demonstration but is not a singular example of the term such as used with references. The term 'caramelized' could have 'examples' presented such as sweetened condensed milk, toffee candy or ale. 'Example' samples can help the panel to understand how a lexicon term is part of a larger matrix of flavours or fragrances within products.

Step 5 is the final development of the lexicon. For this, all discussions around the 'frame of reference', 'references' and 'example' samples are practised using a group of samples. The original frame of reference products can be represented so the panellists can apply the final terms and make sure they are working. Discussions during this step further refine the lexicon and help the panel develop the best possible lexicon (see Table 9.1 for a lotion example). Even so, it should be noted that lexicons are considered working models that will continue to change and develop with new discussion and new samples. For either internal or external publication of a lexicon, the panel often compares two or three products using the refined lexicon to demonstrate the lexicon's ability to describe each sample and discriminate among or between the samples.

Table 9.1 Example lexicon – skinfeel attributes with references for body lotion.

	Definition	Qualitative reference
Rub-out attributes	These attributes are evaluated during the application and spreading portion of the protocol which dictates a set product application volume, skin area, and rub time sequence	Product dominant in the attribute
Wetness	Amount of water perceived while rubbing	Water
Spreadability	Ease of moving product over skin	Baby oil
Thickness	Amount of product felt between fingertip and skin	Eucerin Creme
Oil intensity	Amount of oil perceived while rubbing	Baby oil
Wax intensity	Amount of wax perceived while rubbing	Pomade wax
Grease intensity	Amount of grease perceived while rubbing	Petrolatum
Afterfeel attributes	These attributes are evaluated after the rub-out is completed. They are evaluated immediately and may be done at set post-absorbency time intervals	
Gloss	Amount or degree of light reflected off skin	Baby oil
Stickiness	Degree to which fingers adhere to residual product	Petrolatum
Slipperiness	Degree to which the product residue feels slippery	Baby oil
Thickness	Amount of product residue felt between fingers and skin	Eucerin Creme
Type of residue (%)	Percent of total residue identified as oily, waxy, greasy, silicone, powdery	Baby oil, pomade, petrolatum, silicone gel, talc powder

9.3.1.2 Quantitative References

The Spectrum Method 151-point intensity scale is anchored with intensity references. An intensity reference demonstrates the attribute of interest at a specified strength within the full universal intensity scale. A quantitative reference is generally part of a set of 3–5 references that demonstrate intensity of the stimuli across the range from low to high. For food evaluations, there are published intensity references for basic tastes, aromatics and textures across and beyond the 0–15-point scale (see Table 9.2 for a food example). There are also published intensities for discrete attribute terms for skinfeel and paperfeel methodologies (Table 9.3). As stated earlier, this intensity scale is based on an absolute zero intensity and ratio magnitude estimation relationships between intensity points. Therefore, if an intensity reference is numerically twice as intense as another, the assumption is that the perception of that attribute intensity is twice as high. The intensity anchors on the scale are also consistent across modalities and product.

The Spectrum Method uses these scales to provide boundaries for the sensory experience, allowing panellists to understand strength in context with multiple attributes, products and product categories. References are generated from collective data over several replicates of several panels (Meilgaard et al. 2007). When a specific reference is not available or has changed, new references are

Table 9.2 Universal scale references for aromatics and basic tastes.

Intensity	Aromatics	Sweet	Salt	Sour	Bitter
2.0	Aromatics in vegetable oil	2.0% Sucrose solution in water	0.2% NaCl solution in water	0.05% Citric acid solution in water	0.05% Caffeine solution in water
5	Cooked apple note in Motts Applesauce	5.0% Sucrose solution in water	0.35% NaCl solution in water	0.08% Citric acid solution in water	0.08% Caffeine solution in water
7.5	Minute Maid Frozen Orange Juice from concentrate				
10	Grape note in Welch's Concord Grape Juice	10.0% Sucrose solution in water	0.5% NaCl solution in water	0.15% Citric acid solution in water	0.15% Caffeine solution in water
12	Cinnamon note in Big Red Gum				
15		16.0% Sucrose solution in water	0.7% NaCl solution in water	0.20% Citric acid solution in water	0.20% Caffeine solution in water

Table 9.3 Example skinfeel texture intensity reference scales.

Attribute	Intensity reference
Wetness intensity	
0	Talc
2.2	Petrolatum
6.0	Vaseline Intensive Care lotion
9.9	Water
Thickness intensity	
0.5	Isopropyl alcohol
3.0	Petrolatum
6.5	Vaseline Intensive Care lotion
8.7	Neutrogena Hand Cream

Scale is 0–10 in this study.

identified by evaluating against other reference points and are placed on the scale. Possible references outside the US can also be validated by a trained panel.

9.3.2 Selection of Panellists

The first step in starting a Spectrum panel is screening and selection of potential panellists. An early consideration is from where the panellists will be recruited – either internally within the company/university or externally from the community. Once that decision is resolved, a series of screening and acuity exercises is recommended for selection of a better pool of panellists for training.

9.3.2.1 Source of Panellist Pool

There are generally two sources of panellist pools from which to recruit. Panellists may be recruited from within the company/university or from the local community. There are pros and cons for each source. Inside employees are convenient because they are usually already working in the building or on the workplace campus. The knowledge they gain from the trainings can also be brought back to their own research. Inside employees, however, have to alter their current work schedule and responsibilities to attend the training and panel sessions. Since being a Spectrum panellist is not their primary job function, it can lead to them losing time on the panel. From a work benefit cost analysis, inside employees may cost more per hour than an outside contractor. There could also be a concern that it will be harder to eliminate any biases toward the samples or projects since they usually have more internal knowledge about these projects.

Panellists can also be recruited from a local community. These panellists may be more reliable on a daily basis and have more time to dedicate to both the training and panel sessions, since that is their primary job function. Externally recruited panellists also have fewer overheads, since they are being paid an agreed appropriate amount for the job (their primary job function). Panellists external to the company/university may require more training time since they are not already familiar with the products or with scientific protocol rigor. External panellists generally require schedules to be predetermined and regular; potentially, there is less opportunity for availability on short notice. Whichever source is recruited needs careful screening and acuity testing for a better selection of potential panellists (Meilgaard et al. 2007; Sensory Spectrum 1985).

9.3.2.2 Prescreening of Panellists

A standard panel size recommended for training is 18–20 panellists. At each stage of recruiting and screening, potential panellists are either dropped or do not pursue the project. Usually 60–80 (three times final desired number) people are prescreened with the goal of 35–45 (two times final desired number) people going through the final screening selection step.

One of the first considerations is candidate availability both during training and afterwards for routine panelling. Scheduled sessions for the panel training and future panel sessions should be communicated at this stage to the recruits. The goal would be to have the potential panellist available for the job a minimum of 80% of operating time for a stated time frame (e.g. 2–3 years). A second important step in prescreening covers the health and food considerations of the recruit. Health concerns may include special diets, medical conditions that are affected by diet, medications that affect taste perception, food allergies (if they are part of planned evaluations) and even dentures/dental work if texture evaluation is planned for training and evaluations. For a skinfeel panel, health concerns such as sensitivity or allergic reaction to fragrance, lotions, cleansers or cosmetics should be considered. In addition to health conditions, it is also advised to screen each applicant using basic descriptive ability and interpreting scaling exercises. These are two fundamental skills that will be needed during training and evaluations. Finally, it has been found that adding a layer of questions on interest in foods/personal care products and limiting product aversions (picky eaters/picky with skincare) can help with maintaining longer term panellist employees (Stoer et al. 2002). If panellists are recruited internally from within the company/university, this is the time to secure approval of supervisors for time commitment to the project. The prescreening process is usually completed through a questionnaire that may be administered written, by phone or by an internet interface (Meilgaard et al. 2007; Sensory Spectrum 1985).

9.3.2.3 Screening of Panellists and Final Selection for Training

Recruits selected from the prescreening phase are asked to participate in an acuity screening session and interview in person. The many goals of this screening phase include gauging physiological acuity, ability to follow written and verbal directions, comfort level of working in a group and their interest in the project and to re-emphasize the time commitment. For food evaluation (Meilgaard et al. 2007), physiological acuity testing involves testing for basic taste perception, detection and description of a selection of flavour/aromatics (goal of 80% correct) and intensity testing of flavour and texture attributes in a simple modality (ranking test). For a skinfeel or paperfeel panel screening, the panellists are screened on their ability to describe differences in experiences of use of products in the category of interest (e.g. application of lotions or cleansers on the skin) or of pretreated substrates (e.g. feeling hair swatches washed with different shampoos or treated fabrics). Recruits can also complete intensity ranking tests for different textures substrates (e.g. paper towels, fabrics or hair swatches). During the interview, the food panel candidate is questioned again on general interest in foods, flavour and learning to describe foods as well as another check on time commitment and availability. Similarly, a non-food panel candidate is questioned as related to that field.

Finally, a mock panel session can be held both to give the candidate an experience of a panel session and to allow interviewers to observe group dynamic skills. The top candidates are selected from screening with all parts taken into consideration. A strong candidate is one with the ability and interest to learn the skills needed as a panellist and to perform the duties for an extended period of time (Meilgaard et al. 2007; Sensory Spectrum 1985).

9.3.3 Importance of Panel Leader

The panel leader of a Spectrum trained panel plays a very important role in maintaining panel integrity and in ensuring the quality of the data output. The panel leader for a Spectrum panel is more involved in the ongoing maintenance of their panel compared to some other descriptive analysis techniques. They are responsible for setting guidelines, leading group discussions, confirming test controls are in place, providing references, clarifying attribute definitions and monitoring panel performance for assurance of higher quality data output.

9.3.3.1 Qualifications for a Spectrum Panel Leader

A well-trained Spectrum panel leader ideally has participated in a Spectrum training. This enables the individual to experience the training philosophy and hear questions and discussion from both the trainer and the panellists participating in the sessions. An alternative path to the needed level of understanding is to have the individual study the Spectrum philosophy and observe a working panel in session for several months. The Spectrum panel leader background should include knowledge of lexicon development, the Spectrum scales, and evaluation protocols for the samples planned for testing; ability to manage, teach and communicate with a group; and some personal tasting or tactile evaluation experience. The Spectrum panel leader also adds to the role by having a basic (or greater) understanding of statistics so that they can interpret the panel data and monitor panel performance (Sensory Spectrum 2012).

9.3.3.2 Panel Leader Role in Setting Panellist Guidelines

An important function of the panel leader is to set guidelines and communicate them to the panellists. The guidelines should be clear and detailed so that the panellists understand the job function and what is expected for good job performance. Guidelines include such topic areas as pay, potential pay increases or bonuses, scheduled work hours (including any anticipated flexibility around the schedule), time off allowances and how individual performance will be monitored. The panel leader should practise positive leadership and actively participate in the job role. This may be accomplished by being professional, prompt, well prepared and having a positive attitude during the session. This helps boost panel morale and leads by example.

9.3.3.3 Leading Panel Discussion

One of the main roles of a Spectrum panel leader is to lead panel discussions during training and consensus sample evaluations. The discussions should be conducted so that the panel is working together as a team to come to a uniform, agreed conclusion. The panel leader needs to guide without leading or influencing the outcome. They should guide the discussion with limited positive or negative judging of the attributes suggested or without any attempt to direct the conversation or outcome of the profile/lexicon. The goal is for the terms to be concise and accurate to the descriptive perception and informed by the qualitative references. This can be accomplished by keeping the discussion productive and moving toward the objective of the project while working to achieve agreement among team members' opinions. The panel leader should recognize and acknowledge both individual and group performance during the evaluations. It should take place in a comfortable environment with positive reinforcement and shutting down of negative behaviour (Sensory Spectrum 2012).

9.3.4 Panel Training

A Spectrum panel is trained in stages that build an understanding of the overall method for the panellists. The training includes a series of training weeks interspersed with practice weeks. The first training week includes an introduction to the methodology. Subsequent training weeks focus on product area modules or units that are targeted toward the Spectrum panel needs for future work. After each training week, the panellists practise the new product area modules and scaling intensities. After enough training and practice weeks have been completed to cover the main product areas of interest, the panel completes a validation study to help ensure they are ready to begin work.

9.3.4.1 Initial Training of the Spectrum Method

The first training week for a new panel is also an introduction to taste or skinfeel physiology and the Spectrum Method. The panellists participate in lectures and interactive exercises to cover the components of tasting or skinfeel physiology, as appropriate. A food panel training would include aromatics, basic tastes, chemical feeling factors, texture and appearance. A skinfeel panel training would address components of visual and tactile measurements for skinfeel and appearance. An initial flavour training covers flavour component breakdown so that panellists can begin to understand the complexities of flavour interactions; the idea that a food flavour may have top, middle and base flavour notes; and that there may be an order of appearance when flavours are noticed during the tasting experience. Timing of when to evaluate flavour notes is introduced since they could be evaluated at different stages of chewing and swallowing. An initial skinfeel training week focuses on the methodology around how to properly apply and manipulate samples and includes the experience of different textures and appearances as demonstrated with physical references (either commercially available products or specific formulations).

Scaling is taught from the very beginning with an introduction to scale types and applications. The panellists are shown how their job is to focus on intensity ratings instead of subjective liking and are shown intensity references for Spectrum scales. The Spectrum scale is explained as having an absolute base with cross-modality application within and across foods and skinfeel experiences. Panellists begin to connect their own experience or perception to an intensity so that despite any perceptual variability, they assign scale values similarly (Sensory Spectrum 2010, 2011).

9.3.4.2 Process of Training

The training process, whether a new panel or existing panel being trained in a new category, follows steps designed to both train the panellists on terms for the lexicon and to involve them in the decision process.

The panel is first introduced to a frame of reference group of products for the focused training area (for example, a selection of chicken samples that includes different parts of meat, different cooking processes and chicken-flavoured products). This frame of reference selection of products should be broad so that it covers as many potential lexicon terms as foreseen. It may include both commercial products and prepared prototypes. The panellists evaluate/experience the frame of reference products and as a group come up with terms that will help develop the lexicon. The trainer helps make sure the generated terms are perceivable, discriminating, primary, orthogonal (non-related) and non-integrative (contain combinations of attributes such as creamy or fresh). During this discussion around a specific frame of reference exercise, the trainer explains and shows examples of how the refined terms are relevant across the world of foods or the non-food category of interest and how the terms may show up on other lexicons in addition to the current working one. For example, the term 'woody' can be used as a descriptor for a note that is in coffee as well as chocolate.

The next step is to show external references for the selected lexicon terms. These references are ideally in a simple matrix, have a controlled preparation and are reproducible. References may include chemical flavourants (such as benzaldehyde), simple food ingredients (such as hydrolysed vegetable protein), simple personal care product ingredients (such as silicone gel), controlled processes (coffee at different brew times or dilutions) and sometimes commercial products if needed. The important consideration is to show an array of references so that the panellists can experience different odours, flavours, textures, tactile or appearance aspects within a term.

Finally, through using the terms developed from the frame of reference exercise and external references, the final lexicon terms are refined through group discussions. The trainer makes sure the final lexicon terms are relevant to the product category and not redundant by reviewing definitions with the panel.

This process of building the lexicon is repeated across preselected product categories during the training process. The categories used in training should contribute to the panel's understanding of different attributes (e.g. browning in food; floral in fragrance; stickiness in skin-applied products) that will be relevant for the panel's expected work areas. The panellists practise the new terms and scaling in between training sessions. The training process concludes with a final validation test to measure how the panellists are working together as a group and as individuals in understanding the terms and the process (Sensory Spectrum 2010).

9.3.5 Maintaining the Panel

Maintaining a productive and healthy panel has many facets. One of the main roles is making sure the panel is provided with clear and specific training materials so that the lexicons and references are understood by the whole panel. During sample evaluations, the panel should have clear protocols for sampling and evaluating so that error and confusion are minimized. The panel leader needs to consider all levels of test controls such as guidelines for the panellists, sample preparation, sample evaluation and room environment. It is up to the panel leader to make sure the panellists understand the instructions around all these guidelines and follow them. The panel leader also needs to recognize the role of group dynamics in this teamwork environment and try to minimize both passive and aggressive behaviour by individuals to maintain a positive panel environment.

Part of maintaining a strong productive panel is to monitor both the individual and whole-panel performance (see section 9.3.6.2), making sure the panellists understand both the attributes and the scales. Panellists perform better and take more interest in the job if the panel leader practises good listening skills, demonstrates positive leadership and gives feedback on both individual and group performance (Sensory Spectrum 2012).

9.3.6 Panel Validation

At the end of the practice segment of training, a validation study may be executed to assess panel performance and confirm if panellists generate reliable and consistent data. Validation studies are often used to document performance and identify potential areas of retraining. In addition to recommended use after training, they are also recommended minimally on an annual basis to monitor panel performance and monitor panel and panellist drift.

A validation study's purpose is to determine whether the panel as a whole is providing accurate and precise data and is thereby ready to be used as a calibrated and reliable instrument to measure the sensory properties of products. In addition, validation allows a panel leader to better understand individual panellists' performance and to ensure that panellists provide reproducible data from replication to replication and discriminate appropriately among samples.

For the study, it is crucial to consider the study design and sample selection, along with the analyses to be performed. Care should be taken in identifying criteria for decisions and highlighting a course of action for each possible outcome. Possible decisions include allowing the panel to proceed with formal evaluations or recommending additional training and practice for the whole panel on the scale or specific attributes if the panel is not deemed ready to proceed with formal evaluations. Or it may be decided that the majority of the panel is ready for evaluations with a few panellists needing further training. In this case, those panellists are given feedback on their performance, retrained separately and revalidated. Alternatively, they may be given probationary status to participate in sessions with data collected segregated from panel data and monitored with extra attention from the panel leader to ensure full understanding of the attributes. Probation is lifted once performance improves (Sensory Spectrum 2012).

9.3.6.1 Validation Study Design

To determine how well panellists discriminate among samples, it is important to carefully select samples for a validation study. Generally, 4–6 samples are used, with at least two samples showcasing larger differences and two samples showcasing smaller differences. It is also important that for a validation set as a whole, samples exhibit differences in *most* attributes. Sometimes it is necessary to spike or mock up samples to generate differentiation. For example, if a panel routinely evaluates potato chips and none of the samples chosen exhibit cardboard or painty aromatics (which are common off-notes in oily foods), they may not truly be validated on those attributes. A skinfeel panel that routinely evaluates hand sanitizers would not be validated on stickiness if none of the samples chosen left a sticky residue. The panel leader and sensory team should screen through a set of samples to choose the best validation set.

Ideally, expected product profiles should be known prior to validation testing so panel findings may be compared and contrasted to those product profiles and accuracy may be determined. Training samples with known profiles are worth considering as part of the validation set for that very reason. Once the products are selected, presentation should be randomized (if possible) and panellists should complete at least two or three replications. Usually for a properly trained Spectrum panel, two replications are sufficient unless high person-to-person variability in perception is suspected, as in the case of non-nutritive sweeteners, cooling or warming chemical feelings (such as menthol) or product absorbency into the skin.

9.3.6.2 Validation Analysis and Decision Criteria

Many statistical techniques may be used to analyse the data, including univariate and multivariate data analysis. In any case, the techniques should be chosen to assess panel and panellist ability to replicate as well to discriminate among products. Examples of diagnostics for ability to replicate may include replication effect in an analysis of variance (ANOVA) for the whole panel and standard deviation by product or range between replicates for the individual panellists. Examples of diagnostics for ability to discriminate among products may include product effect in an ANOVA for the whole panel and by individual panellist. Accuracy may be determined by comparing panel profile and individual panellist profile versus a known product profile. If no product profile is available, then consistency among panellists may be assessed (for example, by identifying whether significant product by panellist interactions exist) to highlight whether panellists are perceiving differences among products in a similar way. Panellist consistency and the ability to discriminate can also be determined by plotting the

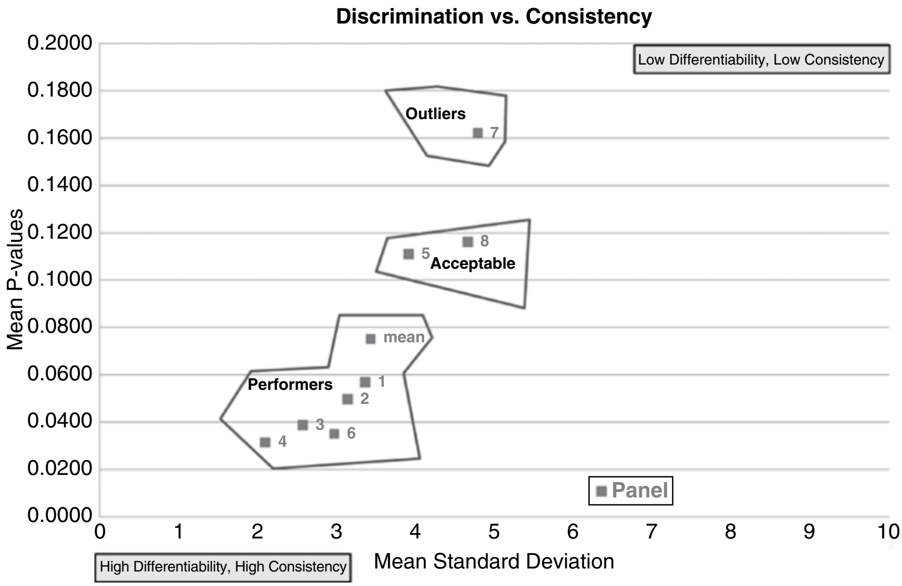


Figure 9.2 Summary plot of panel performance. Panellist 7 is an outlier.

mean P-value (across all attributes) and the mean standard deviation (across all attributes). Figure 9.2 is an example of such a plot.

9.3.7 Lexicon Development

A lexicon is a list of words that fully describe the sensory attributes of a product or product category. Lexicons are based on common terminology agreed on by the panellists through experiencing a wide array of products and attribute references. Lexicons have been developed for a wide range of products (Johnsen & Civille 1986; Johnsen et al. 1987, 1988; Lee & Chambers 2007; Lyon 1987; Retiveau et al. 2005; Talavera & Chambers 2008). The process followed includes these steps.

1 Panellists experience a broad array of products that represent the product category (called the frame of reference step). The array may include commercial products, prototypes and variables such as varied process and age. All modalities of interest (taste, touch, smell, etc.) are explored. An example for a sweet aromatics exercise would include caramelized and vanilla-flavoured products such as cookies, dairy desserts, cereals and candy.
2 Panellists discuss individual and shared experiences using their own language. The panel leader collects all the individual terms and works with the panel to generate a list of terms as a basis for a draft lexicon. Attributes are grouped according to types, similarities and categories. For example generated terms such as toasted, dough, raw grain and flour would be grouped under grain complex.

3 Panellists are exposed to attribute references to clarify terminology. Expected references may be prepared in advance, and the panel can request additional references to be seen at a later session. Examples of references for a sweet aromatics training could include molasses, brown sugar, caramelized sugar, refiner's syrup, different honeys, different maple syrups and different vanillas.
4 Panellists work with the panel leader to refine the attribute list and include definitions of terms and a physical (qualitative) reference for each attribute. Terminology that is redundant, biasing or integrated is removed through discussions. (An example of a finished lexicon is given in Figure 9.3.)
5 The generated lexicon is validated using a pair of samples, with the objective being that the lexicon adequately allows discrimination between the two samples.

9.3.7.1 Tiers, Nuances and Other Lexicon Organizations

The precise nature of the Spectrum Method allows for granular discrimination of product attribute differences that may be difficult for untrained evaluators to describe in a way that provides sufficient guidance to researchers. Often, a consumer or broadly trained descriptive analysis panel can identify major characteristics or differences yet lack language to describe more subtle, shaded differences that can have major implications for production and consumer acceptance.

Two examples are the difference between oily and greasy (words used interchangeably by many consumers for food and personal care) and describing the flavour character of smoke in processed meats (cigarette ash, campfire, phenolic, sweet smoke, woody, etc.).

9.3.7.2 Lexicon Tiers Example – Spirits

Not all descriptors have the same level of precision. Some include a group of more detailed or nuanced attributes. Such descriptors have the added nomenclature of a complex. A complex (or tier) encompasses the overarching understanding of the grouping while allowing for more specific attribute designation (nuances) where possible. Two examples are dairy complex, to include heat-processed dairy notes, milkfat and cultured dairy, and brown spice complex, to include cinnamon, nutmeg, ginger, clove and others. Initially, panellists may record attribute perceptions at the upper tier/complex level of ability and understanding; as experience and knowledge increase, the ability to detect and describe more nuanced notes from the detailed tier increases.

Sensory Spectrum has applied this concept to the development of a tier system for its spirits lexicon. Attributes progress from very general, overarching categories to more specific subcategories and then to specific attribute references. If needed, even more specificity is possible through use of chemical references. Taking one section (Table 9.4) of the spirits wheel as demonstration, the general category

Definitions of Attributes for Green Tea Evaluation		
Attributes	**Definition**	**Reference**
Green	Sharp, slightly pungent aromatics associated with green plant/vegetable matter, such as asparagus, Brussels sprouts, celery, green beans, parsley, spinach, etc.	Fresh parsley water = 9.0 (flavour) 25 g of fresh parsley, rinse, chop, and add 300 mL of water. Let it sit for 15 min. Filter and serve liquid part
Asparagus	The slightly brown, slightly earthy aromatics associated with cooked green asparagus	Asparagus water = 6.5 (flavour) Weigh 40 g of fresh asparagus, wash, dice, add 300 mL of water, cover, microwave for 3 min on high. Serve liquid part
Beany	The brown, somewhat musty earthy aromatics associated with cooked legumes, such as garbanzo beans and lima beans	Kroger Small Green Lima Beans (Kroger Co., Cincinnati, OH) = 5.0 (flavour) Measure juice out of can. Dilute: take 1 part of lima beans juice, and mix with 4 part of water
Brussels sprouts	The somewhat sharp, slightly sour, pungent aromatics associated with cooked cabbage, Brussels sprouts and cauliflower	Brussels sprout water = 6.5 (flavour) Weigh 20 g, wash, dice, add 300 mL of water, cover, microwave for 3 min. Filter and serve liquid part
Celery	The slightly sweet, green, brown, slightly bitter aromatics associated with cooked dried celery leaves	McCormick Celery Flakes (McCormick & Co, Inc., Hunt Valley, MD) water = 6.5 (flavour) Weigh 1.5 g, add 300 mL of water, cover, microwave for 3 min on high. Filter and serve liquid part
Green beans	A viney, green, slightly brown, woody aromatics associated with processed green beans	Del Monte Cut Green Beans (Del Monte Foods, San Francisco, CA) (No Sodium) = 5.5 (flavour) Measure juice out of can. Dilute: take 1 part of green beans juice, mix with 4 part of water
Green herb-like	The aromatics associated with dry green herbs such as bay leaves, thyme, basil	Mixture of McCormick bay leaves, McCormick ground thyme, and McCormick basil = 6.0 (aroma) Mix 0.5 g of each herb. Grind using mortar and pestle. Add 100 mL of water. Mix well. Put 5 mL of herb water in a medium-size snifter. Add 200 mL of water. Cover

Figure 9.3 Green tea lexicon. Source: Lee & Chambers (2007).

Parsley	The clean, fresh, green, bitter, pungent aromatics associated with cooked parsley	Parsley water = 5.5 (flavour) Weigh 15 g, rinse, chop, put in 300 mL of water, cover and microwave for 3 min on high. Filter and serve liquid part
Spinach	The brown, green, slightly musty, earthy aromatics associated with fresh spinach	Spinach water = 6.0 (flavour) Weigh 35 g of spinach, rinse, chop. Add 300 mL of water, cover, microwave for 3 min on high
Brown	A sharp, caramel, almost burnt aromatic	Sethness AP100 Caramel (Sethness Caramel Color, Chicago, IL) colour = 7.0* (aroma)
Ashy/sooty	The light smoky/ashy aroma associated with burning tobacco such as cigarette smoke	Camel Filters (RJReynolds Tobacco Company, Winston-Salem, NC) (Turkish and Domestic Blend) Cigarette smoked filter = 7.0 (aroma) Place 0.02 g of smoked filter in a medium snifter. Add 100 mL of water. Cover
Brown spice	Aromatics associated with a range of brown spices such as cinnamon, nutmeg, allspice	Spice blend = 9.0 (aroma) Place 1/4 teaspoon of McCormick Allspice, 1/4 teaspoon of Cinnamon (McCormick & Co., Inc.) and 1/4 teaspoon of Nutmeg (McCormick & Co., Inc.) in medium snifter
Burnt/scorched	The somewhat sharp, acrid notes associated with burned or scorched vegetables or grains	FMV Wheat Puffs Cereal (Inter-American Products, Inc., Cincinnati, OH) = 7.0 (flavour)
Nutty	Nutty characteristics are: sweet, oily, light brown, slightly musty and/or buttery, earthy, woody, astringent, bitter, etc. Examples: nuts and wheat germ	Diamond Shelled Walnut (Diamond Foods, Inc., Stockton, CA) = 6.5 (flavour) Grind for 1 min on high using blender
Tobacco	The brown, slightly sweet, slightly pungent aromatic associated with cured tobacco	Camel Filter cigarettes (Turkish and Domestic Blend) = 7.0 (aroma) Break cigarette and place 0.2 g of tobacco in a medium snifter. Cover
Almond	Aromatics associated with almonds that may be slightly cherry-pit like	California grown shelled almond = 7.5 (flavour) McCormick Almond Extract = 5.0 (flavour) Place 100 mL of water in medium snifter. Add 1 mL of imitation almond extract. Cover

Figure 9.3 (Cont'd)

Animalic	A combination of aromatics associated with farm animals and live animal habitation	1-Phenyl-2-thiourea (aroma): Eastman Organic Chemicals (Rochester, NY) 1-Phenyl-2-thioures in Fisher Scientific (Fair Lawn, NJ) Propylene Glycol (character reference) 5000 ppm (0.5 g in 100 g propylene glycol). Dip perfume strip (3 mm) in the solution and place in 20 mL test tube and cap
Citrus	The aromatics associated with commonly known citrus fruits, such as lemons, limes oranges, could also contain a peely note	McCormick Lemon Grass = 4.5 (aroma) Weigh 0.1 g of McCormick Lemon Grass. Place in a medium snifter. Add 100 mL of room temperature water. Cover
Fermented	The yeasty notes that are associated with fermented fruits or grains that may be sweet, sour, slightly brown and overripe	Private Selection Burgandy Cooking Wine (Private Selection, Cincinnati, OH) = 7.0 (aroma) Dilute 1 part wine with 1 part water. Serve 1 tablespoon in a 1 oz cup, covered
Floral/perfumy	The somewhat sweet aromatics generally associated with fruit and flowers	Geraniol Pure = 8.0 (aroma) Put 1 drop geraniol in 200 mL of distilled water in a large-size snifter. Cover
Fruity	A sweet, floral, aromatic blend, reminiscent of variety of ripe fruits such as apricots, peaches	Blackberry WONF 3RA654 (McCormick & Wild Inc., Hunt Valley, MA) (character reference) Place one drop of chemical on a cotton ball in a medium size snifter. Cover
Grain	An overall grain impression that may or may not be accompanied by specific grain identities	FMV Wheat Puffs Cereal = 6.0 (flavour)
Medicinal	Aroma characteristics of antiseptic-like products, such as Band-aid, alcohol and iodine	Band-aid adhesive Bandage (Johnson & Johnson Consumer Companies, Inc., Skillman, NJ) = 6.0 (aroma) Place a bandage in a medium snifter. Cover
Mint	Aromatics associated with fresh mint; somewhat reminiscent of toothpaste The sweet, green, earthy, pungent, sharp, mentholic aromatics associated with mint oils	Fresh crushed mint leaf = 7.0 (aroma) Weigh 0.1 g of mint leaves. Crush. Place in a medium snifter. Cover

Figure 9.3 (Cont'd)

Musty/new leather	New leather (like new shoes or purses)	2-6-Dimethylcyclohexanol (character reference): 5000 ppm (0.5 g in 100 g propylene glycol) Place 1 drop of the chemical on a cotton ball in a medium snifter. Cover
Seaweed	The aromatics associated with shellfish, fresh fish and ocean vegetation	Pacific Foods Dried Sea Food (Jin Han International, Inc., Chicago, IL) (Brown seaweed) = 6.5 (aroma) Weigh 1 g, add 300 mL of water; let it sit for 10 min Jin Han International Dried Laver (Green) = 8.0 (aroma) Weigh 1 g, add 300 mL of water; let it sit for 5 min
Straw-like	The dry, slightly dusty aromatics with the absence of green; associated dry grain stems	Walgreen Finest Natural Lecithin (Walgreen Co., Deerfield, IL) 1200 mg (character reference) Break 1 lecithin softgel and place in medium-size snifter. Cover
Bitter	A basic taste factor of which caffeine in water is typical	0.02% 'Caffeine solution' Caffeine (Fisher Scientific) = 3.5 0.035% 'Caffeine solution' Caffeine = 5.0 0.05% 'Caffeine solution' Caffeine = 6.5 0.06% 'Caffeine solution' Caffeine = 8.5
Astringent	The drying, puckering sensation on the tongue and other mouth surfaces	0.03% 'Alum solution' Alum (McCormick & Co., Inc.) = 1.5 0.050% 'Alum solution' Alum = 2.5 0.10% 'Alum solution' Alum = 5.0 0.15% 'Alum solution' Alum = 7.5
Tooth-etch	A chemical feeling factor perceived as a drying/dragging when the tongue is rubbed over the back of the tooth surface	0.1% 'Alum solution' Alum = 4.0 Diluted Welch's Grape Juice (Welch's, Concord, MA) = 6.0 1:1 with deionized water 0.2% 'Alum solution' Alum = 9.0
Sweet aromatics	Aromatics associated with the impression of sweet substances such as fruit or flowers	Fisher Scientific Vanillin = 5.5 (aroma) Place 0.5 g of Fisher Scientific Vanillin in 250 mL of water in large snifter. Cover
*Intensity, 15-point scale: 0 = none, 15 = extreme. Intensities of references are not based on a universal scale and may be relevant only to green tea evaluation.		

Figure 9.3 (Cont'd)

Table 9.4 Tiers and nuances example for spirits.

	Dark rum 3	Dark rum 5
Aromatics (no nuances)		
Sweet aromatics complex	4.0	6.5
Fruity	4.0	2.0
Toasted grain	1.5	0.0
Woody complex	0.0	0.0
Spicy	1.0	2.0
Aldehydic	0.0	0.0
Alcohol	2.8	2.0
Aromatics (with nuances)		
Sweet aromatics complex	**4.0**	**6.5**
Molasses	**3.0**	0.0
Caramelized sugar	0.0	0.0
Buttery	0.0	0.0
Refiner's syrup	1.0	1.5
Vanilla	0.0	**5.0**
Fruity	**4.0**	**2.0**
Orange peel	2.0	1.0
Banana	**2.2**	0.0
Apricot	0.0	**1.0**
Toasted grain	**1.5**	0.0
Woody complex	0.0	0.0
Dried wood	0.0	0.0
Oaky	0.0	0.0
Black pepper/terpene	0.0	**2.0**
Brown spice complex	**1.0**	0.0
Nutmeg	**1.0**	0.0
Aldehydic	0.0	0.0
Alcohol	2.8	2.0

(tier 1) is spicy. It is divided into two complexes (tier 2) of black spice and brown spice. Each of those categories is further divided into specific references (tier 3) with brown spice attributes of clove, cinnamon, ginger, allspice and nutmeg.

9.3.8 Independent versus Consensus Ratings

The Spectrum Method allows for data collection using either independent panellist ratings or consensus panel ratings. For both, the same training and evaluation rigour are used. Independent rating is common for many types of descriptive analysis and is predicated on use of the same lexicon or ballot by all panellists once the evaluation series has begun. Consensus evaluation involves discussion of attributes and intensities for each sample with a focus on the attributes that are needed to clearly define sample profiles. The consensus discussion allows the opportunity to either leave attributes as complexes or break them down to individual perceived attributes. Consensus is not a calculation of average values from all panellists nor the intensity values of the most senior or most vocal panellist. The decision to include a nuanced attribute not perceived by all panellists is dependent on study objective and panellists' familiarity with the attributes. The panel leader may decide to present further qualitative references.

Consensus allows for discussion of complex attributes that may share common elements and also for addition of new terms found in a sample while the session is in progress, with all panellists defining the term in the same way (brown spice may be further defined as cinnamon and nutmeg).

Selection of the desired method of data collection is dependent on the objectives of the research. In general, independent ratings are used when a study requires statistical analysis such as ANOVA. In such cases, N = 6–12 and two or more replications are conducted. Consensus ratings are used when directional and/or nuanced information is needed. In such cases, N = 3–8 and replication may or may not be conducted. If two or more replications are completed with consensus ratings, the resulting data may be statistically analysed.

9.4 New Developments: Combining Spectrum with Other Methods

The Spectrum Method has designed flexibility for combination with other sensory methods and an ability to merge the data with other data sources. The panel as a group is trained both to give individual ratings and participate in consensus discussions. The Spectrum panel may also be used for collaborative language discussions that determine how best to describe fine differences between samples. Both of these components, the experience with group discussions toward consensus and the reinforced trainings and references for attribute language, make expansion possibilities endless.

9.4.1 Combining Spectrum Descriptive Analysis with a DOD/DFC Rating

It is fairly common, either within a specific study or across a monitored line of products from a company, to have a defined control product or sample. Descriptive analysis has historically been used to document similarities and differences between an identified control and experimental or production samples. In addition to providing a full descriptive profile of the samples in the study, a Spectrum panel can easily be trained in a degree of difference (DOD) scale. A standard DOD scale used in some industries is a 10-point integral scale with 0 = no difference and 10 = large difference. The Spectrum panel is trained on this scale by being shown ranges of differences within the product space that includes samples with small, moderate and large differences and having resulting discussions on where

Table 9.5 Example of Spectrum descriptive method used with DOD rating.

Hot chocolate beverage	Control	Test 1	Test 2
Aromatics			
Total impact	7	6.5	6.5
Dairy complex	2.5	4	3.5
Cooked milk	2.5	3	1
Non-fat dried milk/whey	0	1	3
Chocolate complex	4	3	2.5
Chocolate essence	2	1	0.8
Cocoa/earthy/burlap	2	2	1.7
Benzaldehyde	0	0	2
Cardboard	0.5	0.5	0
Basic tastes			
Sweet	9.5	9.5	10
Degree of difference (DOD)	0	3.5	5.5

to place new samples. Table 9.5 illustrates an example of the addition of a DOD rating. Test 1 is assigned an overall difference of 3.5 as it has a higher dairy impact than the control. Test 2 is assigned an overall difference of 5.5 as it has both a non-fat dried milk/whey note and a lower chocolate impact, both of which separate the flavour experience from that of the control in a more noticeable way than the difference noticed in Test 1.

Since the Spectrum panel is used to consensus discussion to reach agreement and is trained in the same descriptive attribute language, the addition of this DOD rating is easily implemented. This method is also referred to as a difference from control (DFC) in sensory evaluation textbooks and courses.

9.4.2 Assigning a Quality Rating with Narrative Support

A trained, functioning Spectrum panel may also have highly detailed and informative discussions that describe sample differences from a qualitative viewpoint only. They are already used to working together with the same language and references and participating in panel consensus discussions. A new tool for a trained Spectrum panel is the introduction of a quality rating scale to further describe product attributes in context of a target and to provide data for specific quality projects.

An example of this application supports a retail store interested in comparing its line of store brand products with competitors and national branded products in the same category. The Spectrum panel can have discussions and provide data on where the main highlighted differences exist qualitatively between their samples and the competition. Then as an extra step, the panel can provide a quality rating on samples within the study.

Table 9.6 Example of quality ratings used with Spectrum descriptive method.

Sample	Sensory quality score	Positive attributes	Negative attributes	Recommendations for sensory quality improvement
Milk chocolate brand A	6.5	High snap Noticeable chocolate essence with winey character	Very low dairy (dairy is cultured dairy rather than cooked milk); astringent, bitter, has distinct woody/ nut shell note	Concept leans towards dark chocolate (very adult); to move towards a milk chocolate, increase cooked milk, increase caramelized, decrease bitterness and astringency
Milk chocolate brand B	8	Smooth melt, dairy notes (cooked milk) and chocolate essence is balanced, has noticeable caramelized notes	Has low tropical floral note, sticks to teeth and mouth surfaces during melt	Eliminate tropical note, decrease stickiness, increase snap
Milk chocolate brand C	6	Chocolate impact, nuttiness, slower melt	Cheesy/butyric dairy note stronger than chocolate note; sticky, grainy/ particulate during melt	Decrease the cheesy dairy so that is more in balance with chocolate; decrease particulate/graininess, decrease stickiness, increase snap

When Spectrum panels have hours of experience in evaluating samples within specific categories and as a result have broad exposure to the product area, comparing product quality is appropriate. Before the study begins, either the panel or project team have a general discussion on whether certain descriptive attributes should be classified as either positive or negative (such as off-notes) and on critical product features. The panel then provides a sample quality scale rating and the narrative reasons by placing defined attributes in positive or negative categories, as illustrated in Table 9.6.

9.4.3 Product Grouping/Sorting/Napping

As stated in other examples, a trained functioning Spectrum panel is skilled at having detailed discussions around samples and attributes and is accustomed to reaching a consensus outcome. Just as product grouping/sorting exercises are used within consumer research to understand how a consumer might group together samples for specific reasons, the same research technique can be executed through a Spectrum panel. The advantage of the Spectrum outcome is that samples are grouped using a clearly defined technical language that can be understood by supporting research teams on the project. This exercise can also help define the main sample outliers and place emphasis on dominant highlighted attributes that lead to samples being grouped in a certain sequence.

9.4.4 Simplifying Understanding of Data from Across a Large Set of Samples

Sometimes it is necessary to decipher and evaluate sample data across a large number of pulls or sampling from production. Production samples, with many potential sources of variability and noise, require pulling samples across a range of different variables such as production runs/shifts, different plants or lines, different lots of raw materials or even countries of origin, etc. A Spectrum panel can evaluate the large number of samples, creating descriptive profiles. These data can then be presented as range plots to simplify the multiple comparisons that could occur in a large study (e.g. 50–100 sample pulls). The range plots quickly demonstrate where the large sources of variation occur in a set of samples and can even be compared with an external standard or control sample. This tool is especially useful in understanding production variation or factors influencing complicated designed experiments.

Figure 9.4 is an example plot where the bars represent the product variation by attribute with a control point overlay. Attribute 11 is an example of an attribute in

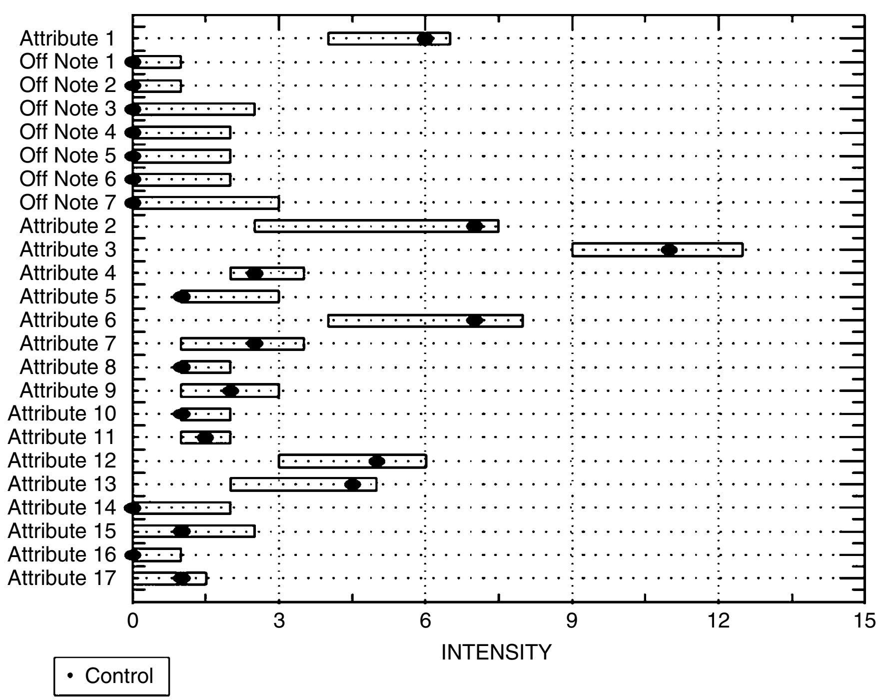


Figure 9.4 Attribute intensity range plot with control overlay.

which the sensory variability around the control is small and not likely to cause consumer complaints or possibly consumer notice. Attribute 2 is an example of an attribute in which the sensory variability is large and could cause consumer complaint. The cause of such a large variability would need to be investigated.

9.4.5 Using Spectrum Descriptive Data for Prediction of Consumer Liking/ Benefits/Performance

Spectrum descriptive data can be part of a larger project that includes consumer quantitative responses. The descriptive data can help with several stages of a study to assess understanding of which product attributes have the largest influence on consumer liking or other consumer response (e.g. holistic consumer terms such as creamy, moisturizing, rich). In a study of this scope, the first place Spectrum descriptive data may be helpful is in determining which samples should be tested by consumers. When gathering samples across a category, it is better to be broad and look at as many samples as feasibly possible, though it would not be cost-effective to submit all the samples to consumer testing. A Spectrum panel can describe a large set of samples and then by using multivariate tools such as factor analysis, the samples can be grouped to help determine how many and which specific samples should be seen by consumers. It can even help determine if there are open spaces where samples are not represented and should perhaps be created for testing.

Another stage of the project where the Spectrum data are critical is at the final data analysis and interpretation. Spectrum descriptive data are merged with consumer response data to create prediction regression models. These models can help determine if an 'ideal' sample already exists or what further tweaks to attributes are needed to reach optimal predicted consumer response. As a final check, the Spectrum descriptive panel can evaluate any prediction model samples to determine if they satisfy the attribute criteria for the most liked product.

Figure 9.5 demonstrates how these final plots show how the consumer liking and perceived benefits (in capital letters) correlate with sensory attributes (in small letters) within a product space. The proximity of sensory attributes to the consumer benefit is an indication of the degree of correlation with that benefit. In this example, there is an efficacious to liking vector with some sensory attributes cuing efficacious and some attributes being those that consumers liked. The business question centres on the strategy of changing the paradigm. In what ways might the company continue to make efficacious products and change the consumer paradigm that good for you tastes bad?

9.4.6 Combining in Context Consumer Qualitative – Within the Sequence Mapping Process

Sequence mapping (Heylmun 2003) allows for understanding of the complete interaction between consumer and product. Moment-to-moment interactions (in context, all the events within purchasing, storing, using and disposing of a product) of consumers are documented along with consumer motivators,

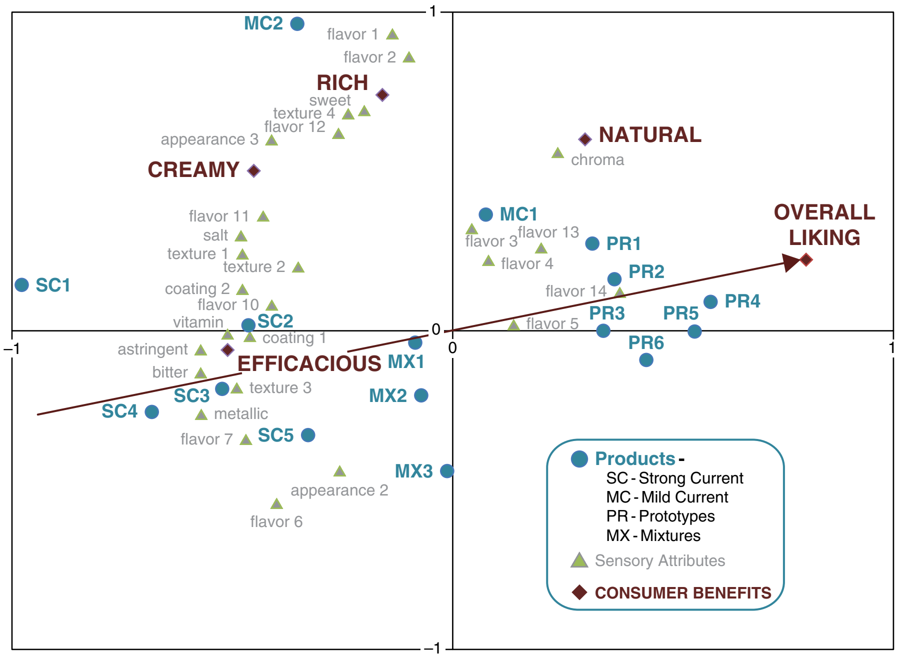


Figure 9.5 Liking drivers plot.

emotions and consumer-stated sensory attributes (at each point of interaction in context). Also included in each stage of interaction are Spectrum Method qualitative attribute descriptions which provide further product understanding. When put together, these maps of product–consumer interaction sequences reveal opportunities for either new products or refinement of existing products. Furthermore, the Spectrum descriptions provide actionable attribute information to the developer.

Each sequence map contains several time points in which the consumer interacts with the product. Figure 9.6 demonstrates the range of data collection for a single step of an exercise focusing on bathing. Review of the rich amount of collected data, by researchers or at times by the consumer group themselves, reveals patterns, images, sensory signals and related emotions that can be leveraged for ideation, development and marketing.

9.5 Statistical Analysis

The Spectrum method typically generates scaled data. Independent of whether data are collected through consensus or judgement by individual panellists, graphical display of the data is highly encouraged and may include bar charts of means, standard deviations, consensus ratings or frequency distributions.

Step Preparing to Bathe	
Event	*Set the atmosphere - lighting*
Details of Event	*Select candles, arrange candles in bathroom, light candles, adjust bathroom light*
Emotions	*Anticipation, relaxation, serenity, letting go, transformation, soothing, escape*
Sensory VOC	*Cosy looking, flickering of candles, spa-like, smells good*
Sensory Descriptive	*Intensity of light, fragrance of candles (floral, vanilla, citrus, lavender)*

Figure 9.6 Example of a sequence map template for describing an event in bathing.

When individual data are collected, they may be treated as continuous data, due to the presence of up to 151 points of differentiation along the scale and the ratio properties of the scale, in analysis and be analysed using parametric tests. Summary statistics may include means and standard deviations, and hypothesis testing procedures include analysis of variance often complemented with means separation procedures (multiple comparison tests). These methods allow researchers to highlight whether significant differences exist among samples for each attribute evaluated and, if differences do exist, to pinpoint those differences among specific samples. It should be noted that in some instances, sensory scientists may choose to identify sensory properties as present or absent instead of providing scaled intensities for those properties, in which case non-parametric tests may be performed to compare samples. Attribute examples include the presence or absence of specific residue types (oil, wax, grease, silicone/plastic, silicone/powder) in the afterfeel evaluations of some skincare products or the identification of the type of residue (fatty, dairy, chalky, oily film), roughness characteristics (gritty, grainy, lumpy, coarse) or aftertaste characteristics in the evaluation of food products. In those cases, data output is often reported as frequencies and significant differences among products are assessed.

Multivariate techniques such as cluster analysis or principal component analysis/factor analysis are also available to summarize the information provided by the panel, when eight or more samples are profiled. Cluster analysis is typically used to identify groups of products with similar characteristics while principal component and factor analysis group attributes into bundles of correlated attributes or sensory dimensions to summarize most of the variability among products.

9.6 Applications with Case Studies

The Spectrum Method is versatile. It has been used in the evaluation of fragrances, malodours, environmental odours, food, beverages, fabric for clothing, paper products, furniture, personal care products (both skincare and grooming) and oral care products. The Spectrum Method can be used to measure product attributes at one point of time as well as over time in a contextual use. In addition, as the Spectrum Method provides descriptions of sensory experiences, the Spectrum Method data can be applied to advertising claims, quality control/ assurance, product development and product ideation.

9.6.1 Case Study 1: Malodour Reduction Advertising Claim

An aircare products company wants to establish a non-comparative claim for home environment odour elimination for a new product line. The sensory department develops a three-part strategy: analytical headspace analysis, consumer assessment to define odour elimination, and expert descriptive analysis odour evaluation of target malodours and fragrance. An independent third party panel is selected for the expert testing. Collaboratively, the company and independent panel develop methodology for evaluation of target malodours using a controlled model room environment. Target home environment malodours are standardized and qualified for fidelity (being a reasonable representation of the malodour compared to an *in vivo* reference such as a malodour as it exists in real time) prior to use. Table 9.7 demonstrates early fidelity testing, with the result that three of the malodours require refinement prior to claims testing. Spectrum evaluations and recommendations guide additional development to generate the set of representative malodours. Expert assessment measures intensity of malodour and fragrance at defined time points before and after introduction of aircare treatment. Each malodour and treatment combination is assessed multiple times. A predetermined success criterion of a percentage reduction level from initial is used to determine efficacy. Reduction level is calculated from the descriptive analysis data. Data are successfully used as support for the claim.

9.6.2 Case Study 2: Developing a Lexicon and Protocol for the Evaluation of Dinner Napkins

A paper product company wants to expand its line and offer premium dinner napkins. Its descriptive analysis panel currently evaluates sensory tactile attributes for paper towels and needs to develop a ballot for dinner napkins. This ballot will be used throughout the product development phase as well as in ongoing testing. The panel reviews a wide range of commercially available paper dinner napkins in order to identify those attributes that describe the tactile differences among the samples. In addition, an evaluation protocol is established. Table 9.8 shows the lexicon developed by the panel.

Table 9.7 Example of malodour intensity and fidelity using Spectrum methodology.

Type	Total intensity	Character	Attribute intensity	Malodor fidelity	Recommendations
Kitchen – Indian food	4	Green onions/ chives	4	**4**	Add heated oil, curry type spices, fish/seafood, increase intensity
Wet dog	6.5	Isovaleric acid Hormonal	4.5 2	**5**	Add lanolin, aldehydes, and wet fur, lower isovaleric acid, add castoreum or slight skatol; intensity is acceptable
Cigarette	7.5	Stale smoke Woody; fermented tobacco	5 2.5	**8**	Meets criteria
Bathroom	4.5	Vinyl Ammonia/urine Hormonal	3.3 1 1	**2**	Reduce vinyl note; increase urine; add mercaptan; aged urine amines; faecal/skatol

Source: Stapleton & Civille (2008).

9.6.3 Case Study 3: Rating Mascara Attributes within Context/Use

A cosmetics company wants a descriptive analysis protocol to assess the similarities and differences among mascara formulations. The protocol needs to account not only for lash variations between individuals but also for noticeable wear attributes over an extended period – 6–8 hours. The latter need complicates testing as it is not reasonable (both in time commitment and in cost) for most external panellists (see section 9.3.2.1 for definition) to be available on site for the required time. The sensory department develops the protocol, ballot and intensity references for a within context/use mascara panel.

The panellists come to the facility to apply the test sample mascaras at the beginning of a scheduled workday. Prior to application, the panellists rate and/or note attribute intensity changes to their untreated lashes. Panellists then apply and rate the samples using the lexicon attributes on which they have been trained. After the session, panellists are instructed to go about their normal daily activities and then return to the facility for the extended-use evaluation at 6 hours post application. For the study, panellists receive clear instructions on application of any other face lotions and make-up as well as a list of prohibited activities or ones to be noted on their ballot (such as swimming, crying). Inclusion of baseline data allows for understanding of

Table 9.8 Lexicon for tactile evaluation of paper dinner napkins using Spectrum methodology.

Attribute	Definition; scale direction
Macro degree of embossing	The area which is covered by large or macro embossing none/low → high]
Macro depth embossing	The extent to which the large or macro embossing is in contrast to the base paper/sample, generally inward (macro = larger identifiable image) [shallow → deep]
Micro depth of embossing	The extent to which the small or micro embossing is in contrast to the base paper/sample (micro = uniform non-identifiable pattern; small dots, lines, etc.). [shallow → deep]
Translucency	The degree to which the sample is translucent and not opaque [opaque → translucent]
Gritty	The amount of sharp, prickly particles [smooth → gritty]
Grainy	The amount of small rounded particles [smooth → grainy]
Lumpy	The amount of bumps and irregularities [smooth → lumpy]
Fuzzy	The amount of pile, fibre and/or fuzz on the surface [no pile/bald → high pile/fuzzy]
Slipperiness (hand friction)	The force required to move the hand across the surface [drag/no slip → no drag/high slip]
Overall roughness on face	The overall roughness felt when wiping sample against face. Includes gritty, grainy, dryness, etc. [smooth → rough]
Tensile stretch	The degree to which the sample stretches from its original shape [no stretch → high stretch]
Force to gather	The amount of force required to gather the sample toward palm [low force → high force]
Stiffness	The amount of pointed, ridged or cracked edges, not rounded/pliable, felt in the palm/fingers [low → high]
Compression resilience	The force of the sample felt against the hands [low → high]
Fullness of body	Amount of paper/material felt in the hand [low → high]
Hand moistness	The amount of moistness/wetness felt on the fingers [dry → wet]
Table moistness	The amount of moistness/wetness felt on the table surface once sample is removed [dry → wet]
Pilling	The amount of pilling or small, rounded balls [low → high]
Integrity of napkin (post sticky assault)	The degree to which sample maintains its original integrity [none/destroyed → high integrity]

Source: Nolan et al. (2009).

individual panellist lash variability and potential data adjustment to focus on changes resulting from application and wear by subtracting untreated lash only values.

Figure 9.7 demonstrates measurable differences in lash appearance of one panellist prior to, immediately after and 6 hours after application of two products. Attributes measured include lash length, density, curling, colour; properties of lash groupings; and transference of moist or dried product to the skin surrounding the eye.

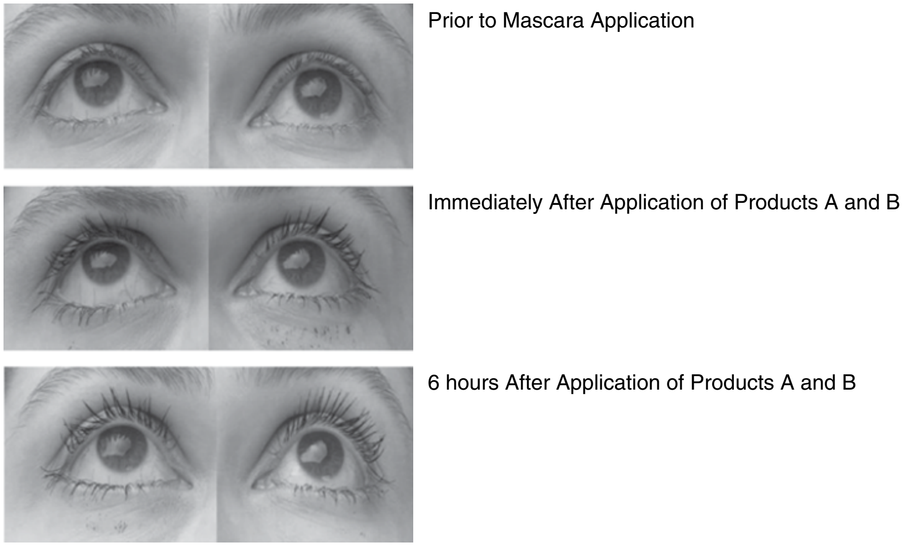


Figure 9.7 Photo of eyes showing mascara evaluation attributes.

9.7 Practical Hints/Tips

Spectrum descriptive analysis panels can provide robust data for many objectives; however, the panel needs to be tended, monitored, calibrated and validated as for any other analytical instrument. As a panel is composed of human judges, part of the tending is motivational as well as related to objective functioning. To this end, the panel leader matters. To develop and maintain a successful Spectrum panel – invest in a good panel leader. A successful Spectrum panel has a panel leader who is skilled at evaluation of the product category of interest (understands the attributes, scales and the protocol), teaching and coaching adult learners and analysing and interpreting panel data.

Once a Spectrum panel is operational, the panel leader should continue to coach the panellists towards evaluation mastery by encouraging them to tease apart small differences in both qualitative aspects (refinement of language) and quantitative aspects (scaling). Providing routine panellist performance feedback between validations is important. Panellists can also be motivated by providing input in the development of new lexicons and evaluation protocols.

References

ASTM (1997) *E1490-92. Standard Practice for Descriptive Skinfeel Analysis of Creams and Lotions*. West Conshohocken: ASTM.

Brandt, M.A., Skinner, E.Z. & Coleman, J.A. (1963) Texture profile method. *Journal of Food Science*, **28**, 404.

Caul, J.F. (1957) The profile method of flavor analysis. *Advances in Food Research*, **7**, 1–40.
Civille, G.V. & Dus, C.A. (1990) Development of terminology to describe the handfeel properties of paper and fabrics. *Journal of Sensory Studies*, **5**, 19–32.
Civille, G.V. & Dus, C.A. (1991) Evaluating tactile properties of skincare products: a descriptive analysis technique. *Cosmetics & Toiletries*, **106**, 83–88.
Civille, G.V. & Lawless, H.T. (1986) The importance of language in describing perceptions. *Journal of Sensory Studies*, **1**, 203–215.
Civille, G.V. & Liska, I.H. (1975) Modifications and applications to foods of the general foods sensory texture profile technique. *Journal of Texture Studies*, **6**, 19–31.
Civille, G.V. & Lyon, B. (1996) *Aroma and Flavor Lexicon for Sensory Evaluation. Terms, Definitions, References, and Examples*. ASTM DS 66. West Conshohocken: ASTM.
Civille, G.V. & Szczesniak, A.S. (1973) Guidelines to training a texture profile panel. *Journal of Texture Studies*, **4**, 204–223.
Heylmun, J. (2003) Creating New Products. 57th PMCA Production Conference, Hershey, PA.
Johnsen, P.B. & Civille, G.V. (1986) A standardized lexicon of meat WOF descriptors. *Journal of Sensory Studies*, **1**, 99.
Johnsen, P.B., Civille, G.V. & Vercellotti, J.R. (1987) A lexicon of pond-raised catfish flavor descriptors. *Journal of Sensory Studies*, **2**, 2.
Johnsen, P.B., Civille, G.V., Vercellotti, J.R., Sanders, T.H. & Dus, C.A. (1988) Development of a lexicon for the description of peanut flavor. *Journal of Sensory Studies*, **3**, 9–17.
Keane, P. (1992) The flavor profile. In: Hootman, R.C. (ed.) *Manual on Descriptive Analysis Testing for Sensory Evaluation*. ASTM Manual 13. Philadelphia: ASTM, Philadelphia, pp. 5–14.
Lee, J. & Chambers, D.H. (2007) A lexicon for flavor descriptive analysis of green tea. *Journal of Sensory Studies*, **22**, 256–272.
Lyon, B.G. (1987) Development of chicken flavor descriptive attribute terms aided by multivariate statistical procedures. *Journal of Sensory Studies*, **2**, 55–67.
Meilgaard, M.C., Civille, G.V. & Carr, B.T. (2007) *Sensory Evaluations Techniques*. Boca Raton: CRC Press.
Muñoz, A.M. & Civille, G.V. (1998) Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, **13**, 57–75.
Nolan, K., Visco, J.V. & Krogmann, A.R. (2009) Creating Customized Descriptive Protocols to Make Consumer Relevant Claims for Household Paper Products. Presentation to the Rose Marie Pangborn Technical Conference, Florence, Italy.
Retiveau, A., Chambers, D.H. & Esteve, E. (2005) Developing a lexicon for the flavor descriptor of French cheeses. *Food Quality & Preference*, **16**, 517–527.
Schwartz, N. (1975) Adaptation of the sensory texture profile method to skin care products. *Journal of Texture Studies*, **6**, 33.
Sensory Spectrum, Inc. (1985) Sensory Spectrum Panel Leadership Workshop. Course Notes. New Providence: Sensory Spectrum.
Sensory Spectrum, Inc. (2010) Sensory Spectrum Flavor Spectrum Descriptive Analysis Workshop. Course Notes. New Providence: Sensory Spectrum.
Sensory Spectrum, Inc. (2011) Sensory Spectrum Look and Feel of Personal Care Products Workshop. Course Notes. New Providence: Sensory Spectrum.
Sensory Spectrum, Inc. (2012) Sensory Spectrum Panel Leadership Workshop. Course Notes. New Providence: Sensory Spectrum.
Skinner, E.Z. (1988) The texture profile method. In: Moskowitz, H.R. (ed.) *Applied Sensory Analysis of Foods*, vol. **1**. Boca Raton: CRC Press, pp. 89–110.
Stapleton, L. & Civille, G.V. (2008) The Challenges of Malodor Evaluations, Qualifying Model Malodors, Evaluating Malodor Counteractants and Measuring Malodor and Fragrance Intensities Objectively. Presentation to the IFT Symposium, New Orleans, LA.

Stoer, N., Rodriguez, M. & Civille, G.V. (2002) New method for recruitment of descriptive analysis panellists. *Journal of Sensory Studies*, **17**, 77–88.

Szczesniak, A.S. (1963) Classification of textural characteristics. *Journal of Food Science*, **28**, 397.

Szczesniak, A.S., Brandt, M.A. & Friedman, H.H. (1963) Development of standard rating scales for mechanical parameters of texture and correlation between the objective and the sensory methods of texture evaluation. *Journal of Food Science*, **28**, 397–403.

Talavera, M. & Chambers, D.H. (2008) Simplified lexicon to describe flavor characteristics of western European cheeses. *Journal of Sensory Studies*, **23**, 468–484.

CHAPTER 10

Quantitative Flavour Profiling

Sophie Davodeau and Christel Adam

10.1 Introduction

Aromas associated with flavours and fragrances offer one of the most direct ways to stimulate human memory and emotion. They activate impulses which are interpreted by the human brain thereby influencing food ingestion, food acceptance and other human behaviours or physiological responses (Weingarten & Gowans 1991). Indeed, everyone has experienced passing a pleasant odour of fresh ground coffee or chocolate and instantly craving the product associated with that smell. Humans also have a great ability to remember events/emotions associated with, and triggered by, odours and tastes, as depicted by Proust (1913).

> The sight of the little madeleine had recalled nothing to my mind before I tasted it … but … as soon as I had recognized the taste of the piece of madeleine soaked in her decoction of lime-blossom which my aunt used to give me…. immediately the old grey house upon the street, where her room was, rose up like a stage set to attach itself to the little pavilion opening on to the garden which had been built out behind it for my parents.

The flavour industry aims to create flavour solutions for food and beverage products that will delight consumers, taking full advantage of their impact on emotions, memory, liking and behaviour. The role of sensory analysis within the flavour industry is to inspire, guide and validate the creation of those flavour solutions. A typical activity would be research into the sensory characteristics that will ensure a high acceptance for a certain flavour type in a specific product (i.e. what are the sensory characteristics most liked by consumers in a vanilla ice cream or even in a specific brand of vanilla ice cream). Flavour creators or flavourists would then utilize this information as a source of inspiration and guidance. The flavourists would use their in-depth knowledge of flavour ingredients, combinations and their performance in applications in order to create the right flavour and the right perception. The success of this process is intimately linked to the ability of the flavour creator to fully understand and interpret the sensory insights (i.e. a sensory description) delivered.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

Flavourists are extremely knowledgeable, precise and sensitive when it comes to flavour perception and this can differentiate them from their counterparts in the food and beverage industry (product developers). Flavour creation is also closer to artistry than food and beverage product development which is often more focused on food technology. The role of flavourists is described as that of an aspiring painter and a musician (Mosciano 2006). For those reasons, flavourists often develop a very personal and intimate way to describe flavours. This description is often a mix between single ingredients, chemicals (e.g. indol, cis-3-hexenol) and broader consumer-like or artist-like vocabulary expressing the general impression of a flavour (fullness, authentic bourbon vanilla, etc.). The role of sensory descriptive analysis is to describe as objectively as possible perception of flavours or products. Within the flavour industry, this role cannot be successful without fully addressing the specificity of the flavourists as users of the sensory insights as well as leveraging their knowledge. Therefore, there was a need for a sensory descriptive method specifically focused to inspire, guide and validate flavour creation which lead to the development of the quantitative flavour profiling (QFP) method at Givaudan-Roure (Stampanoni 1993).

Flavour creation and flavour ingredients development are also processes which take place in the early stage of general product development. Flavour ingredients will be developed or identified in order to become part of the composition of flavours. Flavours will themselves enter into the composition of a product with other product features (brand, packaging, price, etc.). Together, they create a full sensory experience for the final consumer. In these early stages, it is important not to compromise the opportunity to discover great ingredients or create great flavours. An ingredient might very slightly enhance perceived saltiness. However, in combination with a congruent flavour and the right application, significant consumer benefits may be noticed or observed. A small progression in making an orange flavour taste more authentic (i.e. more like the real fruit) will ultimately create a larger consumer response. QFP also addresses this specificity of the flavour industry through comprehensive guidelines that guarantee precision, accuracy and reproducibility. For these reasons, QFP could also be a good general sensory descriptive method for the food and beverage ingredient industries.

In this chapter we refer to flavour as the complex combination of the olfactory, gustatory and trigeminal sensations during tasting. Flavour may therefore be influenced by tactile, thermal, painful and/or kinaesthesic effects (ISO 1992).

10.2 Theory

Quantitative flavour profiling was developed at Givaudan as an enhancement of quantitative descriptive analysis (QDA). The QDA technique uses trained individuals to identify and quantify the sensory properties of a product in order of perception, such as appearance, odour, taste and flavour, etc. The basic steps of this method include the development of a language, the selection and training of sensory subjects, individual data collection of repeated measurements, use of interval scales and analysis of results by univariate and multivariate methods. Straightforward flavour perception visualizations, such as maps or bar charts, facilitate communication between different groups with various levels of expertise, such as flavour creation and application, marketing and sales.

Quantitative flavour profiling deviates from QDA, focusing on the quantification of sensory properties related to flavour. At its beginning, it was mainly focused on flavour (as in retro-nasal odour) and trigeminal perception, and has evolved over the years (since 2009) to also cover taste and mouthfeel perceptions. Today, those aspects play a key role in flavour development in a context where consumers are increasingly looking for products that meet their health and wellness needs without being willing to compromise on 'taste' quality. As with QDA, QFP allows evaluation of more than a single product in a test and can be used for a wide variety of products (e.g. confectionery, dairy, savoury, beverages, etc.). It highlights differences or similarities between samples, determines the perceptual nature of that difference and examines how different sensory attributes correlate with each other.

Quantitative flavour profiling specificity is mainly based on the use of a standardized flavour language, Sense It®, within Givaudan. This language needs to be globally used by all functions (flavourist, application, sensory, consumer understanding, marketing and sales) as well as being integrated into flavour creation and selection in order to ensure universal and aligned communication on flavour profiles. The global language is constantly improved by flavourists in close collaboration with sensory scientists. Each sensory descriptor is provided with a written definition and a physical standard to illustrate the concept of the attribute. These physical references ensure limited influence from a sensory panel leader on the way descriptors are interpreted. This is key to ensure global alignment on flavour evaluation and make sure that we do all speak the same language.

In comparison with classic flavour profiling, therefore, QFP has the key advantage that it facilitates communication between different groups with various level of expertise (from consumers to flavourists). Figure 10.1 illustrates how several people would use different vocabulary and semantic lexicons to describe the same flavour concept whether they are a consumer, a customer (food and beverage manufacturer), a flavourist or a sales person from a flavour company.

In addition, the use of clear definitions and physical references illustrating the concept plays an important role in developing the appropriate terminology and reducing the amount of time necessary to train sensory judges/panellists. It also decreases the judges' variability and allows an exchange of results and comparison of data across time and products. In this way, cultural differences among subjects can be counterbalanced so that sensory panels in different countries can be equally trained.

Figure 10.1 Representation of the various language used to describe the same flavour concept.

Givaudan's trained external QFP panellists are prescreened for their sensory acuity on flavour and taste. They have many years of experience in evaluating products from a wide variety of categories. Today, QFP is used in many different projects such as market studies, flavour optimization and preference mapping to determine drivers of consumer acceptance. The way it is used has evolved over the years in order to adapt to product characteristics which have become more and more complex and to our ways of working in Sensory at Givaudan. This will be described later in this chapter.

10.3 From Traditional to Modern Quantitative Flavour Profiling

10.3.1 The Global Sensory Language: Givaudan Sense It

It is essential to enable effective communication around sensory perception, by reconstructing and harmonizing our language and ways of communicating on flavours. Indeed, to understand each other, a neutral description of sensory perception is required. It removes language subjectivity which relates back to personal context of past experiences, differing perceptions and diverse beliefs. In addition, people are not naturally 'competent' at expressing what they taste and smell as much as they would be in describing visual and tactile aspects of products.

Sense It is the global Givaudan language based on a name (corresponding to an identified flavour direction), a definition (an accurate explanation of what the descriptor's name stands for) and a physical reference (anchor point breaking down the language barrier) illustrating each aroma, taste and mouthfeel characteristic that would be perceived in a product. The use of physical references plays an important role in developing the appropriate flavour terminology on specific projects. References can be single chemical substances, spices/extracts, ingredients or flavours which can be either smelled or tasted. They characterize a specific property of a certain item, and they are made available during all steps of a sensory descriptive study (here QFP) – from language generation to the training and scoring phases. Sometimes, a single reference standard is not enough to illustrate a concept (e.g. mouthfeel language). In this case, a descriptor is illustrated by two references to be evaluated and compared. The difference perceived illustrates the concept.

Descriptors are grouped in flavour families (e.g. orange lexicon) for easy use while performing a QFP test. Today, Sense It contains more than 360 descriptors with reference materials grouped in 140 flavour lexicons. It is constantly evaluated and adapted according to specific needs and product characteristic evolutions on the market. An example of a lexicon is given in Table 10.1. When building a global language or a lexicon, one needs to ensure that the

Table 10.1 Example of orange lexicon.

Descriptor name	Descriptor definition
Aldehydic	Green, fatty aroma associated with citrus oils, fresh chopped coriander leaves or cilantro
Aldehydic-green	Green, waxy aroma associated with green citrus oils
Apricot	Aroma associated with dried light stone fruits like apricots or peaches
Astringent	Drying mouthfeel typically associated with tasting tannin (in water), strong black tea, unripe banana or young red wine
Bitter	Basic taste sensation associated with substances such as caffeine or quinine diluted in water
Blood orange	Fruity, powdery, citrus and berry-like aroma associated with blood orange; often found in blue raspberry and berry-flavoured products, as in Jolly Ranchers, Kool-Aid and Fruit Roll-Ups
Candy-banana	Sweet, fruity, banana-like aroma associated with banana-flavoured candies
Canned fruit	Sweet, fruity, oxidized or metallic aroma associated with canned fruits
Citrus-terpenes	Oxidized citrus, slightly woody and lime-like aroma associated with lime and other citrus fruits
Cooked orange	Aroma associated with oranges that have been subjected to heat, similar to concentrated orange juice
Estery-orange	Fruity, ripe aroma, associated with oranges
Floral-citrus	Woody, rosy and green aroma often found in citrus flavours
Fruity-pineapple	Sweet, fruity aroma associated with ripe pineapple
Grapefruit	Sweet, juicy and citrus aroma associated with grapefruit peel and pulp
Grapefruit-pink	Sweet, juicy, citrus and sulfurous aroma associated with pink grapefruit peel and pulp
Green-fresh	Aromatic associated with green fresh fruit
Green-fruity	Green aroma associated with unripe pears, bananas and apples
Lemon-fresh	Peely, juicy aroma associated with fresh squeezed yellow lemon

language will be exhaustive. This means that the descriptors will enable the perception of the product to be described to its full extent. This completeness of language, coupled with extensive training of the panels, will be crucial to avoid 'dumping bias', which is the effect of restricting the response in profiling as described, for example, by Schifferstein (1996). In this example, the addition of hexenol (fresh-green aroma) to a strawberry flavour caused mean ratings of several sensory attributes to increase unless the attribute 'green' was included in the ballot.

10.3.1.1 Language Development

The objective of language development is to create a core list of unique descriptors for a specific flavour icon. A flavour icon is a flavour which in itself corresponds to a sensory reality (e.g. strawberry flavour) but which is, in practice, broken down into more specific characteristics (e.g. jammy strawberry, wild strawberry, ripe strawberry …).

The language development is usually kicked off by a group of about 6–8 flavourists. In order to profit from the largest expertise, those assessors are selected from different countries and are well recognized for their expertise in the specific flavour icon (e.g. citrus, vanilla, chicken, etc.). The flavourists start from their own experience, providing a first list of key characteristics which are needed to describe the specific flavour icon. Whenever possible, they try to provide a name, a definition and a physical reference associated with each of the needed key flavour characteristics.

In the original QFP method (Stampanoni 1993), flavourists were served with representative sets of coded flavoured products (e.g. different strawberry flavours applied in yoghurts covering specific flavour directions such as green, ripe, cooked, jammy and wild strawberry). Flavourists freely described the flavour of the different samples without hedonic (e.g. good, balanced), generic (e.g. typical, full) or intensifier (e.g. mild, strong) attributes allowed.

Today, as the scope of the projects has evolved from a country-specific to a regional (multiple countries) and global context (multiple regions), the language development only focuses on the flavour icon itself and not on the final application. This allows the development of a standardized language which is flavour relevant, regardless of the country or the application in which the flavour is applied.

Whenever pertinent, external technical experts collaborate with the flavourists in order to provide a complementary view on the topic. For instance, baristas and tea experts participated in the language development of coffee and tea icons by sharing another point of view. They are able to link specific flavour characteristics with product origins, specific product qualities and different steps of the production process. Those characteristics are connected to specific molecule or mixes of molecules which are integrated into the final flavour language and used as references. This external input from fresh minds allows discovery of potentially new references that are not already used by flavourists.

In parallel to the expert panels' exercises, sensory scientists review existing information from all sources of interest: literature review, sensory scientist networks, presentations and posters from symposia. If a language has already been developed from the former ways of working on QFP (Stampanoni 1993), existing Sense It references that have been used during the last 5 years are reviewed and added to the other sources of information.

All sensory descriptors are then compiled into a master core list, containing for each descriptor a name, a definition and a reference standard. The list is then revised by flavourists and sensory scientists in order to eliminate potential duplicates, either combining descriptors with the same name but different references or changing names to keep different references for a same descriptor, complete descriptors for which either a definition or a physical reference is missing and to avoid descriptors which should not be part of the final list (e.g. those that are too specific for particular applications or countries).

The established core list of descriptors is then reviewed during a beta testing by flavourists and application scientists who did not participate in the language development stage to check whether the names, definitions and references are appropriate. They participate in an open discussion moderated by a sensory panel leader who was not actively involved in the language development process. The beta testing objective is to get feedback on the core list of descriptors from different end users (e.g. Is there any gap or redundancy? If a reference is missing, please propose a possible raw material. Are descriptor name, definition and raw materials aligned? Is the reference too strong or too weak? If several references are proposed, please select the most appropriate one for the following descriptor).

In addition, sensory panels undertake a series of exercises focusing on the above aspects. They perform an individual blind evaluation (by smell or by taste, according to the recommended protocol) of references coded with three-digit codes in order to avoid any influence of the name. They describe the reference with their personal comments and share them with the group to define a consensual name. The alignment of the name and its reference is validated by two blind individual recognition tests performed in the sensory booths. The objective is to ensure that the final name corresponds to everybody's perception.

The final validation of the language is conducted using an evaluation of market products, selected in order to cover the full targeted flavour space. It is performed using as much as possible the beta testing reference list and by adding other flavour characteristics if needed (new physical references are then proposed by a flavourist). The individual scoring step of the market products allows the consensual understanding of descriptors to be confirmed. This step ensures that the developed language is adapted to describe the world of flavour for a certain flavour icon (e.g. vanilla, orange, beef, etc.).

Compiling the information from flavourists, application people and the sensory panels allows the adaptation of a very technically oriented language into a more 'consumer' compliant one. After all feedback is collected, the newly proposed standardized language is implemented in the Sense It global language.

All steps described above are an investment into the success of all subsequent QFP conducted using the language created. The proper execution of these steps will ensure that the sensory language is truly global in the way that it can describe the world of flavour perception and be understood by sensory panels, sensory scientists, flavourists and, later on, customers (food and beverage manufacturers).

10.3.1.2 Language Selection on Project

While working on a specific project, the sensory panel leader always refers to the global Sense It language by selecting the flavour lexicon of interest (e.g. orange, chicken, cheese, etc.). The language selection starts with panels of 6–8 flavourists, not directly involved in the project (Stampanoni 1993). The pre-established flavour language is then presented to the sensory panellists so that they can familiarize themselves with the descriptors and in some cases modify or enlarge the descriptor list. This step is to familiarize the panellists with the selected Sense It lexicon(s) by smelling or tasting the references. This process of becoming familiar with an existing lexicon may sometimes be circumvented when the panellists directly generate the first list of Sense It descriptors to characterize the whole product space.

Once they feel confident with the language, they use it to monadically describe each product in the project. During this process, they have the opportunity to add descriptors which either come from other Sense It lexicons or do not exist in the Sense It language at all. The panel leader checks and confirms that the way each descriptor is understood and used is consensual. Flavourists are then asked to validate the descriptor list generated by the sensory panellists. This process ensures that flavourists are still strongly involved in the language generation to facilitate their understanding of the results of the test at the end. They must check and confirm whether the preselected language covers their needs for the project (e.g. specific notes they would work on and would like to have greater detail on). They are also responsible for suggesting possible physical references in case missing descriptors were introduced by the panellists.

This evolution in the methodology has strongly increased the performance of the sensory panellists. Following the original process, they had the tendency to stick to the proposed list without adding descriptors, even though they felt that they would be important for the project. The new process helps familiarization and utilization of the Sense It language in the frame of projects, while maintaining an active flavourist involvement.

10.3.2 QFP Panel Training

The QFP panel training includes two phases. In the first phase, the judges or panellists will be recruited, selected and generally trained to be part of a QFP panel. The judges who have successfully gone through this phase will compose the QFP panel.

This panel then goes through a second phase of being specifically trained for a given project. This last phase is repeated for each project the QFP panel will undertake.

10.3.2.1 QFP Panel Recruitment and Set-Up

For QFP, a panel of between 12 and 16 judges is used. The judges are selected, trained and qualified prior to their participation in the test. QFP panel recruitment includes the following steps.

- Recruitment based on motivation, availability and general ability to perform the job description. This part is typically conducted externally by a recruitment agency.
- Screening tests are used to measure sensory acuity on flavour and taste. Tests are a combination of discrimination and descriptive tests.
- Personal interview: this is used to observe personality and communication skills. It will be crucial for the success of the QFP panel that panel members interact well with the panel leader and with each other. The panel will also be trained on specific and strict languages. Therefore, panellists will need to be open to change and show an aptitude to learn.
- 'Mock' panel: this is to determine whether the potential candidates can work well in a group setting. This is another opportunity to observe communication skills and personality. Potential candidates are presented with products and, with the guidance of a panel leader, express their perceptions. Candidates should be able to detect and describe differences.

The procedure to screen new external panellists consists of three parts.

10.3.2.1.1 Screening Part 1 – Basic Sensitivity Tests

A large number of people are screened on their basic sensitivity skills, such as basic taste recognition, ranking on basic tastes, odour learning and odour identification. A time limit is given for all the tests to check the participants' ability to make fast decisions. The table below lists examples of samples presented for the basic taste recognition exercise.

Ingredient	Concentration
Sucrose	4 g/kg
Sucrose	4.8 g/kg
NaCl	0.6 g/kg
NaCl	0.72 g/kg
Citric acid	0.3 g/kg
Citric acid	0.36 g/kg
Caffeine	0.06 g/kg
Caffeine	0.09 g/kg
MSG	0.3 g/kg
MSG	0.36 g/kg

10.3.2.1.2 Screening Part 2 – Discrimination Tests

Difference tests (triangle tests, warm-up paired comparison) are used to determine whether or not the participants can detect small differences in sensory character. The test can be carried out on products likely to be encountered in sensory testing to assess participants' ability to detect differences in these products. Many of the products assessed by sensory panels are tested outside their normal eating context. For example, margarine, tomato sauce, mayonnaise, soy sauce or olive oil may be tested on their own without a host food such as bread, pasta or meat. It is important that the future panellist is ready to evaluate such products.

10.3.2.1.3 Screening Part 3 – Descriptive Tests

This step is used to see whether participants are capable of describing several different products in a non-hedonic way. At least two different samples are presented to the panellists. In the first step, they are asked to describe the two products, using their own vocabulary. In the second step, Sense It references are presented and panellists are asked to familiarize with them as much as possible.

According to the results of the screening phase, a contract with a probation period is signed.

After the probation period, the new panellists form a new panel or are incorporated into the existing panel. During the probation period, the new panellists are trained on Sense It as well as on the use of a 100 mm unstructured scale to quantitatively evaluate flavour characteristics. The advantages of using line scales have been explained in extensive detail by Stone and Sidel (1993). We aim at achieving very good panellist performance about 1 year after they have been recruited.

10.3.2.2 QFP Panel Training for Specific Projects

The QFP training is usually split into two parts: the language generation (see section 10.3.1.2) and training on the quantitative evaluation of flavour characteristics. Depending on the set of products as well as on the complexity of the project, specific exercises for difficult products and/or attributes are conducted during the training in order to check and maintain a consistent panellist performance. At any time during the training, the panel leader has to ensure that the protocol for product evaluation is put in place and that products are evaluated the same way by all the panellists.

Language generation is always the first step in a project. The details about this phase are described in section 10.3.1.2. The selected project-specific Sense It descriptors/references are always available during the training. They can be resmelled or retasted by panellists whenever needed.

In order to familiarize the panellists with the selected Sense It references, the panel leader can use different techniques. The most common and simple exercise is to conduct several individual sniffing/tasting sessions, where panellists note their personal associations. These sessions are followed by open group discussions in order to share the individual notes and to make sure that the descriptors are understood and used the same way by all panellists.

Once the panellists are more familiar with the references, it is a good exercise to organize quiz sessions, where panellists have to smell the references blind – either individually in the group or in pairs – and link them back to the correct descriptor name.

Once the panel leader sees that the panel is comfortable with the language, the second step is to incorporate intensity measurements. In this step, panellists are asked to score the intensity of each attribute for each product. The aim is to familiarize the panellists with the set of products and to monitor a consistent and consensual performance of all panellists. It is the panel leader's decision whether the training needs to be continued or not.

The training session is moderated by the panel leader. All panellists are invited to participate actively in the discussion. The individual results from all panellists are collected by the panel leader and discussed in the group, in order to come to a consensual agreement.

To score the intensity of a specific attribute for a product, different scales can be used. The appropriate scale should be selected depending on the training level of the panel. For scale training, five indications such as low, low-medium, medium, medium-high, high are recommended (Figure 10.2). Other non-technical words can be used if they better fit the local language. For untrained panels, it is also recommended that numerical scales from 0–100 are used during the training in addition to the five indications (see Figure 10.2).

If the panel is well trained and experienced in the use of the line scale, unstructured line scales can be used during training. In special cases, anchor products can be used to illustrate the intensity level of the attributes. In this instance, the panel leader selects some products that represent a specific flavour

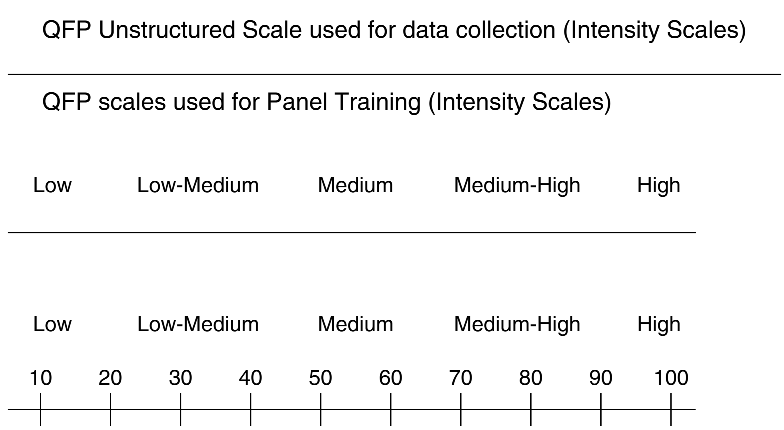


Figure 10.2 Visual analogue scales used in QFP.

direction/intensity. The panel is calibrated on the intensity perception using those anchor products. The calibration can be done on all attributes or just on a selection of key attributes.

As it is important to keep the motivation and performance of the panel at a very high level during the training session, it is recommended to structure the training, starting with simple tasks and slowly increasing the difficulty level. If possible, samples with big differences should be presented first to the panellists; afterwards one can progress to samples with small differences.

In general, it is recommended to focus on only difficult or important attributes during the training. This helps the panellists to stay focused and not to get lost in unnecessary small details. The panel leader has to make sure that all panellists are participating in the discussions. A good atmosphere is key for successful group discussions. All feedback should be considered and panellists should be motivated to express themselves at any time.

At the end of a successful training, the panellists are aligned in the understanding and use of the references as well as the intensity of perception. Their performance is checked by the panel leader and the training is only stopped once panel performance has reached the expected confidence level.

The training conducted before each project aims at maximizing the ability of the panellists to describe samples accurately while minimizing two key sensory judgement biases.

- *Relative scaling*: one of the bases of perceptual psychology is that individuals are better at comparing perceptions than they are at rating them. Indeed, individuals might be very poor at identifying the concentration of salt in a product but will be very good at identifying which product is saltier than another. By using anchored scales during training and extensive exposure to the products, we aim at limiting this effect.
- *Contrast effect*: the contrast effect means that any stimulus will be seen as stronger in the presence of a weaker stimulus (Lawless 1983). By introducing the full sensory spectrum of the products for each project, we aim at minimizing this effect. The panellists will be familiar with the type of differences expected and will concentrate more on describing the product as opposed to comparing with previous evaluated products.

10.3.2.3 QFP Panel Maintenance

With each project, the panel leader continuously checks for panel performance. This can be as simple as observing the training sessions and/or analysing practice data. With each observation or analysis, it is noted whether the panellists understand the attributes and scale, and whether there are some panellists not aligned with the group. This is usually done through the process of repeated tests. Panel leaders can check to determine whether there are significant variations between replications. Feedback sessions are important to the panel, allowing them to see their progress and keeping them motivated to maintain a high level of performance where it is already the case and progress in the areas where development is required.Usually in feedback sessions, there may be comparisons between a panellist's responses compared to that of the entire group and/or a panellist's repeated response. The information in the feedback sessions is usually a mixture of verbal and written communication with the use of graphs and charts. Each feedback session consists of giving positive comments as well as constructive criticism.

10.3.3 Data Collection and Intensity Measurement

10.3.3.1 General Process

Data collection is an individual process during which each judge is presented with the entire product range of the study in a sequential monadic order. All judges will evaluate all the products at least twice. On some occasions, the panel leader might decide to increase this number. The presentation order of the samples is balanced and randomized across judges and repetitions, to minimize order effect and carry-over effect.

One product at a time, the judge is asked to rate the intensity for each descriptor on the list. The intensity rating is given using a continuous visual analogue scale as shown in Figure 10.2, which is an unstructured line scale.

10.3.3.2 Sensory Testing Environment

The data collection process is carried out individually by each judge in a sensory booth in order to eliminate mutual suggestion bias. During this process, there will also be no feedback to the judges from the panel leader or the staff in charge of product serving. The sensory booths are also disconnected from the preparation area so that no distractions or cues can be perceived by the panellist.

Sensory booths are individual booths where panellists can evaluate samples quietly without any interruptions. They are made of specific white material according to AFNOR norms, avoiding reflection of lights and aiming at giving the best context for product evaluation with a minimum of evaluation conditions bias. They are cleaned before each session and are free of odour. Different lights (white, green or red) can be used during product evaluation to ensure that panellists are not too much influenced by the appearance and really focus on the aroma and taste. The panellist communicates with the staff using a light system (different lights to indicate that the evaluation is finished or that the evaluation is ongoing).

10.3.3.3 Product and Sample Preparation

To help remove any biases, samples are usually served in colourless, odourless containers and sample sizes are consistent. Between each evaluation, the panellist waits 5 minutes during which he/she is provided with rinsing material (mostly water and unsalted crackers). This process aims at making the evaluation as monadic as possible and minimizes contrast effect (see section 10.3.2.2). It also minimizes adaptation effect. The physiological adaptation effect, well known for visual (dark/light) or temperature perception (hot/cold), is also experienced in flavour perception, as described by O'Mahony (1986).

Consistency of product preparation and serving is most important. The products as presented to the panellists can only vary in the factors that are being evaluated (taste/flavour/mouthfeel). Any colour differences are masked using red or green lights depending on the general colour of the samples. Products are given in the same quantity (weight/volume) and shape (i.e. full crisps versus crushed crisps) and at the same temperature for each panellist and all products included in a study. Maintaining the temperature of products during evaluation might be necessary to allow evaluation over a period of time (i.e. ice cream, hot soup). Adequate containers are then required.

10.3.3.4 Number of Products, Number of Data Collection Sessions

Data collection can be conducted in one single tasting session or over several sessions, depending on the number of products. The maximum number of products tested in each session is eight. However, this is dependent on the samples (e.g. fewer samples per session for spicy foods or foods with a cooling effect). Data are collected using sensory software (Compusense™ or FIZZ™).

10.4 Statistical Analysis

There are three steps in analysis of QFP data.

- Determine if there are significant profile differences between the products.
- Determine which products are different and in what respect.
- Overall displays of product differences.

Analysis of variance (ANOVA) (Hinkelmann & Kempthorne 1994) is the method of choice to determine if there are significant profile differences between products. The ANOVA model contains six factors as shown in Table 10.2.

- Three main effects: panellists, products and repetitions.
- Three second-level interactions effects.

All the effects are tested at a 5% level of significance. The presence of any panellist-related effects is a reason to consider further training of the panel. Training for scale use is indicated by panellist effects. Product*panellist effects can indicate problems with scale usage, but also with understanding of the descriptors. The panellist*repetition effect is a general indication that training is needed.

ANOVA is the basis for the next steps. The product means can be extracted for further analysis.

Tukey's honest significant difference (HSD) (Hsu 1996) is used to calculate which product pairs are significantly different for a studied attribute, again tested at a 5% level of significance. Calculating significant differences is important in interpretation of QFP data. It allows clear statements regarding product differences.

The final step is visualization of the product means obtained from QFP. The most basic method is the profile plot (see Figure 10.4) which shows which

Table 10.2 Effects and interactions included in ANOVA model.

Effect	Interpretation
Product	Objective of study is to measure product differences
Panellist	Main occurring noise factor: big panellist effects indicate that panellists use different parts of the scale
Repetitions	Noise disturbing factor. Difference between repetitions. This may indicate panel drift or that products change, especially if it happens similarly over several descriptors
Product*panellist	Error factor. Indicates that panellists judge products differently, e.g. different product orders or differences between products
Product*repetitions	Error factor. This effect can indicate that some products change between sessions, especially if it happens over several sessions. It can also indicate trouble with scale usage of the panel
Panellist*repetition	Error factor. This effect shows that some of the panellists can have drift between sessions
Error	This should reflect the experimental error. A large error can indicate some important controlled factors should be included in the model or the experiment suffers from a lack of sufficient control

descriptors are strongest, where products are different and the size of differences between products. It can be annotated with exact product differences obtained from Tukey's HSD, though that can make plots crowded. The biggest disadvantage of the profile plot is that it does not show the overall relation between products and descriptors. To show this overall relation, it is better to rely on multivariate methods such as principal component analysis (PCA) and hierarchical cluster analysis (Krzanowski 1988).

Principal component analysis (Figure 10.3) is the best method to display the relation between descriptors and products. It is a two-dimensional plot where both descriptors and products are displayed together. To obtain this plot, the true data have to be distorted. From the original larger multidimensional set of descriptors, reality has to be compacted to two or three dimensions to be readable. There are a number of variations in how the PCA is calculated and these have influences on the distortions in the plot. First of all, descriptors must be centred on the average. It is also common to scale the descriptors. In the case of QFP data, all descriptors are on a common 'intensity' scale so they should not be scaled unless the focus is on highlighting any difference, whatever the amplitude. Scaling the descriptors consists in dividing the centred averages by the standard deviation observed for the descriptor. In turn, this means that the distances between the products are not respected. However, these distances can be examined using cluster analysis.

Hierarchical cluster analysis is an easy way to depict which products are similar and dissimilar. To build a clustering, there are two steps: first, determine the distances between the products; second, determine the actual clustering. Euclidian

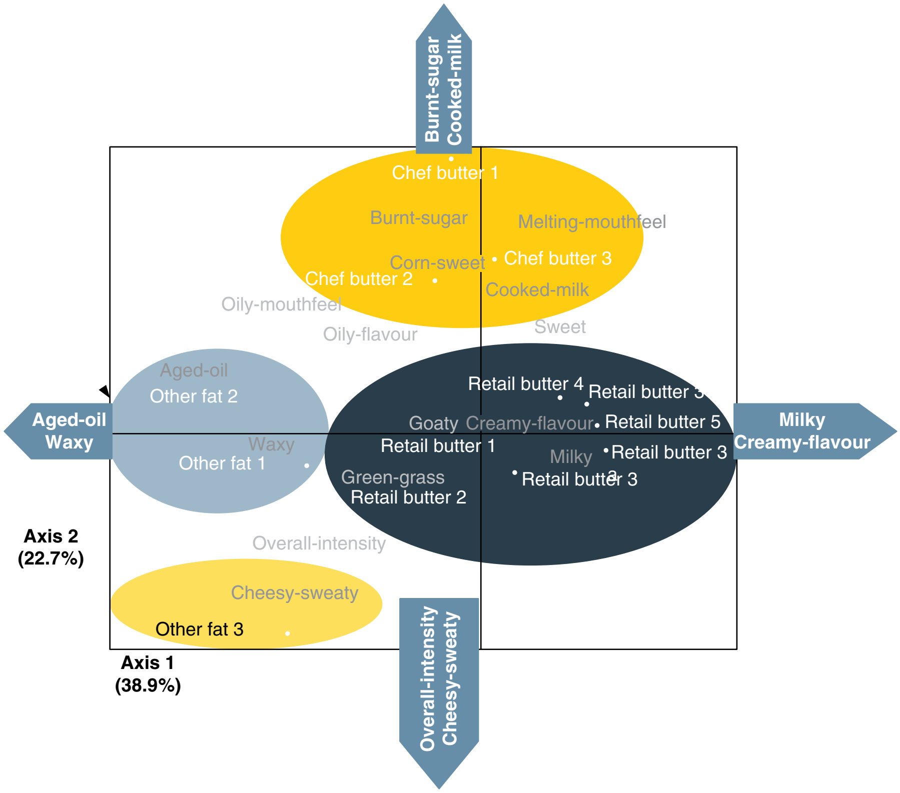


Figure 10.3 PCA on butter and other fat analysed using QFP (PC1, PC2).

distance is used to calculate distances between the products. The clustering method of choice is Ward's method (Ward 1963) on scaled or unscaled averages, as for the PCA. This method results in compact clusters.

10.5 Applications of QFP

Quantitative flavour profiling is designed to provide sensory descriptive information, having four very important characteristics although these do not necessarily appear together in all sensory studies.

- Universality of meaning due to the use of a global language with physical references.
- Applicability to flavour creation using references developed by and for flavour creators.
- Robustness of the description due to extensive training and repetition.
- Precision of the description due to the use of highly experienced panels and specific project training.

However, these qualities mean that QFP also has limitations. It is not a fast way to complete sensory descriptive analysis. Also, it can at times be too precise and comprehensive to truly represent what consumers perceive. For these reasons, the use of QFP is recommended when most of the four characteristics are required for the outcome to be the most appropriate. It is particularly recommended when the outcome is meant as guidance for flavour creation or product development.

As an example, QFP can be utilized when one is optimizing a flavour profile to match a certain existing product, to maximize consumer preference or to maximize certain consumer-driven characteristics (e.g. perceived freshness, association with attributes such as moods, emotions, fit to a concept, etc.). In the case of maximizing consumer preference during preference mapping exercises, for example, precision is very important as one does not know *a priori* the sensory characteristics that might generate consumer liking. In the case of a flavour matching exercise, precision is also extremely important given that the information will need to be translated by the flavour creator in terms of appropriate usage of ingredients to achieve the targeted perception. In both cases, the outcome of the test is guidance for creation and the universality of language as well as its direct transcription into flavour ingredients is key.

Quantitative flavour profiling will also be particularly useful in describing small sensory differences. In the case of wanting to understand the impact of certain molecules or ingredients, it is absolutely necessary to use a sensitive method such as QFP. Indeed, each ingredient might have a very small sensory impact but when added to other ingredients or conditions might end up being very important for the overall perception. Being unable to measure and appropriately describe those small sensory contributions has a negative impact on the ability to innovate and create full flavours and solutions.

In contrast, QFP is not recommended when trying to relate the sensory description with the general perception of consumers that might have been collected during qualitative testing such as focus groups. In these exercises, having a description that is too precise might make interpretation difficult and the outcome is more often used as inspiration for free flavour creation as opposed to specific targeted descriptors. For various studies, one has to compromise between the need for precise and reliable quantitative data and the time and effort necessary to obtain them. For example, when one needs to have only a broad understanding of a sensory domain such as grouping products or flavours with similar characteristics, methodologies like sorting or flash profiling might be more efficient. Lastly, when the number of products is very large (more than 15), such as in studies for category appraisal, market mapping or flavour collection mapping, the use of QFP is time-consuming, like other precise descriptive techniques (e.g. QDA).

10.6 Practical Considerations

Quantitative flavour profiling quality is based on having an expert QFP panel composed of true experts in describing perceptions and using a complex and extensive language. This requires the panel to work for a long time (several years) together in order to gain the necessary experience. The maintenance of a QFP panel therefore requires some compensation to retain panellists.

The use of a global language goes somewhat against nature. Indeed, individuals naturally develop their own way of describing sensory properties. This is even more the case in the food and flavour industries within which professionals come with their own language acquired during training, studies, previous work experience and personal interests. A global language not only requires a lot of effort to build and maintain but also to be understood by all via training and repeated exposure. To make the transition between personal language and global language, panellists are allowed to use their own descriptions or make their own definitions which may help to better understand the references. However, references must remain unchanged as they are the key to the translation of QFP information into flavour and product development.

Panel motivation is a very important factor. Panellists must feel their work is a true contributor to Givaudan business. To ensure panel motivation, feedback sessions are held throughout a project. At the end of the project, either the panel leader or one of the members from the project team will come and present some results. In addition, the panellists are highly integrated into Givaudan activities with visits to the factory, presentations by flavourists, etc.

10.7 Case Studies

10.7.1 Case Study 1: QFP to Guide the Creation of Real Butter Flavours for Cookies

Reproducing the taste of 'home-made', 'made with real ingredients' and 'authentic' is one of the goals of food and beverage manufacturers. However cost, storage, use of certain processes and ingredients often make it difficult to industrially reproduce the taste of a dish prepared at home or 'freshly' prepared just before serving. This research for authentic taste is often motivated by consumers who invariably identify this as a desired characteristic in food products. Using flavours in the development of products which taste as close as possible to a similar 'freshly made' or 'home-made' product is often a great help.

This case study focuses on the development of cookies for the retail market that taste like home-made. More specifically, it focuses on how butter flavours can help the creation of cookies with sensory characteristics as close as possible to cookies made with real butter. In this example, 'real butter' refers to the butter typically used at home or by chefs in restaurants while 'other fat' (OF) refers to the shortening fat used by some cookie manufacturers. The study consists of several steps in which QFP gives key insights ensuring the success of the initiative.

- Step 1: sensory analysis of 'real butter'.
- Step 2: creation of flavours to compensate flavour differences between 'real butter' and 'OF'.
- Step 3: validation of the ability of a new butter flavour to help in the formulation of a cookie that tastes as if it is made with 'real butter'.

Prior to this study, a butter language including 16 descriptors was created by flavourists. The QFP panel was trained on the language until they could identify each of the references. Three out of the 16 descriptors are given in Table 10.3.

10.7.1.1 Descriptive Sensory Analysis of 'Real Butter'

Quantitative flavour profiling is used to describe the sensory properties of 10 real butters (seven retail butters, three chefs butters) and three OFs. The product codes are given in Table 10.4.

Quantitative flavour profiling output consisted of intensity scores (0–100) for each product on each attribute and each panellist. The QFP evaluation included two repetitions of serial monadic evaluation by 12 trained panellists. Average scores per product and per descriptor are calculated and analysed

Table 10.3 Examples of three butter descriptors and their references.

Descriptor	Definition	Reference standard	Reference dosage
Green-grass	Fresh, green, unripe fruit aroma associated with fresh cut grass	Hexenol-3-cis	0.5% in PG
Waxy	Aroma associated with candle wax	Stearic acid	Pure
Burnt-sugar	Sugary aroma associated with brown sugar, burnt sugar, caramel and maple	Homofuronol	20% in PG

Table 10.4 Product codes and types of fat used in QFP.

Real butter		Other fat
Retail butter	Chef's butter	
Retail 1	Chef 1	Other fat 1
Retail 2	Chef 2	Other fat 2
Retail 3	Chef 3	Other fat 3
Retail 4		
Retail 5		
Retail 6		
Retail 7		

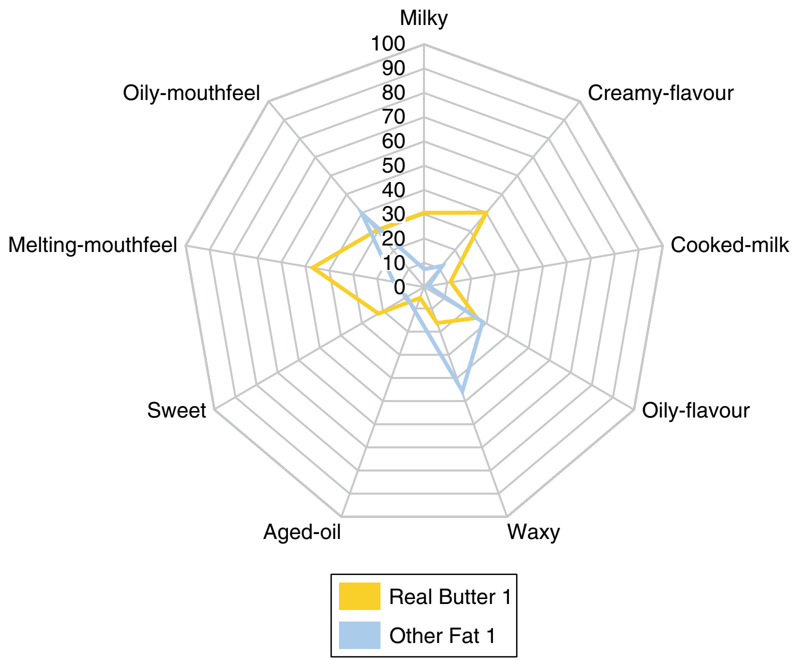


Figure 10.4 Quantitative flavour profiles of two samples tested.

using PCA. This results in a plan where each product is located according to its sensory quantitative flavour profile as shown in Figure 10.3. ANOVA and multiple comparison tests were also run to define the significance of the differences between products on each descriptor.

This representation of the QFP data clearly shows the main differences in sensory profiles between the three types of butter tested. Dimension 1 differentiates between retail butter (more creamy and more milky flavour) and OFs (more waxy, aged-oil flavours). Dimension 2 shows the differences between the chef's butter (more burnt-sugar, cooked-milk flavours) and the rest of the products (stronger flavour overall), especially Other fat 2 (more cheesy-sweaty). However, the PCA plan suffers from one main drawback: information can be explained on other principal components. In order to have a deep understanding of all the differences that exist within the product set, their significance is determined with a Tukey HSD test following a three-way ANOVA.

Figure 10.4 shows the difference between 'Retail 1' and 'Other fat 1'.

10.7.1.2 Creation of Butter Flavours to Improve Perceived Authenticity

This QFP outcome provides clear guidance for flavour creators but also product developers in terms of the flavour notes provided by 'real butter'. It also elucidates flavour notes that are missing in typical ingredients used to replace butter and therefore might end up missing in the finished cookie product.

Using this information, 23 butter flavours were created including:

- five 'butter match' flavours. The objective of these flavours was to reproduce exactly the notes found in five of the real retail butters
- six 'beyond butter' flavours. These flavours included the notes found in real butter as well as notes typically added to butter cookies and known as potential drivers of consumer acceptance. These included notes such as almond and vanilla
- 12 'butter hints' flavours. These flavours were created to accentuate certain characteristics of the real butter but were not exact matches. Those 'hints' were, for example, lactony and caramel.

These 23 flavours were integrated in the recipe of the OF-based cookie and evaluated by a group of experts (flavourists, food scientists, sensory scientists) in order to select the best cookies which mimicked the cookies made with real butter. Five out of the 23 flavours were selected as achieving the goal of enabling cookies made with OF taste as close as possible to cookies made with real butter. The QFP information provided to the flavourists served here as inspiration and guidance.

10.7.1.3 Validation of Authentic Butter Flavours

The five flavours inspired by the QFP information collected on real butter comparison to OF were then analysed into a second QFP. The objective of this QFP was to validate that the newly created butter flavours did indeed deliver by enabling a cookie made with OF to taste as if made with real butter, such as one might use at home or as chefs use to produce cookies. The overall objective is to demonstrate that those flavours can help manufacturers create cookies which taste 'home-made', 'freshly made' and authentic.

For this study, we compared the cookies made with OF and the five new flavours with cookies made with real butter and cookies made from OF and without flavours. The real butter cookies were either marketed as 'made with real butter' or cookies baked internally with real butter (retail or chef butter). OF cookies are either market cookies baked with OF or cookies baked internally with OF. Details of samples are given in Table 10.5.

Table 10.5 Cookies and product codes used in QFP.

Real butter (RB)			**Other fat (OF)**		
Market cookies made with real butter	Cookies baked internally with retail butter	Cookies baked internally with chef's butter	Market cookies made with OF	Cookies baked with OF only	Cookies baked with OF and with added butter flavour
RB market 1	RB retail 1	RB chef 1	OF market 1	OF 1	OF flav 1
RB market 2			OF market 2	OF 2	OF flav 2
RB market 3			OF market 3		OF flav 3
RB market 4					OF flav 4
					OF flav 5

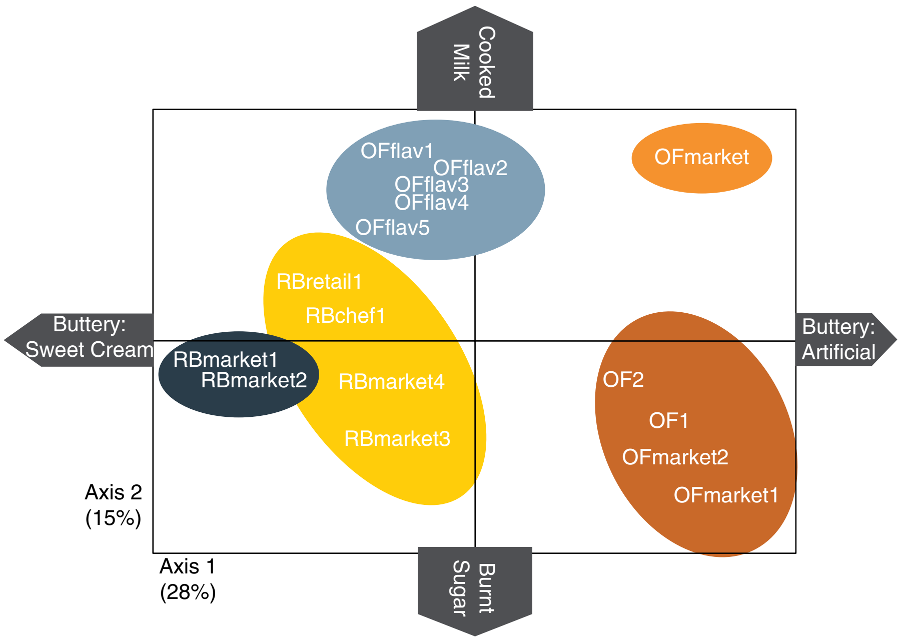


Figure 10.5 PCA on cookies made with butter and other fat analysed using QFP (PC1, PC2).

The QFP consisted of two repetitions of sequential monadic evaluations by 12 trained panellists following the method described in section 10.3. QFP outcomes are the average scores for each product on each descriptor evaluated.

Each product then has an average quantitative flavour profile. The sensory space specificities are then represented using PCA. Figure 10.5 is the graphical representation of the co-ordinates of each product on the two most important principal components (PC1, PC2). The two main dimensions representing the differences in flavour profile can be described as follows.

- 'Buttery' Dimension (PC1): Sweet Cream to Artificial Butter. On this dimension, the cookies made with OF (market or internally baked) and without added flavours are clearly differentiated from the cookies made with real butter (market or internally baked). The cookies made with OF with flavours are clearly closer on that dimension to the 'real butter' cookies.
- 'Brown' Dimension (PC2): Cooked Milk to Burnt Sugar. On this dimension, it looks like the OF products have higher cooked milk notes than the 'real butter' cookies and have a bit less of the burnt-sugar characteristic.

Overall, the QFP clearly helps to validate the successful attempt at creating butter flavour that can be incorporated to non-butter fat while increasing sensory characteristics of real butter. Those butter flavours can be used in the production of cookies in order to increase the perception associated with 'home-made' or 'freshly made'. QFP was used in this example as a great guidance tool for the creation of flavours as well as a validation to demonstrate efficacy of the flavours.

10.7.2 Case Study 2: QFP to Optimize Liking for a Vanilla Flavour

In this example, the aim was to identify flavour characteristics that drive preference for a vanilla ice cream. One might think one vanilla is much like another vanilla. However, a vanilla flavour can go in various sensory directions, from more caramelic notes to more phenolic notes. The objective was to identify which of these directions was particularly appreciated by consumers in specific countries. The final aim was to improve current flavours and create new ones that would deliver optimal consumer preference.

The research is conducted using the external preference mapping method. The sensory descriptive profiles of a series of products are measured. Consumer research is conducted to measure the liking for those same products. A model is then created in order to:

- identify which sensory descriptors are driving preferences for the products
- estimate optimum sensory profiles.

This information would be used as guidance for the creation of a product which would maximize and potentially outperform any of the products tested.

The QFP methodology is used in this example as the precision of intensity scores is important for building a statistical model. The descriptors are also extremely important as the developers (in our case, flavour creators) will have to use this information to optimize the flavour profile. The developers have therefore to understand perfectly the sensation attached to the sensory results obtained. The Sense It references help this translation from a theoretical sensory profile into a real outperforming product.

10.7.2.1 Vanilla Language

The vanilla language consists of 30 descriptors describing the complete sensory characteristics of the products. Texture did not vary as the ice cream base remained constant. Only the taste and flavour could change due to the use of various vanilla flavours in the 20 products.

Each sensory descriptor has a name, a definition and a physical reference designed by flavourists. The sensory panel was trained on the language for two sessions only. Indeed, in this example, the panel had already been trained for some time on the use of the language.

Examples of three of these descriptors are given in Table 10.6.

10.7.2.2 Twenty Vanilla Ice Cream Flavours Tested

The 20 samples represent:

- the sensory dimensions observed in the markets. To do so, preliminary sensory descriptive profiling was conducted on vanilla ice cream from the market studied here. This information gave guidance to flavourists to create formulas replicating accurately the flavour directions observed

Table 10.6 Example of three Sense It vanilla descriptors.

Global name	Reference standard	Concentration (%)	Definition
Anise	Anethol + anisaldehyde 1:1	1% triacetin	Sweet, spicy, licorice-like aroma associated with pastis or anise milk
Dried fruit	Davana oil	1% triacetin	Aroma associated with dried dark fruits like prunes, raisins and figs
Coconut	Aldehyde C-18	2% in TA	Aroma associated with fresh coconut flesh

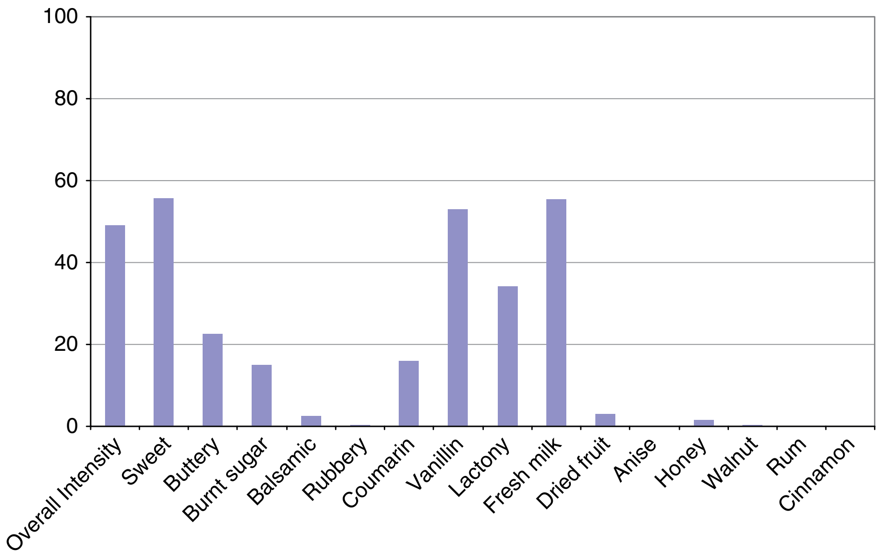


Figure 10.6 Representation of sensory profiles for a vanilla ice cream.

- the sensory dimensions thought to be relevant for the given market using internal knowledge even though those directions might not yet be present in market samples. Those might be sensory dimensions successful in other markets
- free creativity from flavourists in creating new and innovative flavour directions for vanilla ice cream.

10.7.2.3 Sensory Profiles of the Ice Cream

Quantitative flavour profiles of the 20 ice creams are collected using the language and following the QFP methodology. Each quantitative flavour profile can be represented as in Figure 10.6 for which scores of average intensities are represented as bars for each descriptor.

The sensory profiles can also be compared represented in a sensory ‘map’ using PCA. Figure 10.7 represents the main sensory differences measured between the 20 samples in a two-dimensional graphical representation.

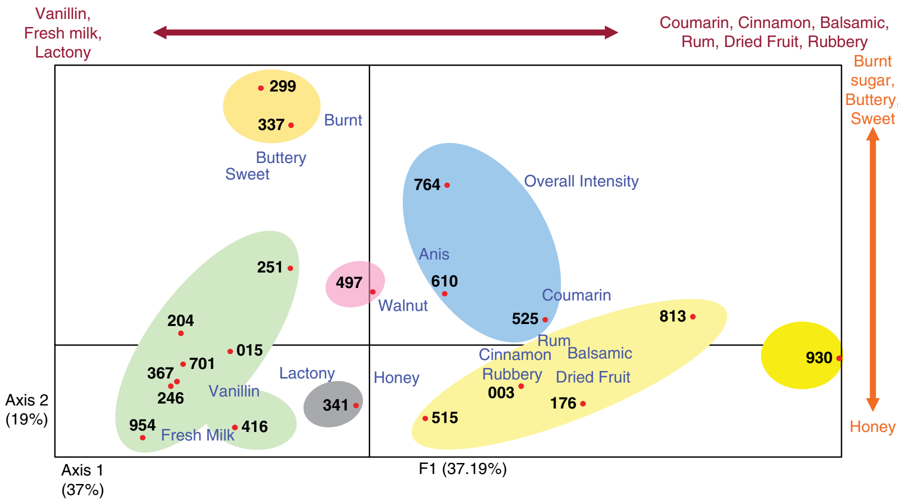


Figure 10.7 PCA map on the 20 vanilla ice cream samples.

10.7.2.4 Identification of 12 Samples for Preference Mapping

The number of samples to be tasted by consumers should be small enough to guarantee the reliability of the collected data. Hence, it is important that a QFP was conducted on those 20 products to identify with a good level of confidence the subset of products that can be representative of the sensory space. To choose this optimal subset, the principle of D-optimal designs is very relevant. This consists of selecting the products using the D-optimality criterion on the co-ordinates of the 20 individuals on the three first axes of the PCA plan. In other words, it consists of selecting the products with the most distant (and therefore different) sensory profiles. This approach was proven to be efficient in minimizing prediction error by the quadratic model of consumer preferences (Ben Slama et al. 1998).

Good coverage of the language is key in that process to ensure that:

- certain perceptions which might be key to consumer acceptance are not overlooked. This would lead to failing to select a sample for the consumer test that might be particularly liked and failing to identify that sensation as driving consumer liking
- the entire sensory profile of the samples is captured. Failing to do so would lead to a poor preference mapping model, as perceptions present in the samples but not measured might be the reason for liking or disliking certain samples.

In our example, the 12 samples were chosen as shown in Table 10.7.

10.7.2.5 Consumer Testing and Liking Scores

A central location test was performed in the country of interest (France) with 120 consumers who rated their liking of the 12 products on a nine-point hedonic scale. The scale used enabled the consumer to express her/his liking by rating

Table 10.7 Sensory characteristics of the 12 samples chosen.

Product codes	Key sensory cluster	Other characteristics
204	Lactony, vanillin, fresh milk	Coffee-like notes
954		
701		Very sweet
299	Overall intensity, buttery, sweet, burnt sugar	Caramel, sugary, furaneol
337		Buttery, coffee
497	Walnut	
341	Honey	
525	Overall intensity, coumarin, anise	Fruity, banana
764		Anise
813	Balsamic, rubbery, coumarin	
176		Strong beany notes
930	Balsamic, dried fruit, rum, cinnamon, coumarin	

each sample from 'Dislike extremely' (score 1) to 'Like extremely' (score 9). The samples were served sequentially and monadically, in the same way as was done for the QFP.

To run preference mapping analyses, it is strongly advised to first standardize the hedonic data. Indeed, consumers do not use the scale the same way.

- Some consumers use a certain area in the scale and others use other areas. It is then important to centre the scores collected around the same value, i.e. 0.
- Some consumers use just a part of the scale whereas others have no problem using the entire scale. It is then important to scale the scores collected to have comparable ranges.

This standardization generates values that can be compared. The resulting data should be considered as preference data.

However, it is crucial to remember that standardization deletes an important part of the information contained in the data, that is, the level of liking. This is an important aspect to keep in mind to avoid misinterpretation of the preference mapping results, for example, estimate optimal profile when none of the products got a score above 5 ('Neither like nor dislike'). Before running preference mapping analysis, it is also very important to check, using ANOVA for instance, that the products have significantly different hedonic scores.

Finally, it is advised to create segmentation of consumers using the standardized liking data. This provides for each segment a list of consumers who share similar preferences toward the product set that was evaluated. In addition to giving more reliable preference mapping models, this enables creation to be guided toward attractive profiles more appreciated than the optimal profile obtained considering all the consumers together. Indeed, a product that can satisfy all consumers can only be a compromise which will not get high liking rates by any consumer. The example described corresponds to one segment identified with the standardized data. The average liking scores for each product given by this specific segment of French consumers are summarized in Figure 10.8.

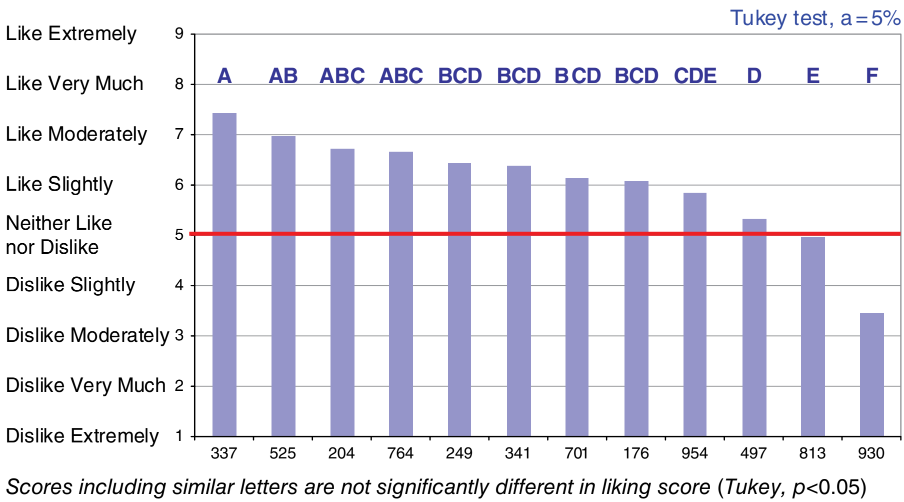


Figure 10.8 Average liking scores (1–9) for French segment 1.

10.7.2.6 Use of QFP Profiles for Preference Mapping

There are two types of preference mapping.

- *External preference mapping*: creation of models to link consumers' preferences to co-ordinates on a sensory map obtained using sensory profiles of the products.
- *Internal preference mapping*: mapping of the hedonic data. Sensory characteristics of each product can be projected on the resulting map for the experimenter to highlight correlations between the preferences and these sensory directions.

In this example, external preference mapping is used. QFP data are averaged and standardized to make sure each flavour direction is weighted the same for the creation of the map with a simple PCA. When dealing with flavours, this aspect is very important. For instance, a honey note in a vanilla flavour can have a strong impact on consumer preferences while its intensity in the overall profile is quite low.

The co-ordinates of each product on the first plane of the PCA are then used to create models. The first plane is commonly used but others might be considered according to the data.

The external preference mapping techniques as described by Danzart et al. (2004) consist then of modelling each consumer's preferences scores (standardized liking data) with the co-ordinates of the products on the map. The models obtained result in surfaces that are compiled to provide a graphical representation as in Figure 10.9. When projecting the colour code of this resulting surface on the PCA map, one obtains Figure 10.10.

From this map representing both sensory profile and consumer preferences, some areas of interest for the development of attractive products can be identified. It becomes interesting to estimate what would be the sensory profile of these optimal products. Danzart et al. (2004) described a simple approach to obtain

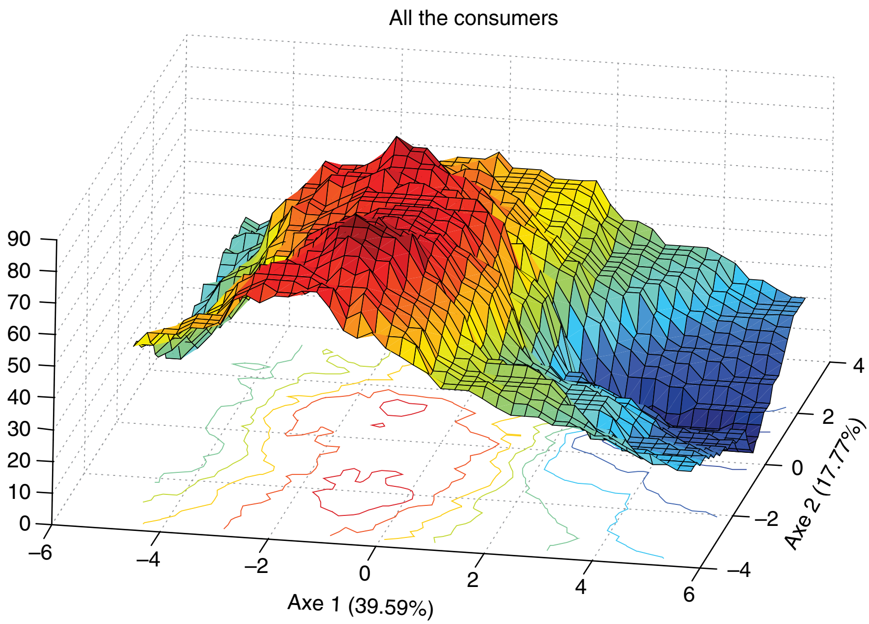


Figure 10.9 External preference mapping, France, consumer segment 1.

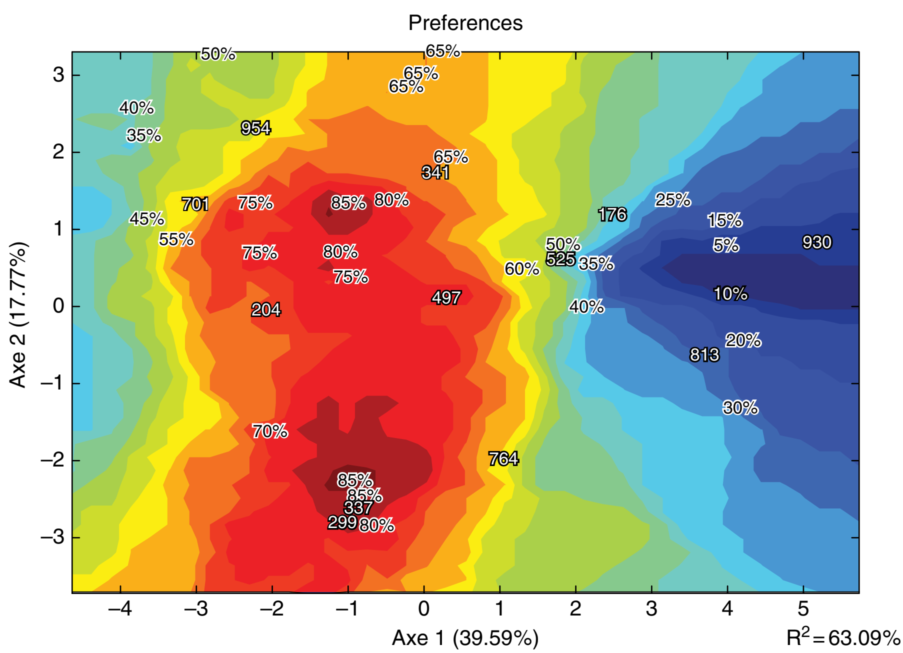


Figure 10.10 Two-dimensional representation of preference mapping, France, consumer segment 1.

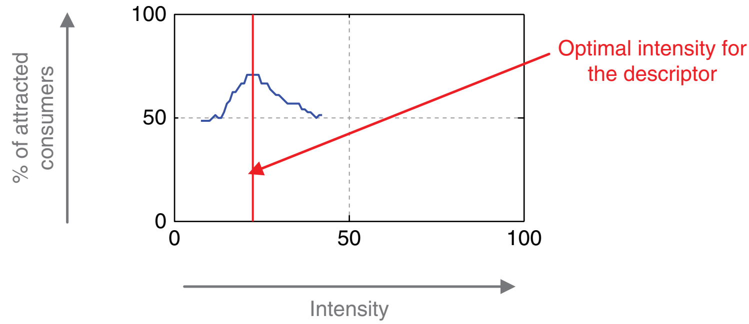


Figure 10.11 Representation of the optimal score for a descriptor in order to maximize liking.

these optimal scores for each descriptor used for the QFP. It consists of selecting three of the actual products that are close to an optimal location. Thanks to the co-ordinates of these four products (three actual products plus the optimal one), the sensory profile scores of each of the three actual products and the barycentre formulas, it becomes possible to estimate the optimal sensory profile.

In addition, using the models created, it becomes possible to go all over the surface according to the intensity of each descriptor. The results obtained are used to create a curve for each descriptor for the area around the selected optimal individual. The curve (Figure 10.11) represents how many consumers would be attracted by the product as a function of the intensity of this descriptor.

All these results and tools can eventually be used to guide the creation of winning flavour solutions.

10.7.2.7 Recommendation for Flavour Optimization

As described in section 10.7.2.6, there are two types of results helpful in the creation of flavours that are appreciated by consumers.

- *The optimal sensory profile*: for one location selected on the preference map, there is one profile estimated. This profile can be read with two complementary approaches. First, for each descriptor, there is a score that corresponds to a certain category of the scale panellists used during the QFP. Hence, the score gives an indication of how this flavour note is supposed to be perceived. The second approach consists of comparing the optimal scores with the ones of an actual product tested in the QFP and the consumer test. However, one should only focus on the descriptors for which the difference between the optimal score and the actual tested product is significant. In consequence, it makes sense to determine whether, for each descriptor, the difference is significant or not. To answer this question, a threshold, that is, a minimum value above which the difference is significant, is computed for each descriptor according to the QFP data. It can also be solved by determining for each descriptor, using the created models, the minimum delta around the estimated sensory profile score that does not

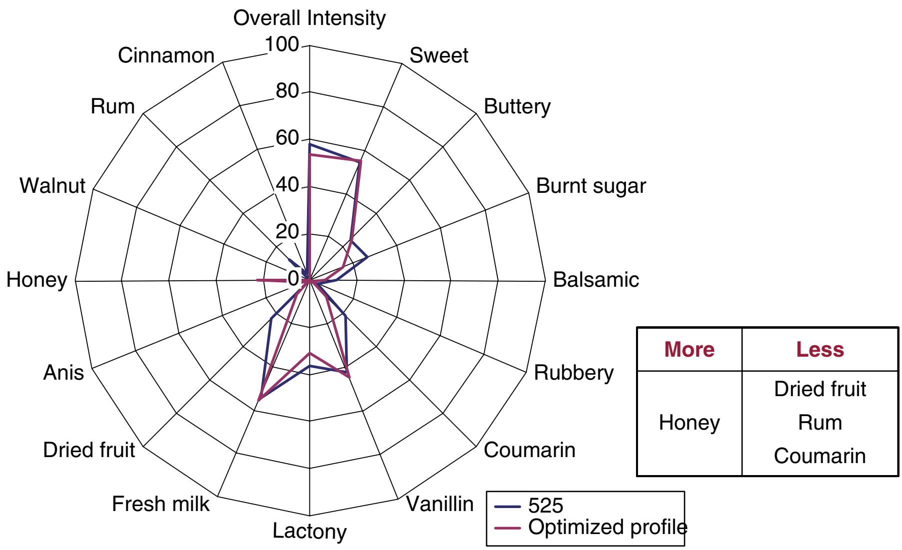


Figure 10.12 Estimated optimal sensory profile maximizing liking for consumer segment 1, France.

impact on the estimated appreciation score. This enables a determination of which descriptor the creation effort should be focused on, starting from the actual product used as comparison.
The optimal sensory profile for the consumer segment of our example is shown in Figure 10.12. By looking at this figure, one can see that taking sample 525, increasing the honey characteristics intensity and lowering rum, dried fruit and coumarin intensities, would lead to a potentially preferred sample.

- *The sensitivity cards*, that is, curves for each descriptor, are also used as indicators of drivers of consumer preferences. This is useful for improvement or free creation of attractive flavours as it gives indications on flavour directions to follow according to consumer expectations. Figure 10.13 shows the sensitivity card for the consumer segment used in our example. This shows the optimal levels for each of the descriptors but also gives insights on how the liking would potentially be affected if the sensory profile deviates from this optimal.

In both examples, the QFP data are fully used to drive the creation of preferred flavours. Having precise QFP scores (intensity of each product and for each descriptor) is necessary to build the models but also to give clues on how to create preferred flavours. The specific language used in QFP that allows the profiles to be understood by flavour creators is also crucial. They will be able to fully understand the sensory characteristics that need to be increased or decreased. Using the references, they will be able to precisely remember and share during the creation process what the desired flavour notes are. This contributes to a faster and more precise creation process.

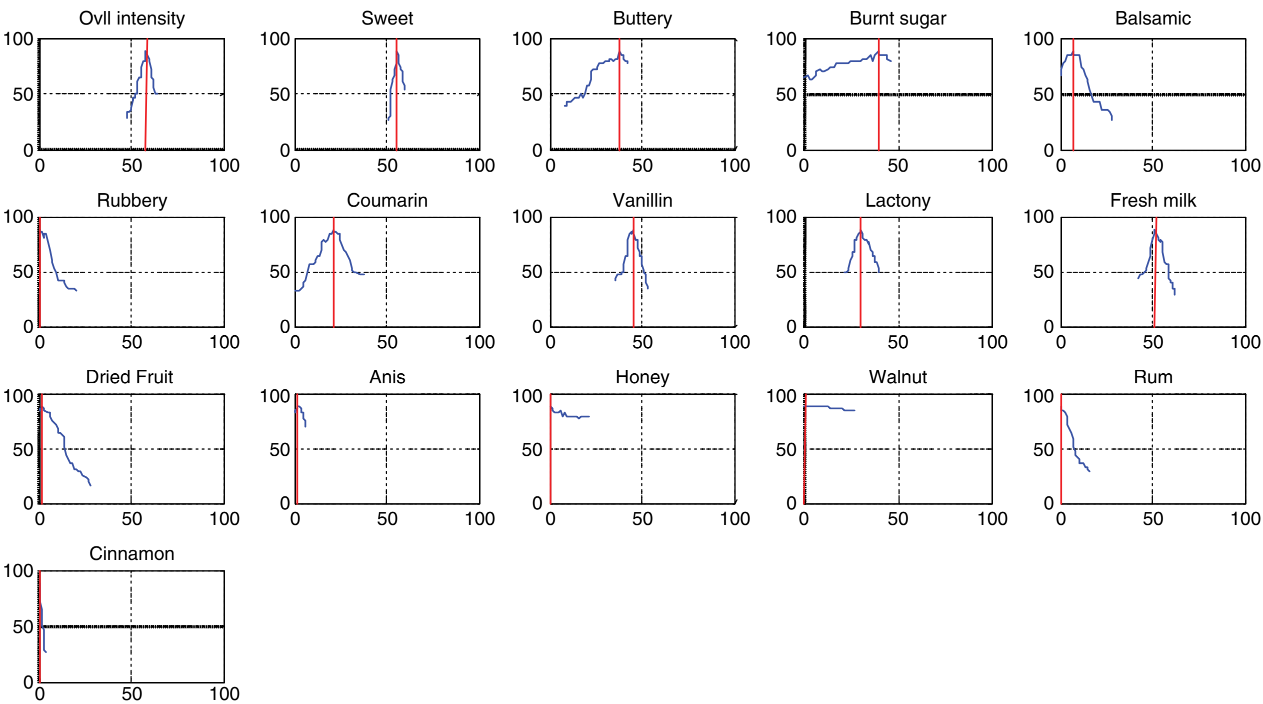


Figure 10.13 Sensitivity card, consumer segment 1, France (x-axis = intensity of the descriptors, y-axis = estimated percentage of consumers liking the product).

10.8 Summary

In this chapter, we have described how QFP is a quantitative descriptive sensory technique fitting the specific needs of flavour evaluation. Through case studies, we have established how QFP addresses the use of sensory descriptive analysis to inspire, guide and validate flavour creation.

The main features of QFP include:

- the universality of its meaning due to the use of a global language with physical references
- the applicability to flavour creation using references developed by and for flavour creators
- the robustness of the description due to extensive training and repetition
- the precision of the description due to the use of highly experienced panels and specific project training.

These features are key for exercises such as:

- the linkage of sensory descriptive data with consumer liking (e.g. preference mapping)
- the optimization of flavours or flavour collections to meet consumer expectations for single or multiple markets
- in-depth understanding of sensory offer in a category (e.g. chocolate ice cream category) or a market (strawberry flavours in France)
- multiple-country sensory evaluations (global analysis of flavour market).

Quantitative flavour profiling is not a fast methodology aimed at rapid decision making. It is not adapted, for example, to using non-trained judges or providing an undetailed overview of a sensory space.

10.9 Future Development

As for many other sensory descriptive methods, QFP looks at each sensory property of a product or a flavour independently from another. The saltiness will be evaluated and given an intensity score, the sweetness will be given an intensity score, so will the fruity character … In the evaluation process, the panellist is not asked at any time to look at these different sensory properties in relation to one another. As an example, the balance between the intensity of sweetness and saltiness will not be considered.

At the same time, one of the key uses of sensory descriptive methods, including QFP, is to try to objectively understand what characteristics of flavours and products drive consumer perception, liking and ultimately choice. To this extent, various methodologies such as external or internal preference mapping attempt to explain consumer perception (e.g. liking) with sensory perception (e.g. sensory descriptive analysis). Sensory scientists often experience the difficulty of fully explaining why consumers express liking for certain products or flavours by solely looking at the sensory descriptive analysis of these products. Indeed, consumer perception of a product or flavour is more complex than single sensations added one to another and consumers do not experience sensory properties independently from each other but all at once for a certain period of time.

The hypothesis that we can obtain a more accurate description of the sensory perception of consumers in order to explain their preferences and choices has been explored, with the inclusion of the temporality of sensations. Methods such as temporal dominant sensations (Labbe 2009) and time intensity look at how we can consider sensations not in a static way but rather as they evolve during consumption time and that the dominance of one versus others may be a key element to consider when trying to fully describe the consumer experience.

Recent QFP developments are looking at how sensory perception is not the addition of single perceptions (sensory attributes in a QFP) but an interaction between all the perceptions to form a holistic sensation. Therefore, QFP aims at integrating more holistic descriptors such as 'balance' which is spontaneously utilized by consumers to express their perceptions but very difficult to describe in an objective manner. The balance definition of a flavour can be, for example, the level to which each of the flavour elements has a similar intensity, with no flavour element dominating the overall profile. A low flavour balance score in a QFP will then be when one or more attributes dominate the overall profile to the detriment of others. A high flavour balance will be when all the attributes have a similar intensity. The difficulty in defining more holistic descriptors is to create the right definitions and the right physical references. These definitions and physical references drive the ability to reach consensus within sensory panels and therefore to utilize these descriptors for objective measurement of sensations as well as guidance for flavour creation and modifications.

References

Ben Slama, M., Heyd, B., Danzart, M. & Ducauze, C.J. (1998) Plan D-optimaux: une stratégie de réduction du nombre de produits en cartographie des préférences. *Sciences des Aliments*, **18**, 471–483.

Danzart, M., Sieffermann, J.M. & Delarue, J. (2004) New Developments in preference Mapping Techniques: Finding Out a Consumer Optimal Product, Its Sensory Profile and the Key Sensory Attributes. Presented at the Sixth Sensometric Meeting, Davis, CA, USA.

Hinkelmann, K. & Kempthorne, O. (1994) *Design and Analysis of Experiments*, vol. **1**. New York: Wiley.

Hsu, J.C. (1996) *Multiple Comparisons*. London: Chapman and Hall.

ISO (1992) *Standard 5492: Terms Relating to Sensory Analysis*. Geneva: International Organization for Standardization.

Krzanowski, W.J. (1988) *Principles of Multivariate Analysis*. Oxford: Oxford University Press.

Labbe, D. (2009) Temporal dominance of sensations and sensory profiling: a comparative study. *Food Quality & Preference*, **20**, 216–221.

Lawless, H. (1983) Contextual effect in category ratings. *Journal of Testing and Evaluation*, **11**, 346–349.

Mosciano, G. (2006) *Successful Flavors – From Formulations to QC to Applications and Beyond*. Carol Stream: Allured Publishing.

O'Mahony, M. (1986) Sensory adaptation. *Journal of Sensory Studies*, **1**, 237–258.

Proust, M. (1913) *Remembrance of Things Past*. New York: Random House.

Schifferstein, H.J.N. (1996) Cognitive factors affecting taste intensity judgement. *Food Quality & Preference*, **7**, 167–175.

Stampanoni, C.R. (1993) The "quantitative flavour profiling" technique. *Perfumer & Flavourist*, **18**, 19–24.

Stone, H. & Sidel, J.L. (1993) *Sensory Evaluation Practices*, 2nd edn. San Diego: Academic Press.

Ward, J.H. (1963) Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, **58**, 236–244.

Weingarten, H.P. & Gowans, S.E. (1991) Sensory control of eating: the meal as a stream of sensations. In: Getchell, T.V. et al. (eds) *Smell and Taste in Health and Disease*. New York: Raven Press, pp. 381–389.

CHAPTER 11

A^{5}daptive Profile Method®

Alejandra M. Muñoz

11.1 Introduction and Fundamental Principles

The A^{5}daptive Profile Method® is a technical descriptive method developed to address issues encountered in several commonly used descriptive practices and to provide solutions for sensory practitioners/panel leaders. In addition, it focuses on custom designing and implementing the most suitable descriptive techniques in order to obtain the most comprehensive and technical qualitative and quantitative product information. Therefore, A^{5}daptive Profile Method's descriptive programmes differ for each user/company. To design each optimal customized programme, attention is given to the specific product categories to be included (which greatly influence the selection of the most suitable descriptive techniques) and the user's specific objectives, needs and preferences.

The A^{5}daptive Profile Method is one of the technical descriptive approaches (ASTM 2011; Muñoz & Bleibaum 2001), and it bases many of its principles on the profile methodology (Brandt et al. 1963; Cairncross & Sjöström 1950; Szczesniak et al. 1963) (see also Chapter 7).

Based on its main philosophical principles and core characteristics, the A^{5}daptive Profile Method developed approaches to:

- provide the most technical and comprehensive qualitative and quantitative descriptive product characterization
- design and implement optimal descriptive practices by considering the programme's needs, the characteristics of the product categories to be included, the type of information desired and the user's philosophical preferences. Therefore, compared to other descriptive methods, the A^{5}daptive Profile Method's training programmes are unique and custom-designed for each user
- provide solutions to problems encountered in many existing descriptive programmes. Through an audit/improvement process, several of this method's techniques, developed to enhance existing practices, are implemented, as needed

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

- design and implement programmes with a holistic view. This method's philosophy is to provide holistic descriptive information, specifically on the following points.
  - The evaluation of all sensory dimensions perceived by the five senses (appearance, taste, smell, touch and hearing), that are most important/relevant for the product categories or of interest to the user, as determined through the A^5 Assessment completed in the design phase. Thus, the A^5daptive Profile Method is suitable for and adaptable to all products, including foods and beverages, personal, oral, healthcare, household, pet care, furniture and other home and office products, etc.
  - The holistic evaluation of products. First, the A^5daptive Profile Method focuses not only on singular attributes, but on integrative attributes that provide the most comprehensive and complete sensory characterization, as needed (e.g. fullness, blendedness, overall impact, etc.). By including these integrative/holistic measures, A^5daptive Profile panels are able to differentiate products which might be identical in their profile but only different in these holistic measures.

    Second, the A^5daptive Profile Method ensures a holistic evaluation of products since it includes, when applicable, other measures to provide the complete/holistic sensory picture of products, such as in-context, temporal and other product information.
- design and implement the unique A^5daptive Profile Method's descriptive practices to ensure the programme's optimal qualitative and quantitative frames of reference, specifically as follows.
  - *Evaluation context*: assessment and establishment of the most suitable product evaluation context (universal or product specific) for the company.
  - *Qualitatively*: assessment of lexicon types and selection of the best option for the user or project, that is, sensory language with only technical/singular attributes, or with umbrella/total impact measures, or with holistic/integrative attributes (e.g. fullness, blendedness, balance, etc.).
  - *Quantitatively*:

    i) use of this method's scale that addresses problems encountered with some other descriptive methods' scoring approaches

    ii) development of customized quantitative/intensity references to best fit the programme's product categories and their intensities and ensure data integrity.
  - *Ballot structure*: attention to the sequence and position of attributes within a ballot to capture all relevant product information based on the display/perception of attributes, and to avoid evaluation and scaling issues, such as duplication and high panel variability.
- develop and establish additional descriptive evaluation tools in a programme to meet specific needs or collect unique information, as needed, such as in-context (in-use) product characterization, overall difference, temporal, consensus measures, etc.

11.2 Methodology

The unique feature of the A^5daptive Profile Method is its approach to customize programmes based on the company's circumstances, philosophy and product characteristics. The customization of descriptive practices applies to new and existing programmes. In both cases an assessment of several parameters (company and product related) is completed as the first step. This process, the A^5daptive Profile Assessment, is described below at two levels:

- for the design and establishment of a new programme
- for an existing programme (i.e. to improve several elements).

11.2.1 The A^5daptive Profile Assessment: New Programmes

Assessment of the company's circumstances and products to develop the customized A^5daptive Profile Method's descriptive programme.

In the A^5daptive Profile Method, the screening and training of assessors is initiated only after the assessment and customization of descriptive parameters is completed. This is a crucial step, since it defines the characteristics of the custom-tailored programme. The A^5daptive Profile Method's programme design, panel selection and training approaches are determined by the descriptive parameters identified as ideal for the specific scenarios in this assessment phase.

The process completed bythe A^5daptive Profile Method involves an assessment of the product categories to be evaluated and their characteristics, user's philosophical preferences, type of information desired, time and budget limitations, and other relevant factors to select or develop the programme's optimal descriptive parameters. This discussion is held by the core programme team (the A^5daptive Profile trainer/panel leader, the manager and members of the sensory/consumer insights group, and other professionals). Based on this information, the optimal A^5daptive Profile Method's programme is designed and implemented.

In the A^5daptive Profile Method, the descriptive parameters developed for each programme are only technically appropriate and strong descriptive approaches and principles. Many of the philosophical principles of this method are based on the profile methods (Brandt et al. 1963; Cairncross & Sjöström 1950; Keane 1992; Muñoz et al. 1992; Szczesniak et al. 1963) (see also Chapter 7).

The A^5daptive Profile Assessment completed for the design of an optimal programme consists of the following steps:

- the A^5 Assessment (see section 11.2.1.1)
- all individual reviews described in sections 11.2.1.2–11.2.1.6, i.e. assessments of:
  - the qualitative product evaluation frame of reference
  - the quantitative product evaluation frame of reference
  - the product use/context and other product parameters and measurements, and
  - the necessary adaptations based on budget and time limitations.

The nature of the discussions held in this meeting and a few details of each of these assessment components are briefly covered below. However, more information and, specifically, the details on how the decisions made in this step impact the panel screening and training approaches are found in the corresponding sections (11.2.4.1 and 11.2.4.2).

Case study I exemplifies the information covered in this section. It presents the A^{5}daptive Profile Assessment process completed by a manufacturer of personal care products in implementing a new programme. This case study illustrates each of the assessment steps, the decisions made in each, and how these selections are considered in the design of the optimal A^{5}daptive Profile Method's descriptive programme. Figure 11.4 is a flowchart of the A^{5}daptive Profile Assessment process for the scenario described in case study I, which aids in visualizing and understanding the whole process.

11.2.1.1 The A^5 Assessment

The selection among the five sensory dimensions (as perceived through the five senses) to be included in the programme.

The core programme team selects the sensory dimensions that are of most interest given project objectives and product categories to be included in the descriptive programme. For example, in foods and beverages, the team would decide which sensory dimensions besides flavour are needed, such as appearance, texture or sound properties. In personal care products, this assessment would determine if only skinfeel (mechanical and tactile) and appearance properties would be included in the programme or if fragrance would also be needed. In cases when only specific sensory dimensions are chosen for the programme, either the company has limited funds and chooses to start a programme initially focusing on the sensory dimension(s) of most interest, or the company's products deliver mostly selected sensory dimensions (e.g. a supplier of flavours might only be interested in flavour evaluations).

In addition, in this assessment a decision is made regarding the scope of each of the sensory dimensions in the programme. Sometimes two or more sensory dimensions require a comprehensive evaluation. Other times, the complete evaluation of only one of the sensory dimensions (e.g. skinfeel) is needed, but only a reduced evaluation of selected attributes in a second or third sensory dimension (e.g. fragrance) is required. For example, a supplier of texture ingredients might be interested in a programme that provides selected information on flavour (e.g. off-notes), but comprehensive data on texture, since flavour is not a key dimension in their products.

The optimal A^{5}daptive Profile Method's training programme is designed by focusing on the sensory dimensions that this assessment unveils as important for the programme.

11.2.1.2 The Qualitative Product Evaluation Frame of Reference Assessment

The nature and complexity of the attributes to be evaluated are discussed by the core programme team. For each lexicon option (see options 1–4 below), the discussions should cover the type of product information provided, advantages and disadvantages over other options, and the training needs. Emphasis should be placed on the advantages of options 2 and 3 over option 1.

Based on the decisions made by the team in this assessment, the trainer is given enough information and background to lead the panel in the development of the desired sensory language (see section 11.2.4.2.3).

In this assessment, the following lexicon types/qualitative frames of reference that can be implemented when establishing an A^{5}daptive Profile Method's descriptive programme are discussed. These options differ in complexity and information provided.

- *Option 1*: technical/singular attributes (e.g. oregano, basil, parsley).
- *Option 2*: option 1 plus umbrella/total impression evaluated for selected attributes (e.g. total herb impression as the corresponding umbrella attribute for the characteristics listed under option 1).
- *Option 3*: option 1 or 2 plus holistic/integrative attributes (e.g. blendedness, fullness, balance, overall flavour, etc.).

  This option is the most complex but most complete qualitative frame of reference. The discussion should address the benefits of this option, such as cases where results show no or very few differences among products when only singular attributes are evaluated. Often, in this case, product differences exist but cannot be captured with singular attributes, only through holistic/complex terms. Case study II presents an example of how product lexicons are enhanced by incorporating more complete and complex sensory language elements.
- *Option 4*: focused lexicon. In some projects and applications, the qualitative frame of reference assessment might lead to the selection of a reduced/focused set of attributes to be included in the training. This scenario might apply in specific applications, such as in quality control (QC), shelf-life and other focused evaluations.

11.2.1.3 Quantitative Product Evaluation Frame of Reference Assessment

11.2.1.3.1 Element 1: Adaptable Scale Based on Intensities

In the meeting with the core programme team, the A^{5}daptive Profile Method's scale's characteristics and advantages are discussed. The scale is a 10-point semi-structured line scale, which is expanded to more points for the scoring of very high intensities. This means that the panel uses the 10-point intensity scale most often, but is trained to use/score higher intensities when needed, by assigning scores higher than 10.

However, the A^{5}daptive Profile Method assesses the type of panel to be trained, user's preferences, etc. to determine if another scale besides, or instead of, its 10-point semi-structured line scale should be considered and taught to the panel.

Scale Use in Research and Development (R&D) Applications

When an A^{5}daptive Profile panel is trained for R&D purposes,the main 10-point A^{5}daptive Profile Method's scale is taught, as well as several other scales, such as verbal, line, five-point scales, etc. While the 10-point scale is used in most projects, with this training, the panel has the ability to adapt and select the most appropriate scale for specific applications/projects.

Scale Use in Other Applications

For specific and focused applications, such as shelf-life or QC measurements, theA^{5}daptive Profile Method selects only one simpler scale to be taught to the panel (e.g. a five-point scale, a fixed 10-point scale, etc.).

11.2.1.3.2 Element 2: Universal versus Product-Specific Context

In the assessment meeting, the A^{5}daptive Profile core programme team discusses the nature, advantages and disadvantages of universal and product-specific rating contexts. The best option for the company is chosen based on the product categories to be included and the objectives of the programme.

A universal quantitative frame of reference or rating context involves the development and implementation of rating scales that can be applied to all product categories to be evaluated. This philosophy allows the comparison of intensities across diverse product categories (Muñoz & Civille 1998).

The universal approach is recommended if (1) the users of the descriptive data are interested in the comparison of intensities across the company's diverse products or (2) when the task of learning intensity references is to be simplified so that the panel learns only a few scales that apply to many categories. However, when each product category is to be evaluated without comparison to other product categories, a product-specific approach is implemented by the A^{5}daptive Profile Method. In this case, the evaluations focus only on products and their intensities within each product category, achieving an evaluation strategy that offers the most discrimination among products of the same category. From the learning perspective, though, when the panel evaluates several categories, this option is more cumbersome, might be more difficult for the panel and takes longer to implement.

11.2.1.3.3 Element 3: Customized Intensity Reference Scales

The A^{5}daptive Profile Method uses quantitative or intensity reference scales, following the practice developed by the original flavor* and texture profile methods (Cairncross & Sjöström 1950; Caul 1957; Muñoz 1986; Szczesniak et al. 1963). However, the A^{5}daptive Profile Method assesses the characteristics of the descriptive programme, the nature of the products and their intensity ranges to develop and implement the optimal intensity reference scales for the programme.

The type of intensity reference scales that are developed based on the programme's characteristics by this method are:

- universal or customized universal scales
- customized product-specific scales
- country-adapted/specific scales
- customized internal product scales.

---

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Case study II presents an example of how intensity references are developed by the A^{5}daptive Profile Method. It demonstrates the importance of assessing the product categories' characteristics and their intensities. This method's philosophy is that universal quantitative references that cover wide intensity ranges are often not the ideal scales for the user.

11.2.1.4 Product Use/Context Assessment (Adaptive Context)

Developing a descriptive programme that provides holistic and complete product information is of great importance in the A^{5}daptive Profile Method. Therefore, it assesses when non-traditional evaluation procedures are needed to provide complete product information. In some product categories, such as in some personal and household products, in-context evaluations are needed to provide a more complete sensory picture/profile, evaluation process and in-use performance of the products.

The importance of these measurements has also been highlighted by other researchers (Cortez-Pereira et al. 2009; Griffiths & Kulke 2002; Kaplan & Okur 2008; Schifferstein 2010). Examples of in-context assessments include the evaluation of complete meals, prepared/complete beverages, and personal care and household products in the context of their use.

When these evaluations and data are needed, as determined in the A^{5}daptive Profile Assessment, the characteristics of the evaluations, needs, training strategies and timing are discussed by the core programme team.

11.2.1.5 Assessment of Other Product Parameters and Measurements (Adaptive Measurements)

Based on objectives, the A^{5}daptive Profile Method also might include other product evaluations or measurements that might be needed in the programme, such as:

- consensus judgements
- overall impact measures
- integrated measurements important for the product or the company
- temporal measurements, etc.

When these evaluations and data are needed, as determined in the A^{5}daptive Profile Assessment, the characteristics of the evaluations, needs, training strategies and timing are discussed by the core team.

11.2.1.6 Assessment of Adaptations Based on Budget and Time Limitations

The A^{5}daptive Profile Method is a technical descriptive method that focuses on implementing the most complete and technically strong descriptive practices. Thus, some A^{5}daptive Profile Method's training programmes (e.g. universal programmes) are lengthy and involved.

However, with budget and time limitations, the training design is modified to limit the programme's scope without compromising quality. Design modifications include focus on selected product categories, staggered training approaches, etc. Section 11.5.2 discusses how limitations are addressed by this method. These ideas and solutions are discussed during the A^{5}daptive Profile Assessment meeting for approval by the core programme team.

11.2.2 The A^{5}daptive Profile Assessment: Existing Programmes

(Assessment/Audit of a Company's Existing Programme)

The same elements as those covered when implementing new A^{5}daptive Profile Method's programmes (see section 11.2.1) are assessed for existing programmes. In this case,the process represents an audit and assessment of the existing programme's characteristics and procedures, data generated, reports, successes, and more importantly the review of all issues/problems thus far encountered. Thus the objectives in this case, compared to those for new programmes, are to determine if the established practices are adequate for the programme, if the generated data have met the company's goals, and to solve the programme's issues. The assessment process might be completed for A^{5}daptive Profile panels, but it is mainly conducted for other descriptive panels/programmes. Only a few issues might be encountered in established A^{5}daptive Profile Method's descriptive programmes, since prior to their implementation the assessment phase is completed (see section 11.2.1), leading to the design of an optimal programme.

The assessment process has numerous benefits for other established descriptive programmes. The A^{5}daptive Profile Method was developed to offer solutions to problems often encountered by descriptive programmes, so it is a beneficial process followed in improving non-A^{5}daptive Profile programmes.

Table 11.1 presents the A^{5}daptive Profile Assessment process followed for an existing programme. All assessment elements described in section 11.2.1 are listed. For each element, the parameters assessed in the audit and the resulting recommendations or needed activities are shown. Case study II exemplifies the process shown in Table 11.1 and covers how a programme can be improved through this assessment and the implementation of the A^{5}daptive Profile Method's philosophy and improved procedures.

11.2.3 Assessment of All Programme Needs Prior to Implementation

Prior to implementing an A^{5}daptive Profile Method's descriptive programme, all its needs are assessed to determine which elements must be addressed and met: management support, sufficient descriptive test requests, required assessors and staff, and facilities.

11.2.3.1 Management Support

Management support is essential for the success of an A^{5}daptive Profile Method's descriptive programme, and without it, the programme cannot be initiated and successfully maintained. The value of descriptive analysis has to be recognized by

Table 11.1 A^{5}daptive Profile Assessment – Existing programme (audit and improvement).*

A^{5}daptive Profile Assessment element**	**Assess:**	**Complete following activities:**
The A^5 Assessment	If additional sensory dimensions not included in the existing programme are needed	Design and conduct training sessions to include additional new sensory dimensions
Qualitative product evaluation frame of reference Assessment (lexicon options)	• If another lexicon option is needed (e.g. umbrella attributes, holistic measures, etc.) • Any issues/problems (e.g. incomplete or misleading data due to lexicon issues, etc.)	• Review existing lexicons, data and reports, and any panel issues • Discuss improved lexicons with the core team, if applicable • Complete sessions to teach a new lexicon development process, or new/improved lexicons, if applicable
Quantitative product evaluation frame of reference Assessment Element 1 (Scale)	• Scale used • Any issues with intensity scoring, use of scale and data	• Review scale used, data and reports, and any panel issues • Introduce a new scale or the philosophy of the A^{5}daptive Profile Method's scale, if applicable
Quantitative product evaluation frame of reference Assessment - Element 2 (Universal vs product-specific context)	• Product categories included in programme • Current product evaluation context vis-à-vis product categories to determine its appropriateness	Design and conduct training sessions, if a different frame of reference/context is more appropriate (e.g. product specific)
Quantitative product evaluation frame of reference Assessment Element 3 (Customized intensity reference scales)	• Appropriateness of current quantitative references relative to intensity ranges of products evaluated • Any issues/problems with intensity scoring, intensity references, use of scale and data	• Review existing intensity references, quant. frame of reference - element 2, scale used, data and reports, and any panel issues • Develop new intensity reference scales to address issues, if needed • Design and conduct training sessions, as needed
Product use/context Assessment	Project requests and data to determine if in-context evaluations are needed	Design and conduct training sessions, if in-context evaluations are to be included
Assessment of other needed measurements	Any additional needed measurements (e.g. consensus, overall impact, etc.)	Design and conduct training sessions, as needed

*See section 11.2.1 for the discussion of this process for new programmes (implementation of an A^{5}daptive Profile Method's programme).
**See case study II, showing an example of this process.

the different management teams, stakeholders and users of the data. The reader is referred to other publications that cover the importance of management support (Gacula 1997; Stone et al. 2012).

11.2.3.2 Sufficient Amount of Descriptive Test Requests

Given the investment made to train and maintain a descriptive panel, it is essential to assess the anticipated project load and panel work prior to establishing an A^{5}daptive Profile Method's programme. It is critical that a sufficient amount of projects requiring descriptive evaluations be expected. Recognition of the descriptive programme's value ensures sufficient descriptive project requests.

11.2.3.3 Assessors

Panel members are an essential element in any descriptive programme. Given the technical nature of the A^{5}daptive Profile Method, it is extremely important to recruit the best candidates. Section 11.2.4.1 presents the process followed to select the best candidates.

11.2.3.4 Facilities

The A^{5}daptive Profile Method places paramount importance on the characteristics, requirements and environmental controls of sensory facilities. The reader is referred to publications that cover the topic (ASTM 2009; Lawless & Heymann 1998) and herein is made aware that proper facility requirements must be met prior to initiating the descriptive programme.

- *Panel room*: required for panel training, orientation sessions, panel reviews and other discussions.
- *Booths and other evaluation rooms/facilities*: most of the general sensory books and publications discuss the characteristics and requirements of booths for the evaluation of foods and beverages. The reader is referred to the publication *Physical Requirement Guidelines for Sensory Evaluation Laboratories* (ASTM 2009) for discussions on special panel booths and evaluation facility features for non-food products.
- *Other facility needs*: in addition to the main requirements above, other facility needs for descriptive analysis include a product preparation area, office space, storage, reception area and other special rooms and facilities, as applicable.

11.2.3.5 Staff

In an A^{5}daptive Profile Method's descriptive programme, the three staff members described in Chapter 7 are required. Only additional staff specific needs for this method are discussed herein, and the reader is referred to Chapter 7 for more details.

11.2.3.5.1 A^{5}daptive Profile Trainer

In addition to the responsibilities described in Chapter 7 (i.e. conduct training sessions, train the panel leader, etc.), the main responsibility of the trainer in the

A^{5}daptive Profile Method is to complete the assessment phase (as described in section 11.2.1) and design the optimal descriptive programme. The assessment is completed in collaboration with the company's core programme team, composed of staff members and the manager of the sensory/consumer insights group, and stakeholders/users of the data.

11.2.3.5.2 A^{5}daptive Profile Panel Leader

An experienced A^{5}daptive Profile panel leader who is familiar with descriptive analysis and has already participated in training programmes can be the trainer and complete the activities described in section 11.2.4.2.2. Without this experience, the panel leader works closely with the trainer and is trained as an assessor, as well as in the tasks of a trainer and panel leader.

TheA^{5}daptive Profile panel leader responsibilities depend on his/her role within the company: (1) with exclusive panel and panel leader responsibilities, or (2) with both panel and project management responsibilities.

Panel Leader with Exclusive Panel Responsibilities

In some organizations, the panel leader only has panel responsibilities and is in charge of all panel activities related to training, maintaining and working with the panel in product evaluations, such as:

- screening and training assessors
- scheduling assessors for panel sessions
- supervising the preparation of samples and references (in some cases being in charge of the preparation)
- conducting panel sessions/acting as panel leader
- resolving discrepancies and being involved in panel monitoring efforts
- recording and compiling panel results.

Therefore, essential skills of the panel leader are good leadership, organizational and communication skills to be able to effectively complete the above activities. When consensus judgements are gathered, the panel leader should have the ability to work with diverse opinions and facilitate discussions that enable the panel to reach consensus.

Panel Leader with Both Panel and Project Management Responsibilities

In addition to the panel responsibilities described above, an A^{5}daptive Profile panel leader might be in charge of project management responsibilities, data analysis and reporting of panel results. In this case, the panel leader responsibilities might also include:

- being involved in project planning and test design as described in section 11.2.5
- interacting with the test requester
- designing the project/tests
- interpreting and reporting of results
- working with/guide panel administrator (if another individual/panel leader takes care of the panel responsibilities described above).

11.2.3.5.3 Technician's Role in the A^{5}daptive Profile Method's Programme

The main responsibilities of a technician working in this programme are similar to those described in Chapter 7.

11.2.3.6 Support in the Acquisition of Products and References

The A^{5}daptive Profile Method extensively uses qualitative and quantitative references throughout the training, in project work and panel maintenance. Therefore, it is critical to seek and ensure support from other technical groups within the company, who may assist in the production or acquisition of certain references needed throughout the training.

It is important that this collaboration be established at the onset of the training programme, and that the technical group(s) be briefed on the programme's needs, the nature of the required references, their presentation frequency, timing, etc.

11.2.4 Implementation of the A^{5}daptive Profile Method

11.2.4.1 Screening of Assessors

The A^{5}daptive Profile Method's screening process is designed based on the characteristics of the programme to be implemented and the parameters selected in the assessment phase (see section 11.2.1). The objective of the customized screening process is to select the best candidates for the programme.

This method's screening process has three phases, which explore:

- non-acuity qualifications
- acuity and scoring abilities
- personality and group behaviour.

11.2.4.1.1 Non-Acuity Qualifications

The parameters investigated in this phase to determine which candidates qualify to participate in the acuity and scoring tests are:

- availability and commitment
- interest in the programme
- health and other factors that might prevent the candidate from participating in the programme, or that might interfere in the product evaluations.

All this information is obtained through a questionnaire and a brief interview. This interview allows further exploration of the information gathered in the questionnaire, and the opportunity for a one-to-one interaction with each candidate to assess basic personality traits.

Availability and Commitment

It is important that the complete programme schedule (for the training, practice and routine project sessions) be reviewed with candidates to assess any potential conflicts. For additional information on how to explore this area, see Chapter 7.

Interest in the Programme

In general, assessors' interest in the programme decreases with time. Therefore, candidates must have a high level of initial interest, and be enthusiastic about the programme and the process. In addition, it is important that the panel leader maintains the assessors' interest in the program by completing different interesting activities and providing motivation on an ongoing basis.

An accurate description of the complete programme, both for training and routine project work, must be communicated to the candidates to assess their reaction to and interest in the programme. In addition, and if possible, candidates are asked in the questionnaire and during the interview to share information and experiences from previous jobs, on activities with family and friends, as well as on hobbies, trips, favourite foods, movies, etc. The responses to these questions and the enthusiasm when sharing the information are assessed, since they provide insights into the candidates' attitudes.

Health and Other Potential Limitations

The nature of the questions asked to explore general health and potential limitations depends on the type of evaluations to be completed and the products/ingredients to be evaluated. It is recommended that the questions be reviewed by the company's human resources or legal group to ensure their appropriateness. Once the questionnaire is approved, to the extent possible, overall health, allergies and sensitivities, intake of medications potentially affecting acuity and mental focus/concentration, etc. should be explored to determine any potential issues and conflicts.

In the A^{5}daptive Profile Method the questionnaire or interview is adapted based on the sensory dimensions to be evaluated (thus senses involved and potentially affected), and on the type of products and ingredients. For example, and as described in Chapter 7, allergies, frequent colds and sensitivity to high intensities of chemical feeling factors must be investigated as they represent problems in the evaluation of flavours and fragrances. Issues interfering in the perception of texture-related attributes in foods and non-foods should be investigated, such as excessive calluses, carpal tunnel syndrome, dentures/implants, etc.

Attitude Towards Product Use Restrictions and In-Context Evaluations

In all training programmes, assessors must follow certain pre-session instructions, which might include refraining from using perfumes and colognes, smoking, eating or drinking highly flavoured foods, etc. Furthermore, there might be additional restrictions in the evaluation of personal care products. For example, there are restrictions on the use of own deodorants, mascaras, make-up, lipsticks, lotions, etc. prior to sessions when any of these categories are the test products. Candidates' attitudes towards these and other restrictions should be explored in the screening process.

In addition, when the A^{5}daptive Profile Method's programme includes in-context evaluations, candidates' potential reservations towards these evaluations should be explored. Individuals expressing any concern with above or similar product use limitations or the nature of in-context evaluations should not be selected to participate in the programme.

11.2.4.1.2 Acuity and Scoring Abilities

Sensory acuity and scoring ability exercises are designed based on the A^{5}daptive Profile Method's programme's characteristics to be implemented and the decisions made in the Assessment phase (see section 11.2.1). The factors that determine the type of screening exercises required include:

- the sensory dimensions (e.g. appearance, aroma/fragrance, texture/tactile characteristics) and the type of products to be evaluated
- the scope of the programme (i.e. involving diverse product categories/universal programme versus one product category/product-specific programme).

Acuity Exercises Based on Sensory Dimensions and the A^{5}Assessment

The programme's relevant screening exercises are designed based on the sensory dimensions to be evaluated, and are chosen from the ones presented below.

- *Taste*: when this screening test is applicable (e.g. in foods, oral care products, etc.), candidates' basic taste acuity is investigated, such as acuity to sourness, saltiness, sweetness and bitterness. See Chapter 7 for more details of this test.
- *Aromatics, aroma/odour and fragrance* (olfaction).
  - Odour recognition test. A series of odorants (8–15 test odorants) are presented blind, and candidates are asked to identify the odour/smell of each sample. Odorants should be moderately strong in intensity and somewhat common, such as spices, fruits, simple floral notes, etc. Also it is important that a few of the odorants be related to some of the key characteristics of the programme's product categories. Candidates are asked to provide a description (that can be specific or general) or an association. For example, correct answers for lime citrus oil might be citrus, lime, lemon, furniture polish, etc. If possible, another set of samples is presented with other more complex materials and odours, such as those present in finished products (e.g. beverages, foods, fragrances, lotions, household products, etc.).
  - Ability to integrate concepts for holistic/integrative attributes. When holistic measurements/attributes are included in the A^{5}daptive Profile Method's training programme, it is important that a screening exercise be included to assess the ability of candidates to integrate characteristics in evaluating overall/holistic attributes. Interestingly, since candidates are not trained, the completion of some simple holistic measurements is often an easy task, since as 'consumers/untrained individuals', people are accustomed to integrate perceptions. This exercise usually includes one or two pairs of samples that vary in a holistic measurement, such as overall flavour intensity or blendedness, or in a specific holistic attribute of interest in non-foods, such as softness, moisturizing, etc. The pair of samples should be extremely different in the holistic measure for untrained candidates to understand and complete the exercise without much difficulty.

- *Touch/texture*: the acuity tests in this sensory dimension are specific to the product categories included in the programme. In descriptive programmes testing foods and beverages, the A^{5}daptive Profile Method includes 2–4 simple texture exercises. These attributes should be:
  - simple: mastication attributes are complex (e.g. moistness of the mass, cohesiveness of mass). Therefore, these attributes are not usually included in this method's screening exercises
  - product relevant: a few of the critical texture dimensions of the products to be evaluated are included in the screening exercises, such as crispness/crunchiness for snacks, softness/firmness in many food products, thickness in juices and other liquids, etc.

    In descriptive programmes testing non-foods, such as personal or paper products, the touch/texture characteristics included in screening exercises might be one or two mechanical characteristics (e.g. spreadability, stiffness) and one or two tactile attributes (e.g. roughness/grittiness). Alternatively, the A^{5}daptive Profile Method includes one simple attribute from each of the product's stages of evaluation (e.g. appearance, pick-up, rub-out and afterfeel in skinfeel evaluations).
- *Appearance*: in most cases, the appearance acuity screening test that the A^{5}daptive Profile Method uses is the colour blindness test (Ishihara 1969). In addition, specific appearance acuity exercises might be designed focusing on key product characteristics, such as colour intensity, uniformity of colour, shininess, etc.

Acuity Exercises Based on Product Categories Included (i.e. Universal versus Product-Focused Programme)

The A^{5}daptive Profile Method designs acuity screening differently for universal and product-focused programmes.

Because a universal programme incorporates diverse product categories, it is sufficient to test acuity for the relevant general sensory dimensions chosen in the assessment phase (e.g. smell/olfaction, taste, etc.), and outlined above. However, in a product-focused programme, the screening exercises are conducted testing relevant variables and intensities of the specific product category(ies) of interest. Acuity for these sensory characteristics can be tested either through discrimination or attribute rating tests. For the latter, a simple ballot listing key and simple sensory characteristics is used for candidates to rate those products' attributes. In this case, it is important that the products show medium to large differences, and that the attributes be simple and easily understood by untrained participants.

Exercises to Explore Abilities to Rate Intensities and Use Quantitative References

In the A^{5}daptive Profile Method the scoring ability of candidates is assessed in two ways.

- *Visual scoring exercises*: candidates score areas, sizes, volumes, etc. of shapes, images or pictures.
- *Exercises to score sensory characteristics*: candidates rate the intensities of simple sensory characteristics (e.g. a basic taste, a general flavour or fragrance, a simple texture, skinfeel/hairfeel attribute, etc.) using a scale.

In addition, embedded in the scoring tests are exercises to test the ability of candidates to use quantitative references to score intensities, because this is a key practice in the A^{5}daptive Profile Method. In one or two screening exercises, candidates are familiarized with the concept of intensity references and are asked to rate the intensities of simple attributes relative to a quantitative reference. The reference sample is identified with an attribute intensity (e.g. total overall intensity in a fragrance) and candidates are asked to remember that intensity. They then evaluate test samples and are asked to rate the specific attribute intensity (e.g. total overall intensity) relative to the identified reference's intensity.

11.2.4.1.3 Personality and Group Behaviour

During the training of an A^{5}daptive Profile panel, assessors work in a group and continuously interact with one another in activities, such as lexicon development, review of references, development and improvement of ballots, review of data, etc. Therefore, it is imperative to select panel members who enjoy the interaction and are able to work well with other participants and the group.

In general, panel members should be respectful of other people's opinions, flexible and willing to compromise, interested in the programme and activities throughout the training, comply with restrictions on product usage delineated at the beginning of the programme, avoid controversial or dominating behaviour throughout the training and in panel activities, and overall have a positive attitude and the ability to work with the whole group in a positive and productive way.

In this method's screening, an interview and one or two group activities are completed to explore general personality and group behaviour. While these activities could be part of the process described in 11.2.4.1.1, the A^{5}daptive Profile Method recommends that they be completed once a smaller group of candidates has been identified and passed the previous screening phases (see sections 11.2.4.1.1 and 11.2.4.1.2). The objective is to focus on the best candidates identified thus far, and base the final selection on personality and group behaviour.

If the interview has not been conducted as described in 11.2.4.1.1, the trainer or panel leader meets with each candidate for one-to-one interactions. Besides further exploring health and other issues from the written questionnaire, the interview gives the panel leader a chance to examine how the candidate behaves and responds to the different questions, and to get a general idea of the person's personality.

There are several group activities that can be conducted with the smaller groups of best candidates, such as a group discussion on a given topic, participation in a mock panel session, etc. In this panel session, candidates interact with one another and discuss the products provided. This gives the panel leader or trainer the opportunity to observe behaviours, group dynamics, individual and group participation and the outcome of discussions.This final and very valuable information is taken into consideration to choose the assessors for the descriptive programme.

11.2.4.2 The A^5daptive Profile Method's Training Approach

Traditional descriptive methods use one fixed training methodology. However, the A^5daptive Profile Method uses several approaches, because each training is custom designed to establish the optimal programme. Therefore, in this section, the A^5daptive Profile Method's training approach is covered as follows:

- the presentation of several training approaches that give the reader an understanding of how this method custom designs each programme (see section 11.2.4.2.1)
- the discussion of this method's philosophy, methodology and training techniques across the different training approaches (see section 11.2.4.2.2).

11.2.4.2.1 The Diverse and Customized Training Approaches of the A^5daptive Profile Method

To exemplify this method's approaches, Figure 11.1 presents a schematic overview of three different A^5daptive Profile Method's programme designs and training approaches. The key characteristics of each programme are as follows.

- *Universal panel/programme*: the term 'universal' in a descriptive programme refers to a panel capable of evaluating a variety of product categories. Thus, the training has to cover the key elements of all categories. The A^5daptive Profile Method strategically designs a universal training programme to be able to use the learning and elements of one product category as the basis for the next product category. Thus an effective training programme is conducted, minimizing time and effort.
- *Product-specific panel/programme*: the term 'product specific' in a descriptive programme refers to a panel trained in one (or a few related) product category/ies. Thus, the training only covers the elements of that specific product category. The A^5daptive Profile Method implements this approach in the following scenarios.
  - Descriptive results needed for only one product category: this might be the case for a company manufacturing only one product category, or when there is a special focus on one key category.
  - Focused evaluations: an A^5daptive Profile panel is trained to evaluate selected attributes of interest, such as in QC/sensory programmes, shelf-life studies, etc.
  - Product screening: in this application, product attributes are chosen to train a panel in order to screen a variety of products based on those attributes.
  - Other applications where a focused evaluation is needed.

As shown in Figure 11.1, product-specific programmes are shorter and focused. They also require a less involved panel screening and occasionally a simpler scale might be taught/used.

- *In-context panel/programme*: the A^5daptive Profile Method establishes an in-context panel/programme when the evaluations are conducted in the

STEP 1. **A^5DAPTIVE PROFILE ASSESSMENT PHASE**

For all training programs, the first step is the completion of this assessment phase (section 11.2.1.) and the custom-tailored design of the descriptive program

STEP 2. **SCREENING AND SELECTION OF ASSESSORS**

STEP 3. **PANEL TRAINING (three examples shown below)**

UNIVERSAL*
A^5daptive Profile Method®

PHASE I FLAVOR (could be Texture) — ***Initial training and practice sessions***

- ► Fundamental Concepts
- ► Lexicon and ballot development process to be followed for all product categories
- ► Scoring principles and use of intensity references (review of customized universal scales for flavor characteristics)
- ► Practice and guidance in ballot development and evaluation procedures for simple categories (from Universal product set)

PHASE II FLAVOR (could be Texture) — ***Advanced training and practice sessions***

- ► Evaluation procedures for complex categories (from Universal product set)
- ► New intensity scales, as needed
- ► Practice and guidance - product evaluations

PHASE III (could be Phase IV) — ***Advanced training and practice sessions***

- ► Advanced concepts
- ► Integrative attributes
- ► Other measurements (e.g., consensus)
- ► Practice and guidance - product evaluations

PHASE IV TEXTURE (could be Flavor) — ***Initial training and practice sessions***

Very similar to phase I, but focused on second sensory dimension (e.g., Texture)

PHASE V TEXTURE (could be Flavor) — ***Advanced training and practice sessions***

Very similar to phase II, but focused on second sensory dimension (e.g., Texture)

Panel validation

PRODUCT FOCUSED/SPECIFIC
A^5daptive Profile Method®

PHASE I — ***Initial and focused training and practice sessions***

- ► Fundamental Concepts
- ► Review of product category variables and product category specific lexicon/ballot development and evaluation exercises
- ► Scoring principles and use of intensity references (review of customized product specific scales for aromatics, texture, skinfeel, handfeel etc.)
- ► Practice and guidance in simple evaluations

PHASE II — ***Advanced training and practice sessions***

- ► Advanced concepts
- ► Integrative attributes
- ► Other measurements (e.g., consensus)
- ► Practice and guidance - product evaluations

Panel validation

IN-CONTEXT
A^5daptive Profile Method®

PHASE I — ***Initial training and practice sessions***

- ► Fundamental Concepts
- ► Review of product category variables and product category specific lexicon/ ballot development and evaluation exercises
- ► Scoring principles and use of intensity references (review of customized universal or product specific scales for aromatics, texture, skinfeel, handfeel etc.)
- ► Practice and guidance in simple evaluations

PHASE II — ***Advanced training and practice sessions***

- ► Advanced concepts
- ► In-context evaluation procedures and attributes
- ► Practice and guidance - product evaluations

PHASE III (if needed) — ***Advanced training and practice sessions***

- ► Advanced concepts
- ► Integrative attributes
- ► Other measurements (e.g., consensus)
- ► Practice and guidance - product evaluations

Panel validation

* This example is presented for foods (i.e. flavor and texture). However, a similar approach is followed for the development of universal panels for non-food products (e.g., paper, household products etc.,). In this training, the first phase focuses on flavor. However, the training can also focus on texture first, then flavor.
In personal care products, the first phase may cover fragrance, and the next ones may cover skinfeel, hair feel, etc.; or a different order may be followed.

Figure 11.1 Schematic overview of three different A^5daptive Profile Method's programmes.

context or conditions of actual product use. These types of programmes are important to consider in the evaluation of personal care, household and other non-food products where product characteristics evaluated in a controlled environment (i.e. in traditional descriptive laboratory settings) do not capture the complete profile and performance of the products.

Figure 11.1 illustrates the training philosophy of the A^{5}daptive Profile Method and the main characteristics of the approaches chosen as examples, and the differences and similarities among the diverse training approaches. A brief discussion of these topics follows.

Training Philosophy of the A^{5}daptive Profile Method and Main Characteristics of Diverse Approaches

- *The philosophy of designing customized descriptive training programmes and techniques based on assessments and decisions*. Section 11.2.1 explained the process followed in the A^{5}daptive Profile Assessment to design the optimal descriptive programme. Based on this philosophy and process, customized descriptive training programmes are designed to meet different needs. Figure 11.1 shows three completely different descriptive programmes designed to meet the different objectives and needs of three groups. This illustrates how adapted/custom-tailored programmes differ in length, flow, focus and content to meet the specific objectives.
- *The philosophy and importance of covering foundations and fundamental concepts of descriptive analysis and product evaluation*. In all A^{5}daptive Profile Method's programmes (see Figure 11.1), the fundamental concepts of sensory science, descriptive analysis and the sensory dimensions of focus (e.g. aroma/smell/fragrance/aromatics, taste, texture/skinfeel/hairfeel) are covered initially. Assessors learn and apply these concepts in initial stages,and develop the bases to understand and learn more advanced concepts covered in later stages.
- *The philosophy of a sequential introduction/learning of descriptive techniques*. In addition to the A^{5}daptive Profile Method's training philosophy being adaptive and flexible, it is sequential. This method's training approaches focus initially on the most simple product categories (for universal panels) or simple concepts (attributes and evaluation principles), followed by more complex product categories and attributes, advanced concepts and evaluation practices. Figure 11.1 shows how the complexity of products and concepts increases as the training progresses, both within a phase and across training phases.
- *The philosophy of customizing technically sound qualitative and quantitative descriptive frames of reference*.
  - Qualitative training modules: as shown in Figure 11.1, training and practice modules which focus on terminology are common to all training approaches, given the importance that this method places on lexicon development. The complexity of the sensory language in each programme (i.e. singular, umbrella, holistic or focused attributes) is determined based on the user's philosophy, desired outcome and nature of the product categories (see section 11.2.1.2).

  - Quantitative training modules: carefully designed quantitative/scoring training modules are taught across all diverse training approaches (see Figure 11.1). They are designed to establish a technically strong descriptive scoring system and the intensity reference framework to be used throughout the programme. The most suitable scale and scoring context are chosen, and the most appropriate intensity references based on the programme's needs and characteristics of the product categories are developed and taught to the panel.
- *The philosophy and objectives of practice sessions*. The A^{5}daptive Profile Method emphasizes the importance of practice sessions and guidance to the panel. After each training period (see Figure 11.1), practice sessions are conducted to ensure that the panel applies and practises the principles and techniques learned, and becomes more comfortable with the tasks. In addition, it is imperative that feedback be given to the panel, concepts be retaught and issues be addressed, as needed.

Differences and Similarities among A^{5}daptive Profile Method Training Approaches

The differences among this method's training approaches are as follows (see Figure 11.1).

- *Scope and nature of product evaluations*: the A^{5}daptive Profile Method designs the programme that will meet the company's objectives and thus provide the type of data and information needed. As a result, training programmes differ in scope and their training modules are unique (i.e. different designs to teach the sensory dimension of interest, depth (i.e. complete versus reduced characterization), specific evaluation techniques, such as in-context evaluations, consensus process, etc.).
- *Focus*: training programmes have a different focus depending on objectives.
  - Focus based on product categories. Universal programmes cover, as the name indicates, a large/diverse set of product categories, while product-specific programmes focus on one or two selected product categories. Therefore, the latter training programmes are very focused and concentrate on a limited number of/selected attributes and intensity ranges.
  - Focus based on evaluation or data collection procedures. Training programmes might differ in the specific evaluation procedures they focus on, such as controlled/traditional evaluations, in-context evaluations, consensus, complete evaluations or evaluations focused on selected stages (e.g. focused only on the product application stage in non-foods), etc.
  - Focus based on attributes. To meet specific objectives, a training programme might focus only on selected attributes of interest (e.g. QC, shelf-life programmes, etc.).
- *Qualitative and quantitative frames of reference*: training approaches might differ:
  - qualitatively, based on the type and complexity of lexicons developed and used
  - quantitatively, based on the type of scale and intensity references used.
- *Total duration*: the duration of each programme differs and is dependent on its scope and focus. For example, a universal programme is comprehensive

and thus long. In comparison, a product-specific/focused training programme is much shorter since it focuses on one or a few specific product categories of interest and sometimes only on specific attributes.

The similarities among this method's training approaches are (see Figure 11.1):

- the completion of the A^{5}daptive Profile Assessment phase, which determines the specific programme characteristics, design and training approach
- the design and completion of a thorough screening and selection of assessors
- a sequential training philosophy and approach. Initially, the fundamental and basic concepts of lexicon and ballot development, scoring and product evaluation are taught. Then, the more complex product categories, attributes or concepts are covered
- the design and execution of practice sessions conducted after each training phase
- the importance given to both the qualitative and quantitative components. Thus, the necessary training modules to develop strong qualitative and quantitative frames of reference are designed and implemented
- the use of intensity references.

11.2.4.2.2 The A^{5}daptive Profile Method Training Process Across All Training Programmes

As shown in Figure 11.1, training and practice sessions, which increase in depth and complexity, are completed in sequential phases (e.g. phases I, II, III, etc.) in the A^{5}daptive Profile Method's descriptive training programmes. A brief summary of the activities completed in each of these phases is described below.

Initial A^{5}daptive Profile Method's Orientation/Training Session (Phase I)

In this first phase the panel learns the following programme elements.

- General concepts: sensory methods and applications, with emphasis on descriptive analysis.
- Fundamental concepts of the sensory dimensions of focus to be evaluated in the programme (based on the A^{5}Assessment),which might include one or all of these sensory dimensions.
  - Appearance
  - Smell: aroma/aromatics for foods, beverages, oral care, OTC and other products; odour or fragrance for non-food products
  - Taste: the five basic tastes
  - Touch: mechanical, somesthetic/geometrical and fat/moisture attributes (texture, skinfeel, handfeel)
  - Hearing: the auditory attributes relevant to the product categories (e.g. pitch, intensity, etc.).
- Evaluation procedures and controls (especially important in texture, skinfeel, handfeel programmes).
- Qualitative component: lexicon and ballot development.
- Quantitative component.
  - Scoring principles and the characteristics of the selected scoring approach (i.e. ranking, rating, etc.).

  - Intensity/quantitative references. Section 11.2.4.2.3 presents a detailed discussion of the A^{5}daptive Profile Method training approach on the qualitative and quantitative components.
- Exercises involving lexicon/ballot development and product evaluations.

Phase I practice sessions are designed for the panel to practise the concepts learned in the first orientation/training phase. Four to six hours per week of practice are recommended. Results are summarized and statistically analysed. During this time it is important that the panel leader (or sensory scientist) establishes the data analysis system to be used in routine project work and in monitoring panel performance.

Feedback is provided to the panel in two ways: 1) immediately after each or several of the evaluation sessions, and 2) upon completion of each practice exercise as follows.

1) Immediate feedback after each practice evaluation session may not be possible. However, this feedback is important because assessors will have just completed the session's evaluations and can remember their perceptions. In addition, samples are available for review. This feedback primarily focuses on the panel's ability to identify the main product differences, extreme outliers and panel agreement.

2) Feedback upon completion of each practice exercise (all evaluation sessions). This feedback is provided once all evaluations in the practice exercise are completed, the data have been analysed, and assessors' individual performances have been assessed. Therefore, this feedback is more thorough. It provides more information on the product sensory differences and similarities found by the panel but, more importantly, on individual performances relative to the whole panel's performance. Comparison of each practice exercise's data as the training progresses shows the panel's progress and provides information on any issues that need to be revisited (i.e. attribute definitions, qualitative and quantitative references, scaling issues, etc.).

Advanced A^{5}daptive Profile Method's Orientation/Training Sessions (Phases II, III, etc.)

The main objectives of the advanced training sessions are to cover more complex concepts and product categories, and to introduce the programme's more focused and advanced tasks, attributes or evaluation procedures (see Figure 11.1). Therefore, the activities in these sessions include the following.

- Work with more complex products, flavour or fragrance families, complex texture/skinfeel/handfeel attributes and scales, etc.
- Introduction and evaluation of complex sensory attributes (e.g. balance, fullness, etc.).
- Development of complex ballots.
- The consensus process, in-context evaluations, etc. (if applicable).
- Development of new or improved intensity references, to complement the sets introduced in Phase I.
- Review of issues encountered by assessors, etc.

A practice session follows each of the advanced orientation/training sessions. Similarly to Phase I practices, the objective of these sessions is for the panel to apply and practise all principles learned in the previously conducted orientation/training session(s).

As in Phase I, a series of review exercises and evaluations are planned by the trainer for assessors to apply the more advanced concepts learned. The panel develops and uses more complex lexicons and ballots, completes more advanced product evaluations and strengthens group interaction and collaboration. Ultimately, the most important objective is for the panel to continue building confidence in the process and in their skills.

Panel Validation

Once all the training phases are complete and the data show that the panel has achieved an acceptable level of performance, an A^{5}daptive Profile panel should be validated. This descriptive method recommends that the data from the last practice sessions be assessed and/or a validation study be designed and conducted, as explained below.

Panel Performance Measures. The best way to validate an A^{5}daptive Profile panel is to assess the recently trained panel's performance data. This method recommends that the panel data from several exercises conducted at the end of the training be assessed (e.g. panel data from a series of exercises during the latter part of the practice period). With this approach, several sample sets and/or product categories can be included to better assess and validate the panel.

Several panel performance measures can be considered to determine the training level and performance of the recently trained A^{5}daptive Profile panel, such as:

- scale usage
- discrimination ability/sensitivity
- repeatability
- agreement
- reproducibility.

See section 11.6.1. on panel performance for more information.

Validation Study and Comparison with Trained Panel or Consumer Data. Another option to assess the panel's readiness to initiate project work is to design and conduct a validation study. The product categories, the specific samples and the measurements (e.g. singular, integrative, in-context attributes, etc.) are selected. This study should include one or two replications.

In general the validation study should show product differences in the attributes where differences exist, good panel agreement and expected outcome.

Alternatively, the approach explained in Chapter 7 can be used to validate an A^{5}daptive Profile panel.

- Comparison of results from the recently trained panel with results from an experienced/trained panel.
- Comparison of results from the recently trained panel with consumer attribute data for those attributes for which consumers have demonstrated their understanding and the ability to differentiate among products.

Assessment of Panel Performance/Validation Study Results and Additional Training and Calibration (If Needed)

The results from the validation study or the series of evaluations completed at the end of the practice sessions are assessed. If the data show good panel performance and/or the panel is validated, the panel can initiate project work.

If the results show that additional training is necessary, this author recommends conducting a panel session first to review and discuss the validation/panel performance results with the panel. In this (these) session(s), the panel leader and panel review the issues unveiled by the assessment/validation study. In addition, the panel is asked to provide input on any issues encountered and through these discussions, information not necessarily shown by panel scores and the corresponding statistical analysis is unveiled. For example, often, the difficulties/issues may not be related to the actual attributes, definitions or evaluation procedures, but to sample presentation, environmental and other non-product issues. The assessors are often the only ones who can provide insights and suggestions on the modifications that evaluation procedures and other elements might need.

The validation study's data together with the panel input allow the panel leader to address the issues that need to be modified and to design the necessary retraining/calibration exercises and accomplish the following:

- address issues with sample preparation and presentation
- review specific attributes' procedures and definitions
- review and improve intensity reference scales
- modify parameters related to product evaluations such as rest periods, number of samples, fatiguing effects, etc.
- resolve other issues as needed.

Following the additional training/panel sessions, new evaluations are designed and completed (i.e. a second validation study).These evaluations' results are assessed to determine if the issues encountered in the first validation/evaluations have been addressed. If these additional training sessions have been successful, the panel is ready to initiate project work.

11.2.4.2.3 Characteristics of the A^{5}daptive Profile Method's Training Approach on the Qualitative and Quantitative Descriptive Components

As described in section 11.2.4.2.2, the qualitative (lexicon development) and quantitative (intensity scoring) descriptive components are taught as key tools in the first training session. These concepts are further reinforced in the practice and ensuing training sessions.

The A^{5}daptive Profile Method approaches the training on these two elements in a unique way compared to other descriptive methods, as described below.

Training on the Qualitative Component: Development and Types of Lexicons

In the A^{5}daptive Profile Method, the lexicon development process has three steps.

- Review of samples/product category and individual sensory language development.
- Review of references and development of consensus/group lexicon.
- Evaluation of products, discussion of results and lexicon validation.

The sensory language development process above has been described in detail by Muñoz (in Moskowitz et al. 2003) and the reader is referred to this publication for additional information.

During the training of an A^{5}daptive Profile panel, this process is taught for simple product categories to facilitate the learning process and make the panel feel comfortable. During the first exercises, the focus is on the process, the panel interaction and the guidance that the trainer offers. At later stages, in training and practice sessions, the focus shifts to the outcome and the quality of the lexicons generated.

Section 11.2.1.2 described the way the A^{5}daptive Profile Method's qualitative frame of reference assessment and resulting decisions determine the nature/complexity of the lexicon/attributes to be established and used by the panel. Therefore, the training approach for the development of the sensory language is based on these decisions, and is specific to each lexicon option as described below.

Training Option 1: Technical/Singular Attributes. For this option, the general process described by Muñoz (in Moskowitz et al. 2003) is taught during the training.

The A^{5}daptive Profile Method pays special attention to the following during training.

- *Technical nature*: this descriptive method's objective is to develop the most technical panel terminology/lexicon. Therefore, during training, an extensive array of qualitative references is presented. The objectives are to discuss attributes in detail and choose the most appropriate, specific and technical terms (e.g. benzaldehyde, diacetyl, cohesiveness of the mass, force to gather, etc.) to be able to provide specific guidance to product formulators in future project work.
- *Order of attributes*: this descriptive method pays special attention to the order in which the attributes are placed on the ballot and thus evaluated. The factors that determine the position of the attributes are taught during the training: relevance, attribute relationships and order of appearance (and other time-related issues).

The panel is taught the process to assess and decide on the order of attributes, as follows.

- Once attributes are agreed upon, discussions should be held regarding the product's attribute relationships, perception profile over time and relevance within the category and the project. These factors determine the final characteristics and nuances of the ballot and the best position of the attributes.
- The panel is to check if the final ballot captures all relevant product information, and if potential duplication has been avoided (since it leads to panel variability and other panel and evaluation problems).

Training Option 2: Technical/Singular Attributes Plus Umbrella/Total Impression Measures. In this option's training, the panel is shown the benefits of the umbrella/total impression attributes and cases where they should be included in the lexicon. As the name indicates, umbrella/total impression attributes capture the total impact of the flavour or texture family. Examples include total herbs, total fruit, total meat, total residue (in skinfeel evaluations), etc. Case Study II illustrates the benefits of umbrella attributes and how their inclusion can address panel issues.

During this option's training, the techniques to determine the need for umbrella/total impression attributes is taught during the second phase of the sensory terminology development process (i.e., in the review of references and development of consensus/group lexicon). The panel is taught to hold a discussion in this phase to determine which umbrella attributes are needed, and the singular terms captured and to be evaluated under each of these umbrella/total impression measures.

Training Option 3: Holistic/Integrative Attributes Added to Option 1 or 2. As shown in Figure 11.1, the training covering this concept and the evaluation of these attributes are conducted in later phases of the programme, given their complexity. Often, this training component is not implemented until the panel has worked on projects.The need for these terms cannot be recognized and learned until the panel has had enough experience completing product evaluations and the review of these results.

This option, being the most complex, is best implemented in two phases. In the first phase, the panel is trained following the standard A^{5}daptive Profile Method's training approach focusing on singular, or singular and umbrella attributes (options 1 or 2). The second phase is completed once the panel is trained and experienced. Training exercises are designed and completed to cover these complex attributes, where samples and products are used that show low and high intensities of these holistic attributes. It is important that the panel learns these characteristics in more than one product category to become experienced in integrating different singular attributes in the evaluation of the holistic attributes. For example, blendedness should be taught for several of the company's product categories where this holistic attribute is most relevant (e.g. juices, sauces, etc.). Through these exercises, the panel learns how, in diverse product categories, different sensory attributes play a role in the evaluation of the integrative attributes.

In this option's training it is important to cover cases where product differences are perceived, yet the data show no or very few differences among products with the exclusive evaluation of singular attributes. In these cases the need and value of holistic attributes are demonstrated since it is shown that differences could not have been captured with only singular terms.

Case study II presents an example of how existing lexicons can be improved by incorporating more complete and complex elements in the sensory language.

Training Option 4: Focused Lexicon. The lexicon development training process is different for this option, as shown in Figure 11.1. Given the focused scope, the otherwise involved lexicon development process and thus training are shortened, specifically step 2 (review of references and development of consensus/group sensory language).

Because the characteristics of the focused lexicon have been discussed by the core programme team during the assessment phase (see section 11.2.1.2), a shorter sensory language may already have been developed. In this case, during training, the trainer shares with the panel an initial list of attributes to be considered for the final sensory language. The panel reviews the range of products (step 1 of the lexicon development process) and has input in the development of the final sensory language (shortened step 2 – the trainer and the panel make decisions on the final focused lexicon through the review of products and references).

Training on the Quantitative Component: Scaling, Quantitative Frame of Reference and Intensity References

The characteristics of the programme's quantitative component are determined in the A^{5}daptive Profile Assessment (see section 11.2.1.3). Decisions made during this assessment determine the nature of the three quantitative elements discussed below and taught to the panel during training.

Element 1: Adaptive Scale. During the training, A^{5}daptive Profile panels learn two aspects of scaling.

- Main scale: this scale is used during the programme's product evaluations. For most R&D applications, the 10-point A^{5}daptive Profile Method's scale is taught, and how it is adapted to rate higher intensities when applicable.
- Other scales and scoring approaches are taught to the panel during training. Thus, the panel has the ability to adapt and select the most appropriate scale for specific applications/projects.

Element 2: Universal versus Product-Specific Context. The training on the quantitative component/scaling differs when a universal versus a product-specific panel is developed.

During the training and practice exercises, A^{5}daptive Profile universal panels learn the philosophy of universal scaling and how the same intensity scales are used to score intensities for diverse product categories. Conversely, product-specific panels learn different intensity scales for the different product categories, and are trained to independently evaluate and rate the product categories. This strategy is most effective when only one or a few product categories are included in the programme, or for categories whose attributes are perceived at low to medium intensities, where a universal scale poses limitations.

Element 3: Customized Intensity Reference Scales. This is one of the unique characteristics of the A^{5}daptive Profile Method, since this method strives to develop and use the most effective quantitative reference scales. This method's viewpoint on intensity scales coincides with its overall philosophy of developing and implementing the optimal descriptive programme's parameters and techniques.

Often, within a given sensory dimension, one general universal scale cannot be effectively used for *all* products and/or applications. Therefore, the A^{5}daptive Profile Method develops or custom-designs quantitative scales as described below. These customized/optimized intensity scales are taught to the panel during the training. Each panel uses these optimized scales in all training and practice evaluations, as well as in project work.

Universal scales have been developed to encompass the complete range of intensities in a given sensory dimension (e.g. basic tastes) or an attribute (e.g. moistness, spreadability, etc.) (ASTM 2011; Muñoz 1986). These scales are successfully used when diverse product categories are evaluated by the panel. However, they are traditionally used without assessing if these quantitative scales, which encompass the complete range of intensities, are appropriate. While a panel might still be universal (i.e. evaluate diverse product categories), the general universal intensity scales might not be appropriate for that programme.

The A^{5}daptive Profile Method develops universal adapted intensity scales when the focus on a range of intensities is beneficial for a programme. Once developed, the panel is taught the customized universal scales to be used in product evaluations. Case study II shows an example of how customized universal intensity scales are developed and successfully applied.

The A^{5}daptive Profile Method develops *product-specific* intensity scales, when needed, which focus on the product categories of interest and their intensity ranges. Specific references in these customized intensity scales demonstrate the relevant/focused intensity ranges for the product category(ies) to be evaluated. Products within the category that span the intensity range of interest are selected as references.

There are other cases requiring the developmentof customized intensity reference scales. Examples include country-adapted scales (Hough et al. 1994), project-specific scales, QC scales, etc.

This method considers the specific objectives and properties of the product categories to be evaluated to determine the need and the characteristics of the specific and optimal intensity scales for the programme or project. These specific needs are discussed during the A^{5}daptive Profile Assessment phase (see section 11.2.1.3).

11.2.5 Panel Operation and Project Product Evaluations in an A^{5}daptive Profile Method's Programme

The A^{5}daptive Profile trainer works with the panel leader close to or upon completion of the training to review the required activities in descriptive projects. Discussions include the role of descriptive tests in sensory/consumer projects, necessary interaction with test requesters, planning for test products, test designs, etc. The review of all these activities falls beyond the scope of this chapter, so this discussion focuses only on the planning and execution of descriptive studies. However, it is emphasized that the A^{5}daptive Profile Method considers that descriptive projects should not be designed and executed in isolation, but always in the context of the overall project. This way, a descriptive programme successfully aids in the company's projects and the decision-making process of technical and business aspects.

11.2.5.1 Descriptive Projects in the Context of the Overall Sensory/ Consumer Projects

This method emphasizes the need to consider the descriptive studies as part of the overall company's projects, because their objectives and needs affect and define the descriptive project. Thus, this method recommends that the descriptive objectives, sample needs and amounts, and timing be discussed during the initial stages of the overall sensory/consumer project planning.

Once the overall project objectives are clearly defined and all test samples are available, this method recommends that the panel leader promptly initiates some of the descriptive project's activities. Sometimes, despite the test samples not being available, the panel leader can start working with the A^{5}daptive Profile panel in some of the pre-project activities described below, if appropriate. Once the test samples become available, the panel leader should screen those products to then design and complete the descriptive pre-project and project activities and evaluations.

11.2.5.2 Pre-Project Activities

Activities to be completed prior to formal project evaluations depend on the panel's previous work and expertise with the product category being evaluated. Less pre-project work and time are needed when the panel has conducted previous evaluations and has experience with the product category.

11.2.5.2.1 Panel Activities without Previous Experience with the Product Category

Table 11.2 shows all the pre-project panel activities to be completed when the panel lacks exposure to the product category. Several sessions might be needed to complete these activities and for the A^{5}daptive Profile panel to be ready to initiate formal project evaluations.

Table 11.2 A^{5}daptive Profile Method's pre-project activities.

Panel pre-project activity	Previous experience with product category: NO	Previous experience with product category: YES
Terminology/lexicon and ballot	• Need to develop for the new product category (as per process followed during training. See section 11.2.4.2.3 • Process is speedier, since at this time the panel is trained and has developed enough expertise in this activity	• Only review and, if needed, adjust existing ballot • To determine needed ballot adjustments, one or a few test samples might be reviewed, possibly with non-project prototypes
Review of references	Those required for the lexicon/ballot development a) qualitative references to demonstrate or clarify attributes b) quantitative/intensity references as needed, for calibration purposes	Only selected references are reviewed, if needed a) no qualitative references if panel conducts frequent evaluations in this category b) quantitative/intensity references as needed, for calibration purposes
Practice product evaluations and discussion of results (samples should be carefully selected. Set might include one or selected test products, and other non-project prototypes or products)	Practice evaluations needed. Through the discussion of results the following is accomplished: • refinement of lexicon • decisions on additional/needed qualitative and quantitative references to review • review of product issues affecting the evaluation techniques, and needed adjustments • feedback to the panel • review of assessors' questions	Optional. Only required, if: • no recent product evaluations have been conducted, or • considerable changes were made to the ballot in reviewing some of the test samples If conducted, the same objectives as described in the process without experience are accomplished (e.g. lexicon refinement, panel feedback, etc.)
Lexicon/ballot fine-tuning and review of selected references based on practice results	• Review of additional qualitative and quantitative references selected by the panel • Final adjustments made to the lexicon/ballot	Same as in process without experience
Final product evaluation practice	Optional. Only needed with considerable lexicon/ballot changes made after the discussion of practice results	None needed

11.2.5.2.2 Panel Activities with Previous Experience with the Product Category

When the A^{5}daptive Profile panel has experience and has previously worked with the product category being evaluated, the process described in section 11.2.5.2.1 is simplified and much shorter (see Table 11.2).

11.2.5.3 Product Evaluations

In this method, the project descriptive evaluations are initiated once the pre-project panel sessions have been completed.

11.2.5.3.1 Individual Evaluations

The individual product evaluations are completed following established protocols, using the lexicon and ballot developed or refined in pre-project activities, and under strict test controls and protocols. Samples are presented following sound sensory practices and the chosen test design (e.g. balanced order of presentation, design with the constant reference, BIBD, etc.).

In the A^{5}daptive Profile Method, replications are recommended to be able to study interaction, batch and other effects of interest in the analysis of variance (ANOVA). Data are summarized and statistically analysed, as described in section 11.6.

11.2.5.3.2 Consensus

When consensus results are obtained, individual data are not summarized and statistically analysed. However, the individual evaluations are completed as described above.

Once all assessors have completed their evaluation, a group discussion follows to review each of the attributes and agree on their intensity scores. In addition, the panel discusses the main differences and similarities among products once consensus on all attributes has been reached.

11.2.5.4 Data Analysis/Review and Interpretation (Project Product Evaluations)

11.2.5.4.1 Statistical Analysis and Interpretation

A^{5}daptive Profile data (assessors' individual attribute intensities) are summarized and statistically analysed, using an ANOVA approach (see section 11.6). The test design determines the specifics of the ANOVA. The product/sample effect/source of variation is the most important component in the analysis. Other effects are analysed depending on the design, such as session and the interaction effects of interest.

Additional analyses are completed depending on objectives (see section 11.6).

11.2.5.4.2 Consensus Data

When the panel generates product information through consensus, the results are summarized and reported as consensus intensities. In this case, no statistical analysis is completed.

11.2.5.5 Reporting of Results

In Chapter 7, Muñoz and Keane stated that descriptive reports are not method-specific, and presented their ideas and recommendations on the contents of a descriptive analysis report.

In the A^{5}daptive Profile Method the reports follow the outline presented in Chapter 7, placing importance on the background, overall sensory/consumer project objectives and test design. This method considers that the descriptive study is part of the overall sensory/consumer project and thus the former should be designed, conducted and reported as one of the overall sensory/consumer project's components.

In addition, the A^{5}daptive Profile Method stresses the importance of the report's executive summary, which often is the only component that is carefully reviewed by the user and management. Thus, the executive summary should concisely present the objectives, main findings, conclusions and recommendations. Also, it should emphasize the contributions of the descriptive results to the overall project's objectives, and discuss the descriptive findings' impact on the technical and business decisions.

11.3 Advantages and Disadvantages of the A^{5}daptive Profile Method

11.3.1 Advantages

The A^{5}daptive Profile Method has many advantages (Figure 11.2). Among others, it was developed to address issues encountered in several commonly used descriptive practices. In addition, this method's techniques focus on establishing a descriptive capability that provides the most comprehensive and technical qualitative and quantitative product information.

11.3.2 Disadvantages

11.3.2.1 Time Investment in the Design Phase

Time and effort are involved in the design phase and in working with the core programme team to design and establish the optimal A^{5}daptive Profile Method's descriptive programme. While the outcome of the design phase offers many advantages, its completion represents an additional step that must be incorporated into the programme schedule.

11.3.2.2 Length of Programme

Investing the time required to establish an A^{5}daptive Profile panel is considered a disadvantage by some professionals. However, this time investment is needed to develop an A^{5}daptive Profile Method's technical descriptive programme with the best qualitative and quantitative descriptive elements. Universal programmes, in particular, are long and elaborate.

The length of the programme increases as the scope of the following elements is expanded: product categories, in-context evaluations, complex lexicons (e.g. inclusion of holistic attributes) and customized quantitative references.

Adaptive and flexible approach in the design and establishment of the optimal descriptive program

The optimal program is designed and the most suitable descriptive practices for the program are implemented by considering the company's needs, user's preferences, and characteristics (and intensity ranges) of the product categories to be evaluated (see section 11.2.1)

Holistic descriptive approach, evaluation techniques and lexicons

- A holistic comprehensive sensory evaluation: The evaluation of all perceived sensory dimensions or the most important/relevant dimensions of interest to the company
- Development and implementation of all needed evaluation techniques for a holistic/complete product characterization, e.g. :
  - Complex attributes in addition to singular attributes (e.g., total intensity, fullness, blendedness, overall impact, etc.), as needed
  - In-use product evaluations for a complete product characterization. The implementation of in-context evaluation techniques,when needed (e.g., in-context use of personal, household and other products).

Customized/adaptive methodology for any product category or sensory dimensions

The A^{5}daptive Profile Method custom designs and develops descriptive protocols, product application and evaluation techniques, and lexicons for any product category (e.g., traditional food texture techniques are adapted to develop those for color cosmetics (e.g., mascara, lipstick, etc.), hair and home care products, writing instruments, etc.)

Technical, detailed and actionable product guidance and accurate product documentation for diverse applications

A^{5}daptive Profile data offer to developers very technical and detailed sensory information to be able to successfully target their formulation or reformulation efforts (e.g., information on dairy fat, smoothness/lack of grittiness, etc., instead of "creaminess")

Customized and technically strong qualitative and quantitative frames of reference

- Customized/optimal qualitative frame of reference
  The optimal lexicon/qualitative frame of reference is established for each user (see section 11.2.1.2.) (e.g., singular, singular and umbrella, complex/integrative attributes, etc.)
- Customized/optimal quantitative frame of reference
  Special attention is given to scale preferences, use of the data, and the intensity ranges of the product categories to determine the elements of the ideal quantitative frame of reference: scale, context of evaluation (universal versus product specific), and type of customized quantitative/intensity references scales

Implementation of other relevant or needed measurement techniques

The A^{5}daptive Profile Method develops and implements other product evaluation measurements that might be needed to provide the most comprehensive product characterization (e.g., consensus, difference, temporal measures etc.)

Figure 11.2 Advantages of the A^{5}daptive Profile Method.

11.3.2.3 Panel Maintenance

The A^{5}daptive Profile Method pays special attention to maintaining the panel's skills in both the qualitative and quantitative descriptive areas. Therefore, time is required to:

- acquire and prepare all the qualitative and quantitative references needed in project work and panel maintenance
- schedule the necessary panel sessions prior to projects and those required in the panel maintenance programme.

11.4 Applications of the A^{5}daptive Profile Method

This method's data have many applications and provide useful information to support the projects and efforts of different groups within an organization, especially research and development (R&D), quality control/assurance, and marketing and market research.

The A^{5}daptive Profile data have the same applications as those described in Chapter 7.

11.5 Practical Considerations

11.5.1 Do's and Don'ts

Figures 11.3a, 11.3b and 11.3c present the do's and don'ts that are key in the implementation and operation of a successful A^{5}daptive Profile Method's programme. The information has been divided into Do's and Don'ts for:

- preliminary activities (prior to training) (Figure 11.3a)
- panel training (Figure 11.3b)
- project work and panel maintenance (Figure 11.3c).

11.5.2 Programme Constraints and Possible Solutions

Often, constraints might hinder the establishment of a complete or optimal A^{5}daptive Profile Method's programme, or the execution of the most complete/sound product evaluations.

In these cases, some modifications may be considered. However, they have to be carefully assessed to ensure that the quality of the programme is not compromised. Section 11.2.1.6 covered how constraints and limitations are reviewed during the A^5 Profile Assessment. Possible solutions to address these limitations should be approved by the core programme team in that phase.

Below are a few scenarios on how constraints drive modifications of an optimal A^{5}daptive Profile Method's programme.

A note of caution: while some suggestions are given on how to address restrictions and constraints, it is emphasized that the practices below do not represent ideal scenarios. Therefore, they should be carefully assessed prior to their implementation. In addition, sensory practitioners facing constraints are encouraged to build on these modified/shortened and less than ideal practices (i.e. add-ons, improvements and more complete and ideal practices should be implemented as soon as constraints lessen).

11.5.2.1 Budget Constraints

The following training (11.5.2.1.1) and project work (11.5.2.1.2) alternatives/modifications to an A^{5}daptive Profile Method's programme might be considered with budget constraints.

Do not

- Start any planning or program design until you have completed the A^{5}daptive Profile Assessment
- Start the training unless you have
  - Management and stakeholders' support
  - A qualified trainer or panel leader to effectively conduct the training
  - A sufficiently large pool of qualified candidates
  - Motivated and interested candidates
  - Communicated the commitment and the training/project work schedule to assessors
  - Communicated main do's and don'ts to candidates and eliminated those who express a concern with disallowed practices (e.g., no scented products, no application of deodorant or make-up priot to testing similar products, etc.)
  - All needed training procedures, products and references
  - Enough expected project work to justify the descriptive program's investment

Do

- Understand the philosophy of the A^{5}daptive Profile Method prior to initiating the program
- Ensure management supports the program and all its needs, and understands the benefits
- Engage stakeholders from the onset to ensure support
- Identify partners in product development or outside resources to ensure support with the supply of needed references
- Complete all the recommended steps in the A^{5}daptive Profile assessment to design the optimal program
- Design the program and prepare well in advance prior to starting the training to ensure a successful program
- Recruit a qualified panel leader or trainer to successfully lead the training
- Complete the thorough screening process to ensure the selection of the best candidates
- Recruit a sufficiently large panel pool
- Communicate commitment, schedules and main do's and don'ts to assessors
- Initiate product and reference acquisition immediately after training scope has been determined, and program is approved and designed

Figure 11.3a Do's and don'ts – Preliminary activities (prior to training). The A^{5}daptive Profile Method.

11.5.2.1.1 Training Alternatives

- Consider another descriptive method that is not as technically strong but requires less or no training.
- Select only 2–3 key product categories and complete an initial shortened training focused on these selected categories. The programme can be expanded at a later time, by introducing additional categories.
- Reduce the training time (and costs) by shortening or eliminating one or several of the training activities, as follows (it is recommended that these by-passed activities be completed at a later time once the panel has completed the basic training).
  - Use of a reduced set of qualitative references to demonstrate attributes or quantitative/intensity references (or temporarily postpone the use of intensity references).

Do not

- Allow a long break in between sessions while training
- Allow uncommitted and problem assessors to continue participating in the program
- Conduct long sessions and create fatigue
- Move to next training exercise until issues encountered are addressed
- Be afraid of making modifications on training plans, if designed program is not working well or as expected
- Introduce advanced concepts (e.g. consensus training) unless the panel has mastered the previously learned concepts

Do

- Always be prepared for the training sessions
- Always be on time to reinforce commitment to assessors
- Have all needed samples and references for the training sessions
- Stress uniform evaluation techniques/controls to assessors from the onset (i.e., sample size, smelling, tasting, applying product)
- Stress do's and don'ts to assessors
- Provide frequent feedback to assessors during training
- Always complete all required practice sessions prior to introducing a new concept or product category
- Ensure completion of make-up training and practice sessions
- Be positive, supportive and provide positive feedback to assessors
- Foster a positive and relaxed environment and friendly relationships
- Be creative with examples and references that help the panel learn

Figure 11.3b Do's and don'ts – Training. The A^{5}daptive Profile Method.

- ∘ Consider the completion of a focused training (e.g. subset of attributes, shortened evaluations) (*Note*: caution should be exercised in the selection of attributes to ensure that all the products' critical attributes are included).

11.5.2.1.2 Project Work Alternatives

- Eliminate one/several of the replications.
- Focus on selected/key attributes (*Note:* sensory professionals must ensure that the key attributes or those showing differences are included).
- Collect product information through group consensus, instead of collecting and analysing individual data.
- Screen products and have the panel characterize only the products with perceivable differences.

Do not

- Initiate project work until panel has been validated
- Start a project unless
  - the project and test objectives have been clearly delineated
  - enough amount of and adequate samples are available for testing
  - the pre-project activities have been conducted
- Overlook the importance of panel performance and feedback

Do

- Ensure strict test controls (e.g., environment, product, serving conditions)
- Complete the necessary pre-project activities to ensure panel has reviewed applicable references and established the optimal lexicon/ballot
- In reports, ensure that the descriptive findings relate back to the overall project, and the technical and business objectives
- Follow-up with project managers and test requestors after report has been issued to answer questions and discuss descriptive panel contributions and other future program applications
- Establish a panel performance program to be able to monitor the panel and provide feedback
- Provide feedback to panel on projects, contributions to company and performance
- Motivate your panel

Figure 11.3c Do's and don'ts – Project work and panel maintenance. The A^{5}daptive Profile Method.

11.5.2.2 Time Constraints

11.5.2.2.1 Training Alternatives

A few of the recommendations provided when facing budget constraints apply (see section 11.5.2.1). It is recommended that once the urgent project has been addressed, sensory professionals plan on expanding the training programme by capitalizing on the issues which required the urgent training.

- With real time constraints, do not complete any training but collect other product information that may help the test requestors (i.e. difference/similarity, acceptance data, etc.).
- Subcontract descriptive evaluations for the specific project.
- Consider another descriptive method that is not as technically strong but requires a shorter or no training.

- Complete an initial training programme focused on the product category(ies) of interest or to address the urgency. The programme should be expanded at a later time, by introducing additional categories.
- Reduce the training time by shortening or eliminating several of the key training activities, as described in section 11.5.2.1.1.

11.5.2.2.2 Project Work Alternatives

See the suggestions given in section 11.5.2.1.2.

11.5.2.3 Lack of Support: Product References Materials for Training and/or During Panel Maintenance

The following alternatives/modifications in the acquisition of reference materials might be considered with time constraints.

- Use commercially available products as training references. Screen before using.
- Prepare own references using easily accessible ingredients.
- Hire a consultant to design and prepare needed training references.
- Engage trained assessors in acquiring/preparing references.

11.5.2.4 Staff Limitations

The following ideas might be considered with time constraints.

- Hire required staff.
- Hire temporary help.
- Train non-sensory staff on sensory principles and descriptive analysis in order for them to assist in training programmes and project work.
- Hire one or several interested and responsible assessors to complete selected pre- and postpanel activities.
- Hire a consultant to assist in the training activities and project work.

11.5.2.5 Attrition and Small Panel Size

The following recommendations might be considered when faced with a small sample size due to attrition.

- Develop and implement a programme to regularly add assessors to the existing panel.
- Assess the reasons for attrition and address internal circumstances or issues that may be causing attrition.
- Motivate assessors.
- Assess salaries (for outside panels) and determine if appropriate for region and task.

11.6 Statistical Analysis of A^{5}daptive Profile Data

Descriptive projects differ in objectives, focus (products or panel), characteristics and number of samples, necessary evaluation procedures, etc. These characteristics determine the test design, the test execution specifics, and in turn the statistical analyses needed for the resulting data, as follows.

An A^{5}daptive Profile panel leader should be familiar with three general statistical analysis areas needed to analyse the programme's descriptive data, as follows.

11.6.1 Panel Performance and Monitoring Analyses

Statistical tools are needed to monitor the panel and assessors' progress during training, in validating the panel, and in monitoring the panel performance once trained. In these analyses, the focus is the panel and the individual assessors. As presented in the discussion of panel validation, the A^{5}daptive Profile Method recommends the following measures to monitor the panel's performance.

- *Scale usage*: ability of a panel or assessor(s) to accurately rate an attribute within a specific area or point on a scale. This measure is particularly important in this method since intensity references are used. They are intended to standardize the scale usage among assessors.
- *Discrimination ability/sensitivity*: the ability of a panel to perceive, identify and differentiate products when differences exist.
- *Repeatability*: ability of a panel or assessor to repeat its/his/her own ratings from one session to another one.
- *Agreement*: ability of a panel or assessor to rate/order products similarly to other panels or assessors on a given attribute.
- *Reproducibility*: ability of a panel or assessor(s) to reproduce the results of other panels or assessors.

The reader is directed to Chapter 4 in this book and other publications that have discussed the importance of these and other panel performance measures, and/or have covered the statistical approaches for these measures (ASTM 2017; Hunter & Muir 1995; McEwan et al. 2002; Næs & Solheim 1991; Sinesio et al. 1990).

11.6.2 Procedures for the Analysis of Routine Project Descriptive Data

In these analyses the main objective is to study the product/sample effects and other test variables included in the design. Since the focus is on the samples and other product variables tested, the interest in assessors becomes secondary, since it is assumed that they are trained and calibrated. However, an assessment of interaction effects involving assessors is always recommended.

The test design and the factors/sources of variation of interest in the study determine the statistical analysis needed to analyse A^{5}daptive Profile data. In most cases, ANOVA is used to analyse the data.

The main ANOVA sources of variation analysed in A^{5}daptive Profile data are:

- assessors (A)
- sample/product (P)
- replication (R)
- session (or day) (if applicable) (S)
- other sources of variation
- interactions (e.g. Ax P, PxS, etc.) depending on interest, the test design and model used.

Chapter 5 discusses statistical analysis philosophies and ANOVA analyses used for descriptive data.

11.6.3 Procedures for the Analysis of Descriptive Data from More Complex Projects

As discussed in section 11.4, A^{5}daptive Profile data have many applications. The test objectives and execution, and the nature and number of products evaluated by the panel, differ in each project and determine the necessary statistical analysis.

In multifaceted or more complex projects, a large number of test samples, more complex designs and the inclusion of other data sets (e.g. consumer, instrumental) might be part of the research and descriptive projects. In these cases, more advanced statistical tools are needed to analyse the descriptive and other data sets, requiring the application of experimental design, data relationships, multivariate and other advanced statistical analyses. This discussion falls beyond the scope of this chapter but the reader is referred to Chapter 5 for additional information and to other publications (Gacula & Singh 1984; Lawless & Heymann 1998; Næs et al. 2010).

In many organizations, panel leaders have support from the internal statistics group for the completion of some or all of the data analyses described above. In this case, panel leaders are not required to complete the analyses and be familiar with statistical programmes and procedures. However, panel leaders must have basic background in the statistics concepts that apply to descriptive data and their analyses to be able to effectively interpret these analyses and apply the findings.

11.7 Case Studies

Two case studies are presented that illustrate some of the most important characteristics of the A^{5}daptive Profile Method.

Case study I illustrates how the A^{5}daptive Profile Assessment is completed, the way the different descriptive parameters' options are assessed, and the decisions made throughout the process. This information is the basis for the design of an optimal descriptive programme. Additionally, this case study shows how decisions differ depending on the sensory dimensions included in the programme, and thus the importance of the A^5 Assessment.

Case study II shows an example of a descriptive programme audit, and the way the A^{5}daptive Profile Method assesses the parameters of an existing programme and provides solutions to occurring programme's problems.

11.7.1 Case Study I: A^{5}daptive Profile Assessment and Design of an A^{5}daptive Profile Method's Descriptive Programme

A manufacturer of body lotions and creams wished to implement an A^{5}daptive Profile Method's descriptive programme for the evaluation of its products.

As per this method's philosophy, the assessment of the descriptive parameters' options (A^{5}daptive Profile Assessment) was completed to design the optimal programme for the company, as described in section 11.2.1.

Figure 11.4 is a flowchart which clearly illustrates the complete process followed and the decisions made in each step. This process is briefly described below.

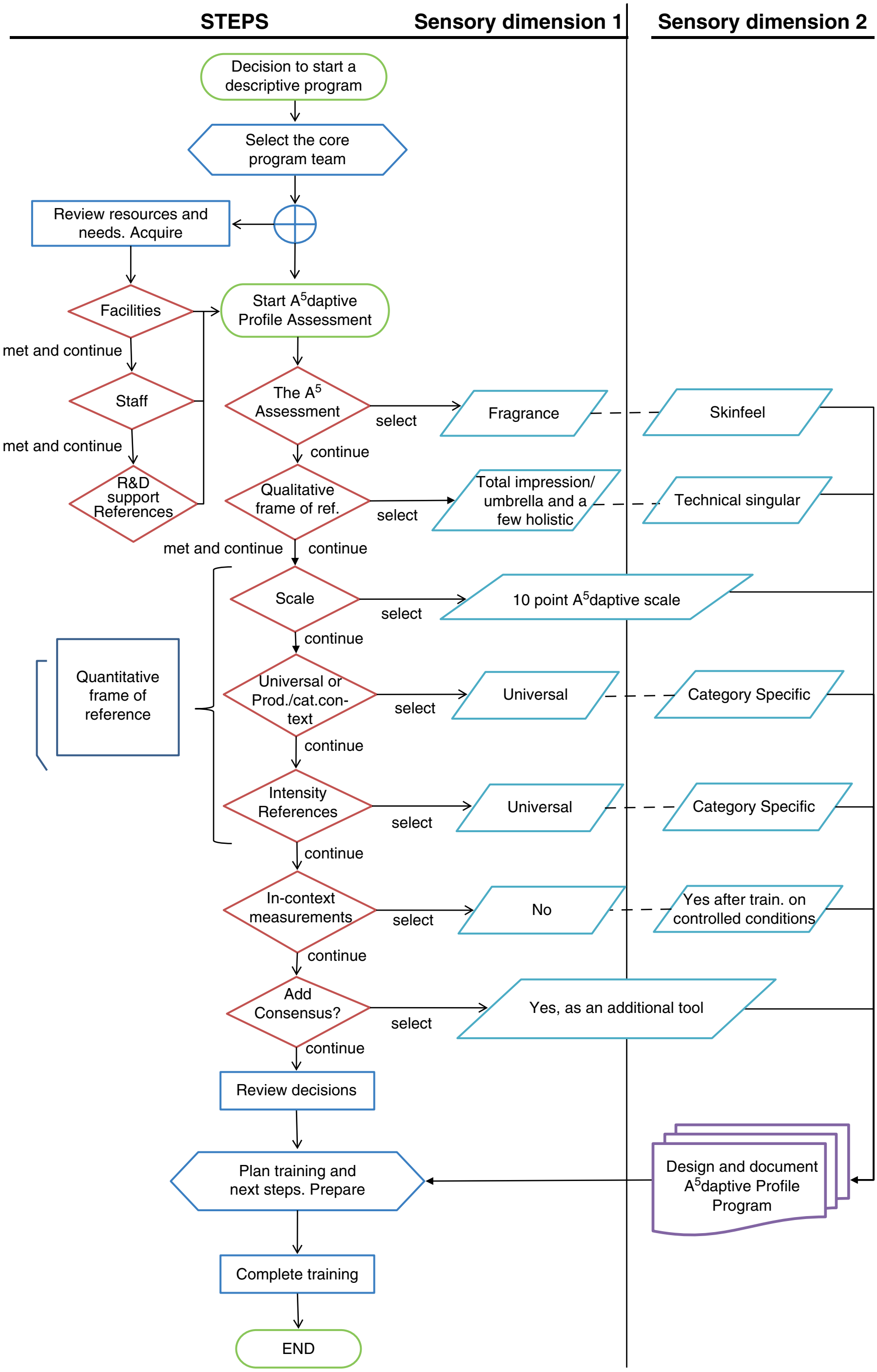


Figure 11.4 Flowchart of the A^{5}daptive Profile Assessment process (case study I).

11.7.1.1 Formation of the Core Programme Team

A meeting was held to determine the key members of the core programme (CP) team. The following professionals were invited to be part of this team: the A^5daptive Profile trainer, the sensory and consumer insights group director, an experienced sensory professional involved in the product categories' sensory projects, the panel leader being trained, the technician chosen to provide support in this programme, and two product developers/formulators in charge of the R&D work for the product categories of interest. The A^5daptive Profile trainer guided the discussion and offered a perspective on advantages and disadvantages of each option to ensure that the best decisions for the programme were made.

11.7.1.2 Review of Resources and Needs

Concurrent with the completion of the A^5daptive Profile Assessment, all other programme needs were evaluated vis-à-vis available and current resources. Facilities, staff and R&D support needs were assessed. The CP team determined that all requirements in these areas were met (see Figure 11.4).

11.7.1.3 A^5daptive Profile Assessment

In this phase the following was completed: the A^5 Assessment (to decide which of the five sensory dimensions to include in the programme) and all other assessments (of the qualitative and quantitative frames of reference, use/context evaluations and other product parameters and measurements), as described below.

- *A^5 Assessment*: in this assessment decisions were made regarding which of the five sensory dimensions were needed for the programme. The CP team decided that in addition to training the panel on skinfeel, a basic training programme on fragrance was needed. Therefore, based on the A^5 Assessment, the descriptive training programme was designed to incorporate these two sensory dimensions (see Figure 11.4).
- *Qualitative frame of reference Assessment*: the different lexicon options that the A^5daptive Profile Method recommends (see section 11.2.1.2) were reviewed by the CP team. It was decided that different lexicons were required for the skinfeel and fragrance evaluations. For skinfeel, technical singular attributes were needed. For fragrance, and given the general nature of necessary information, the panel was to be trained mainly on umbrella/total impression attributes (i.e. total citrus) and a few holistic measures (see Figure 11.4).
- *Quantitative frame of reference Assessment*: the three elements described in section 11.2.1.3 were assessed by the CP team to make decisions on the quantitative descriptive parameter options described below.
  - Scale. For the selected sensory dimensions, skinfeel and fragrance, the 10-point A^5daptive Profile Method's scale was chosen by the CP team as the main scale to be taught to and used by the panel (see Figure 11.4). In addition, and as typically done in this method's training, the panel was to be introduced to other scales during the training (see section 11.2.1.3.1).
  - Universal versus product/category-specific context. For fragrance evaluations, a universal frame of reference/context was chosen by the CP team.

This decision was made because the fragrance intensities of the company's products and prototypes, and of other commercial products, vary widely and are often strong. However, for skinfeel, a category-specific context was chosen by the CP team. For the company, the focus was only on lotions and creams and not on other personal care products (see Figure 11.4).

- ◦ Intensity/quantitative references. The decisions on context (i.e. universal versus product-specific context) dictated the type of intensity references to develop and use: for fragrance, a universal odour/fragrance intensity scale, while for skinfeel, product/category-specific intensity scales (see Figure 11.4).

- *Product use/in-context Assessment*: as described in section 11.2.1.4, often, and particularly in the area of personal care and household products, a thorough descriptive product characterization cannot be completed in booths. Guided by the A^{5}daptive Profile trainer, the CP team considered it important to add an additional phase in the skinfeel training programme, to be implemented after the panel had successfully worked on projects for about 18 months. In this phase, the procedures and protocols established and used under controlled conditions/booth evaluations would be adapted for in-context evaluations (see Figure 11.4).
- *Assessment of other product parameters and measurements*: the last element assessed by the CP team was the need for other measurements, and thus evaluations, to be completed by the panel. The CP team planned an additional phase to train the group on consensus. While the A^{5}daptive Profile panel was to complete individual evaluations on a regular basis, in some projects and for some product evaluations, consensus would be needed. Therefore, a module to train the panel on the process to reach consensus was to be conducted 10 months after the panel would have finished the core descriptive training (see Figure 11.4).

11.7.1.4 Design and Documentation of the A^{5}daptive Profile Method's Descriptive Programme

All decisions made by the CP team throughout the A^{5}daptive Profile Assessment were reviewed and discussed. The final A^{5}daptive Profile Method's descriptive programme was designed based on the decisions made, and all documentation was finalized (i.e. training programme design, protocols and procedures) (see Figure 11.4).

11.7.1.5 Implementation of the A^{5}daptive Profile Method's Descriptive Training Programme

Once all preparations were completed, the training programme was initiated. The recruitment and selection of candidates were completed first (see section 11.2.4.1), followed by the training programme (see section 11.2.4.2.2).

A successful A^{5}daptive Profile Method's programme was implemented. It was the ideal descriptive programme for the company, meeting all the expectations and providing the necessary product information. This success was possible due to the thorough assessment of conditions and options completed in theA^{5}daptive Profile Assessment, the optimal programme designed and the sound training completed.

11.7.2 Case Study II: Improvement of a Descriptive Programme: The A^{5}daptive Profile Method's Audit and Improvement Process to Address the Programme's Issues

A manufacturer of baked products used a descriptive programme for 2 years. The sensory and consumer insights group had encountered a few issues with the descriptive data, which raised some concerns. This case study presents the A^{5}daptive Profile Method's audit completed to unveil issues and provide solutions, and the implementation of improved practices.

In summary, the main issues that had been encountered were as follows.

- *Discrimination/sensitivity*: overall, low sensitivity had been observed in this panel's data. Often, no significant differences in attributes had been shown by the panel results, when in fact noticeable and perceivable differences existed among products.
- *Other panel measures*: low panel agreement and repeatability were other issues of concern found with the panel data.
- *Information provided to R&D*: the R&D groups had expressed some concerns regarding the descriptive data. In some projects, the data did not provide all the information required. Researchers had pointed out a few important attributes that were not being measured by the panel.
- *Motivation*: because of the issues encountered above, the panel was demotivated. Panel data and project results had been shared with the panel, and several attempts had been made to address the issues in panel maintenance sessions. However, the panel had experienced frustration and lack of motivation because the results had not improved, despite their and the panel leader's efforts.

Given the lack of success with previous efforts, the company's sensory and consumer insights group decided to consider the philosophy of the A^{5}daptive Profile Method, complete a programme review/audit, and implement any necessary improvements. The audit's process followed, information reviewed, issues unveiled, recommendations and implemented improvements are described below.

11.7.2.1 Review of the Descriptive Programme's Documentation

All programme documentation was reviewed and assessed with special attention to lexicons and ballots, quantitative references, panel data and reports, and selected products previously evaluated by the panel.

11.7.2.2 Descriptive Programme Elements Needing Improvement

The review of the descriptive programme's documentation unveiled several issues needing attention.

- *Area 1*: qualitative frame of reference/lexicons
- *Area 2*: quantitative frame of reference
- *Area 3*: pre-project preparation

For each of these areas, the issues unveiled and the A^{5}daptive Profile Method's practices recommended as solutions are presented below.

11.7.2.2.1 Area 1: Qualitative Frame of Reference/Lexicon Component Issues Encountered

The audit/review unveiled that several lexicons needed improvement, especially flavour lexicons, as per the standards and philosophy of the A^{5}daptive Profile Method.

For illustration purposes, Table 11.3 shows two examples of flavour lexicons generated and used by the panel prior to the audit. Improvements in the following elements were identified and reported to the sensory and consumer insights group.

- *Incomplete lexicons*: the review of representative products showed that more attributes (especially aromatics) were needed to fully characterize these products' flavour.
- *Inaccurate/technically incorrect terms*: additionally, it was determined that several attributes were not identified and thus not labelled accurately, based on the perceived flavour characteristics of the products. Examples included chocolate, buttery, etc. The evaluation of attributes not perceived in products, yet listed on ballots, creates problems in individual evaluations and might result in misleading conclusions.
- *Lack of total impression assessments*: the review indicated that the panel had not been trained to evaluate total impression attributes for certain key flavour families, as practised by the A^{5}daptive Profile Method (e.g. total wheat). Including total impression measures in the ballot, and thus grouping characteristics according to flavour families, has many advantages. Often, the lack of total impression measures may result in misleading product information and panel performance problems.

Other lexicons needing improvement were identified. However, only the two examples shown in Table 11.4 are discussed to limit the scope of this case study.

Table 11.3 Examples of flavour descriptive lexicons prior to the A^{5}daptive Profile Method's audit (case study II).

	Chocolate cake	Fruity cake
Aromatics		
	Toasted wheat	Toasted wheat
	Chocolate	Vanilla
	Milky	Pear
	Vanilla	Apple
	Other sweet aromatics	Strawberry
	Buttery	Cinnamon
	Nutty	Nutty
	Artificial	Milky
Basic tastes		
	Sweet	Sweet
	Salty	Sour
	Bitter	Salty

Table 11.4 Improved flavour lexicons based on the A^{5}daptive Profile Method's audit (case study II).

Chocolate cake	Fruity cake
Aromatics	
TOTAL SWEET AROMATICS	TOTAL FRUITINESS
Vanillin/vanilla	Apple/pear
Caramelized	Berry
Other	*Describe character*
Cocoa	Grape
TOTAL WHEAT	Other fruit
Raw	OTHER SWEET AROMATICS
Cooked	Vanillin/vanilla
Toasted	Caramelized
Dairy	Other
Describe character (NFDM, cooked milk, diacetyl, butyric, etc.)	TOTAL WHEAT
	Raw
Cooked egg	Cooked
Nutty/woody	Toasted
Oil aromatic	TOTAL BROWN SPICES
Fruity	Cinnamon
Describe type	Clove
Baking soda	Other
Other	Nutty/woody
	Dairy
	Describe character (NFDM, cooked milk, diacetyl, butyric, etc.)
	Oil aromatic
	Other
Basic tastes	
Sweetness	Sweetness
Saltiness	Sourness
Bitterness	Saltiness

NFDM, non-fat dry milk.

Recommendations and Solutions

The information shown in Table 11.4 was shared with the sensory and consumer insights group, summarizing solutions to the issues explained above. This table shows improved flavour terminology for the two products shown in Table 11.3, following the A^{5}daptive Profile Method's philosophy.These improved lexicons would address the issues identified as follows.

Recommendations Regarding the Incomplete Lexicons. Based on the review of ballots and the products themselves, the sensory and consumer insights group was informed that more terms than those being evaluated by the panel (see Table 11.3) were needed to fully characterize these products' flavour. As illustrated in Table 11.4, missing attributes included the following.

- Attributes associated with the base/wheat flavour. Processed grain products develop and display different characters of grain flavour, such as raw, cooked and toasted. Thus, the raw and the cooked grain flavour characteristics were missing in the current sensory terminology.
- Other sweet aromatics. Vanilla is not the only sweet aromatic perceived in cakes, and often might not even be the perceived character. In an improved lexicon (see Table 11.4), a family of sweet aromatics (i.e. the umbrella attribute: total sweet aromatics) should be included to capture the different attributes in this flavour family.
- Other attributes. Other missing attributes included dairy, baking soda, cooked egg and other fruity and brown spice notes.

These findings explained the R&D groups' concerns on missing attributes in the panel's descriptive evaluations.

Recommendations Regarding the Inaccurate/Technically Incorrect Terms. The A^{5}daptive Profile Method's audit also showed that there were a few flavour terms that were incorrectly identified, named and thus evaluated. Examples included (for the specific products discussed, see Table 11.3): chocolate, milky and buttery.

Table 11.4 shows solutions to these issues, such as.

- Instead of chocolate, the appropriate term for the perceived attribute in the particular 'chocolate' cake manufactured by the company is cocoa.
- A more appropriate term for the baked products' perceived dairy note is NFDM (non-fat dry milk) and not milky (a general term that mostly relates to cooked milk).
- No butter/baked butter aromatic is perceived in these cakes, so 'buttery' is an inappropriate term to describe this product's flavour character. An oily flavour note and other fat/butter-related flavour characteristics (e.g. butyric acid) are perceived.

Recommendations Regarding the Lack of Total Impression Assessments. As explained in section 11.2.1.2, the A^{5}daptive Profile Method incorporates holistic measures, when appropriate, to achieve a more complete descriptive evaluation of products. One relevant holistic measure is total flavour impression.

There are many advantages in grouping related flavour attributes under a total impression/umbrella measure. One of the main advantages is that the total impression assessment provides valuable information on the total flavour family's impact, despite individual assessors differing in the perception and scoring of the family's specific flavour notes. This is possible because in the A^{5}daptive Profile Method the total impression attribute is initially rated, and assessors are instructed to temporarily disregard the nuances or specifics of the flavours it encompasses. Once the total flavour impression attribute is scored, the specific flavours of that flavour family are evaluated.

The improved lexicons shown to the sensory and consumer insights group (see Table 11.4) illustrated these benefits and how the total impression attribute could be evaluated for several flavour families, such as sweet aromatics, wheat, fruitiness and brown spices. For example, with the ballot used prior to the audit, no valuable or conclusive information on fruitiness, an important attribute for the product, was obtained (Table 11.5). Assessors differed in the fruity character they identified, in that a few assessors perceived only pear or only apple or only strawberry, while others perceived any combination of these fruity notes. Thus, the data showed lack of agreement and repeatability in these terms and gave no conclusive information on the perceived fruitiness type.

Assessing total fruit impression first, as shown in the A^{5}daptive Profile Method's improved ballot (see Table 11.4), would allow assessors to rate the intensity of total fruit regardless of character first, and then rate the intensity of those specific fruity notes assessors perceive. This means that regardless of the individual differences in perceived fruity notes, assessors would be in agreement and reliable information would be obtained in the total fruit impression measure.

Table 11.5 shows the data obtained after the recommended changes were implemented. The data allowed gathering a score from all assessors (N = 14) for total fruitiness impression, which was not possible with the previous ballot. In addition, the improved structure of the ballots yielded more accurate and reliable data for fruitiness. The results showed that mainly apple/pear and berry notes were perceived in the product. All assessors rated apple/pear and two-thirds of the panel rated berry. The previous results did not provide a clear profile and a good indicator of the specific fruit characteristics and their

Table 11.5 Total impression measures in flavour lexicons, following the A^{5}daptive Profile Method's philosophy (case study II).

Attribute	Number of assessors with scores ≠ 0	Average intensity
Fruitiness evaluated prior to audit		
Pear	N = 7*	3.4
Apple	N = 7	6.2
Strawberry	N = 4	*not calculated with N = 4*
Fruitiness evaluated as per the A^{5}daptive Profile Method		
TOTAL FRUIT	N = 14	6.7
Apple/pear	N = 14	4.8
Berry	N = 10	3.2
(describe character)		*(N = 4. Strawberry)*
		(N = 6. Raspberry)
Grape	N = 6	*not calculated with N = 6*
Other fruit	N = 5	*not calculated with N = 5*
(describe character)		*(dry fruit, peach, apricot)*

*Panel size N = 14.

intensities. In addition, with the previous ballot, since only four assessors rated strawberry, no average could be calculated for this attribute.

Incorporating the total impression attribute for other flavour families offers the same advantages as described for fruitiness. Thus, the flavour and fragrance lexicons developed in the A^{5}daptive Profile Method always incorporate the evaluation of total impression for many flavour or fragrance families, as needed.

Discussion on the Effect of Qualitative Issues on Data. Additional feedback was given to the sensory and consumer insights group regarding the effect of flawed terminology (qualitative issues) on the quantitative information/data. It was noted that the issues in the panel data would be addressed by implementing the improvements:

- in the qualitative frame of reference (i.e. the lexicons), and
- in the quantitative frame of reference explained in Area 2.

11.7.2.2.2 Area 2: Quantitative Frame of Reference Component Issues Encountered

Many project/product evaluations yielding very few or no significant differences among samples were identified in the audit of all the projects and data chosen for the review. It was observed that the scores occupied a very narrow range across products evaluated in most descriptive projects.

As an example, Table 11.6 was reviewed with the sensory and consumer insights group. This table illustrates that only a few significant differences were

Table 11.6 Texture descriptive results of a pair evaluation prior to the A^{5}daptive Profile Method's audit (case study II).

Attribute	Sample # 845	Sample #291
First bite		
Firmness	4.5	4.8
Cohesiveness	2.3	2.9
First chew		
Denseness	3.1	3.2
Chew down		
Moisture absorption	12.4	13.0
Cohesiveness of mass	10.2	11.4
Roughness of mass	4.7	4.8
Moistness of mass	5.8	6.0
Residual		
Loose particles	5.2 a	3.1 b
Toothpack	7.3 b	9.4 a
Mouthcoat	3.6 b	4.1 a

Means followed by a different letter are significantly different at the 0.05 significance level (Fisher's LSD).

found in the texture evaluation of two products of one of the categories shown in Table 11.3 (fruity cakes) in a project conducted prior to the audit. The samples were reviewed vis-à-vis these results. It was confirmed that noticeable differences could be perceived in these products, especially in firmness, cohesiveness, denseness, moisture absorption and cohesiveness of the mass. However, the panel's data failed to identify significant differences in these key texture characteristics for baked products.

Other product evaluations yielding very few or no significant differences were also identified. However, only the above example is discussed to limit the scope of this case study.

Two issues were discovered in thoroughly reviewing the fruit cake descriptive data.

- In some of the above texture attributes, the individual scores across all assessors for the two samples were very similar (low sensitivity). While slight differences were shown, assessors' scores for each pair of products should have shown a larger difference in range than those shown by the data.
- Low agreement and repeatability were noticed in some of these attributes' data, explaining why the ANOVA failed to show significant differences.

Recommendations and Solutions

The audit included review not only of the data and products but of the quantitative references used by the panel, which were universal references. It was determined that the use of these universal scales was one of the main reasons for the lack of sensitivity/significant differences in some of these key attributes.

Therefore, following the A^{5}daptive Profile Method's philosophy of developing/customizing quantitative references to address the needs of the user and the products' characteristics, recommendations were provided on how several of the texture universal quantitative references could be modified. Instead of using the broad texture universal scales, new and improved quantitative/intensity scales should be developed and used for the company's products. These scales should be product category focused, but broad enough to capture the texture space occupied by the relevant company's product categories.

Table 11.7 illustrates the suggested modification for two of the key texture attributes' intensity scales in baked products presented to the sensory and consumer insights group. Universal quantitative scales used in the evaluation of product attributes occupying a narrow intensity range yield the above problems. Therefore, these problems are addressed by adapting universal quantitative scales as per the A^{5}daptive Profile Method's philosophy and as shown in Table 11.7.

11.7.2.2.3 Area 3: Pre-Project Preparation Component

Issues Encountered

The descriptive protocols were also reviewed in the audit, showing that panel preparation sessions prior to project work were only scheduled when a new product category was evaluated. These sessions were conducted to develop the ballot.

Table 11.7 Customized/modified quantitative texture scales, following the A^{5}daptive Profile Method's philosophy (case study II).

	Universal scale (Muñoz 1986)*		A^{5}daptive Profile Method universal-adapted (for Baked Goods Co.)	
	Scale value	**Product**	**Scale value**	**Product**
Firmness scales				
	1.0	Cream cheese	1.0	Cream cheese
	2.5	Egg white	4.7	Pastry MF432LS
	4.5	American cheese	7.2	Cake MF67AX
	6.0	Olive	10.0	Olive
	7.0	Frankfurter		
	9.5	Peanut		
	11.0	Almond		
	14.5	Hard candy		
Denseness scales				
	0.5	Whipped topping	0.5	Whipped topping
	2.5	Marshmallow topping	2.9	Meringue cookie
	4.0	Nougat	5.8	Cake MF127TB
	6.0	Malted Milk Ball	9.1	Nougat
	9.0	Frankfurter	11.0	Malted Milk Ball
	13.0	Fruit jellies		

*Selected information is presented from tables published in Muñoz (1986).

However, for most projects, no panel sessions were conducted prior to the evaluations. The rationale for this practice was that if ballots and evaluation procedures had been previously developed, they did not require any modifications for the project evaluations. This was another improvement area identified.

Recommendations and Solutions

Guidance was provided regarding the activities that should be completed prior to project evaluations. As indicated in section 11.2.5.2, the A^{5}daptive Profile Method recommends that panel sessions be conducted prior to projects, even in studies involving product categories for which ballots and procedures have been developed.

Recommended panel activities in pre-project sessions were provided for both scenarios: with and without previous work/experience with a product category (see section 11.2.5.2).

11.7.2.3 Implementation of Improvements

A plan was developed to implement the recommendations for the three main problem areas unveiled in the descriptive programme.

Panel sessions were conducted to retrain the panel in the following elements of the A^{5}daptive Profile Method.

11.7.2.3.1 Retraining on the Qualitative Frame of Reference/Lexicons

- Review of the characteristics of A^{5}daptive Profile Method's lexicons and the process to generate improved sensory terminology.
- Review of previous issues and the development and use of improved lexicons (see Tables 11.4 and 11.5).

11.7.2.3.2 Retraining on the Quantitative Frame of Reference: Customized Intensity Reference Scales

- Review of previous issues
- Review of the rationale behind and teaching of the A^{5}daptive Profile Method's customized intensity reference scales
- Completion of product evaluations using the customized intensity reference scales

Practice sessions were designed for the panel to apply and practise the new concepts learned. The data generated in these practice sessions were summarized and analysed to assess improved results.

11.7.2.3.3 Implementation of Pre-Project Activities

The recommended pre-project panel activities to be conducted in future projects were also reviewed with the panel.

In addition, one-on-one sessions were conducted with the panel leader to review the recommendations on pre-project activities, panel issues discussed above, improved practices and panel maintenance activities.

11.7.2.4 Review of Improved Data and End of Audit

The data from the practice sessions conducted upon completion of the retraining sessions were assessed. They showed that the main issues encountered in both the previously used qualitative (lexicons) and quantitative frames of reference were addressed. The improved descriptive practices yielded:

- complete lexicons providing the thorough characterization of the products' singular and umbrella/holistic flavour and texture attributes
- accurate/technically correct terms in the descriptive terminology
- data showing realistic (in magnitude) and significant product differences
- high panel agreement and repeatability.

This information was shared with assessors to demonstrate the value of these improvements and to recognize their efforts.

11.8 Summary

This method was developed after years of practising the original profile methods and other modified/ derivative profile methods and uncovering some elements that could be further advanced. Thus, the A^{5}daptive Profile Method provides effective new recommendations to sensory practitioners involved in descriptive analysis and aids in solving some issues/concerns encountered in certain technical descriptive programmes.

The main characteristics of the A^{5}daptive Profile Method presented in this chapter were:

1) A technical descriptive method which has adapted several of the solid foundations of the profile methodology.

2) An approach with the core philosophy that descriptive techniques should be custom-designed and adapted for every application to be able to implement a best-in-class descriptive programme.

3) A comprehensive and holistic method, which incorporates, as needed, additional attributes (e.g. overall impact) or evaluation procedures (e.g. in-context) to provide the complete/holistic characterization of products.

A summary of these main characteristics follows.

1) The A^{5}daptive Profile Method is a technical descriptive method, which has adapted several of the flavour and texture profile methods' philosophical principles. This characteristic makes the A^{5}daptive Profile Method a descriptive approach which provides comprehensive technical qualitative and sound quantitative product information.

Qualitatively, this method's strengths arise from developing and using highly technical sensory lexicons which offers many advantages in the use of descriptive data, such as clear product guidance to product developers, effective interpretation of consumer data, etc. In addition, and only unique to this method, A^{5}daptive Profile panels acquire special skills for the development of sensory language, such as:

- ability to assess the best lexicon type/combination of elements among options learned: singular attributes, singular and umbrella/total impression attributes, and the two previous options with holistic attributes (e.g. fullness, blendedness, etc.)
- ability to assess the ballot structure to determine the best order of terms to avoid duplication and decrease panel variability.

Quantitatively, this method's strengths arise from its:

- adaptable scale, which is expanded for higher intensities, and
- use of quantitative/intensity reference scales, which are customized/adapted, avoiding problems encountered with set and mostly universal intensity reference scales used in other descriptive methods.

2) Custom-designed approach and techniques (Adaptive approach)

As shown throughout this chapter, the A^{5}daptive Profile Method includes, as a key initial step, the thorough assessment of the company's needs, applications, product categories and all circumstances to design the best-in-class unique/customized elements of the descriptive programme. These elements include the types of lexicon categories to be taught to the panel (e.g. technical/singular, holistic attributes, etc.), the scoring options and the adaptable scale to be provided to the panel, the customized intensity references and how they are adapted for the products and applications in the programme, the best product evaluation context to use (e.g. universal/product specific, traditional and/or in-use product evaluations, etc.), and the inclusion of consensus and when it should be used by the panel (by itself or in combination with traditional data collection approaches).

3) Comprehensive/holistic characterization
The A^{5}daptive Profile Method applies a holistic product evaluation approach.

- A holistic/comprehensive sensory evaluation (i.e. the evaluation of all sensory dimensions perceived by the five senses relevant or of interest to the user).
- The evaluation of holistic/complex attributes that, together with individual attributes, provide the most complete product sensory characterization (e.g. blendedness, fullness, overall impact, etc.).
- The holistic product experience and evaluation by combining traditional with in-use product evaluation practices for the evaluation of products in context, when needed.

11.9 Future Directions

11.9.1 Development of New and Improved Descriptive Methodology

This chapter presented the characteristics of the A^{5}daptive Profile Method, which is a relatively new traditional descriptive method. As discussed, it was developed to address issues encountered with several traditional descriptive practices and to provide a more comprehensive/holistic product evaluation approach by incorporating additional evaluation dimensions (e.g. holistic attributes, in-context evaluations, etc.). Thus new ideas were developed by this method which contributed to advancing the field of descriptive analysis. Similarly, other researchers are encouraged to contribute to the growth of this field, by developing other new ideas and suggestions.

11.9.2 Application of the A^{5}daptive Profile Method's Techniques

More sensory practitioners, involved in descriptive analysis, are encouraged to implement several of the A^{5}daptive Profile Method's suggestions to improve their current programmes. Of special value are this method's unique approaches in the following areas.

11.9.2.1 Customized Programme Design and Techniques

The core philosophical principle of the A^{5}daptive Profile Method is the assessment of all the parameters, limitations, company characteristics and preferences to adapt approaches and custom design the best-in-class descriptive programmes. Often, a descriptive method/philosophy is established disregarding the elements above, which leads to a suboptimal programme.

Practitioners are encouraged to complete similar assessments as done by this method, to design and adapt/modify techniques as needed in implementing an optimal programme or improving an existing one.

11.9.2.2 Optimizing the Programme's Qualitative Component

Based on the A^{5}daptive Profile Method techniques, sensory practitioners are encouraged to consider the use of:

- additional qualitative elements besides singular attributes in product evaluations (e.g. overall impression measures, complex terms, etc.)
- technical terms that provide notable benefits in the use of descriptive data for research guidance, consumer data interpretation, etc.

11.9.2.3 Optimizing the Programme's Quantitative Component

Based on the A^{5}daptive Profile Method techniques, sensory practitioners are encouraged to consider:

- the use of the suggested adaptable scale that is expanded to higher intensities
- the development and use of customized intensity reference scales
- diverse scoring techniques provided to panels, such as the use of consensus and several other scoring techniques besides scaling.

11.9.2.4 Holistic/More Complete and Relevant Product Evaluations

Sensory practitioners are encouraged to consider the philosophy of the A^{5}daptive Profile Method regarding a holistic approach in descriptive product evaluations for the benefits provided.

First, the evaluation of products in a controlled/traditional way and only focusing on singular attributes might not always provide a complete and comprehensive characterization of products. Second, the lack of product differences shown by the descriptive results (when differences exist) might be the result of many factors, such as insufficient panel training, high product or panel variability, environmental and other external factors influencing product integrity, etc. However, when none of the above applies, the A^{5}daptive Profile Method considers that the lack of differences shown by results might be the lack of relevant/holistic product information.

Therefore, sensory practitioners are encouraged to adopt the holistic philosophy inherent in this method, and consider in-context evaluations and the assessment of holistic measures such as overall impression attributes, complex holistic measurements (e.g. blendedness, fullness, etc.), as needed.

11.9.2.5 New and Challenging Applications

The A^{5}daptive Profile Method has been used in many food and non-food applications. Many challenges exist in using descriptive techniques in new applications, such as in novel/innovative products, and particularly in non-food products due to the limitation of evaluation sites (e.g. in skinfeel evaluations), the elaborate set-ups (e.g. in some household, automotive products, etc.), or fatiguing effects (e.g. in fragrances), etc.

This method's versatility lends itself to the successful use of its techniques in any new/complex application. Sensory practitioners are encouraged to incorporate the A^{5}daptive Profile Method's new ideas/suggestions to successfully develop and adapt traditional techniques to address new and challenging product evaluation scenarios.

References

ASTM International (American Society for Testing and Materials) (2009) *Physical Requirements Guidelines for Sensory Evaluation Laboratories*, 2nd edn. Kuesten, C. & Kruse, L. (eds.). West Conshohocken: ASTM.

ASTM International (American Society for Testing and Materials) (2011) *Standard Guide - E 1490-11. Two Sensory Descriptive Analysis Methods for Skin Creams and Lotions*.West Conshohocken: ASTM.

ASTM International (American Society for Testing and Materials) (2017) *WK8435. Standard Guide for Measuring and Tracking Sensory Descriptive Panel and Assessor Performance*. West Conshohocken: ASTM.

Brandt, M.A., Skinner, E.Z. & Coleman, J.A. (1963) The texture profile method. *Journal of Food Science*, **28**, 404–409.

Cairncross, S.E. & Sjöström, L.B. (1950) Flavor profiles – a new approach to flavor problems. *Food Technology*, **4**, 308–311.

Caul, J.F. (1957) The profile method of flavor analysis. *Advances in Food Research*, **7**(1), 1–40.

Cortez-Pereira, C.S., Baby, A.R., Kaneko. T.M. & Robles Velasco, M.V. (2009) Sensory approach to measure fragrance intensity on the skin. *Journal of Sensory Studies*, **24**(6), 871–901.

Gacula, M.C. (1997) *Descriptive Sensory Analysis in Practice*. Connecticut: Food and Nutrition Press.

Gacula, M.C. & Singh, J. (1984) *Statistical Methods in Food and Consumer Research*. New York: Academic Press.

Griffiths, P. & Kulke, T. (2002) Clothing movement – visual sensory evaluation and its correlation to fabric properties. *Journal of Sensory Studies*, **17**(3), 229–255.

Hough, G., Contarini, A. & Muñoz, A.M. (1994) Training a texture profile panel and constructing standard rating scales in Argentina. *Journal of Texture Studies*, **25**, 45–57.

Hunter, E.A. & Muir, D.D. (1995) A comparison of two multivariate methods for the analysis of sensory profile data. *Journal of Sensory Studies*, **10**, 89–104.

Ishihara, S. (1969) *Ishihara's Test for Color-Blindness*. Tokyo: Kanehara Shuppan Co.

Kaplan, S. & Okur, A. (2008) The meaning and importance of clothing comfort: a case study for Turkey. *Journal of Sensory Studies*, **23**(5), 688–706.

Keane, P. (1992) *The flavor profile. In: ASTM Manual Series MNL 13. Manual on Descriptive Analysis Testing*. West Conshohocken: ASTM.

Lawless, H.T. & Heymann, H. (1998) *Sensory Evaluation of Food. Principles and Practices*. London: Chapman and Hall.

McEwan, J.A., Hunter, E.A., van Gemert, L.J. & Lea, P. (2002) Proficiency testing for sensory profile panels: measuring panel performance. *Food Quality and Preference*, **13**. 181–190.

Moskowitz, H.R., Muñoz, A.M. & Gacula, M. (2003) *Viewpoints and Controversies in Sensory Science and Consumer Product Testing*. Connecticut: Food and Nutrition Press.

Muñoz, A.M. (1986) Development and application of texture reference scales. *Journal of Sensory Studies*, **1**, 55–83.

Muñoz, A.M. & Bleibaum, R.N. (2001) Fundamental Descriptive Analysis Techniques. Exploration of Their Origins, Differences and Controversies. Workshop presented at 2001: A Sense Odyssey, Fourth Pangborn Sensory Science Symposium, Dijon, France.

Muñoz, A.M. & Civille, G.V. (1998) Universal, product and attribute specific scaling and the development of common lexicons in descriptive analysis. *Journal of Sensory Studies*, **13**, 57–75.

Muñoz, A.M., Szczesniak, A.S., Einstein, M.A. & Schwartz, N.O. (1992) The texture profile. In: *ASTM Manual Series MNL 13. Manual on Descriptive Analysis Testing*. West Conshohocken: ASTM.

Næs, T. & Solheim, R. (1991) Detection and interpretation of variation within and between assessors in sensory profiling. *Journal of Sensory Studies*, **6**, 159–177.

Næs, T., Brockhoff, P.B. & Tomic, O. (2010) *Statistics for Sensory and Consumer Science*. Chichester: John Wiley & Sons Ltd.

Schifferstein, H.N.J. (2010) From salad to bowl: the role of sensory analysis in product experience research. *Food Quality & Preference*, **21**(8), 1059–1067.

Sinesio, F., Risvik, E. & Rizidbotten, M. (1990) Evaluation of panelist performance in descriptive profiling of rancid sausages: a multivariate study. *Journal of Sensory Studies*, **5**, 33–52.

Stone, H., Bleibaum, R.N. & Thomas, H.A. (2012) *Sensory Evaluation Practices*. New York: Academic Press.

Szczesniak, A.S., Brandt, M.A. & Friedman, H.H. (1963) Development of standard rating scales for mechanical parameters of texture and correlation between the objective and sensory methods of texture evaluation. *Journal of Food Science*, **28**(4), 397–403.

CHAPTER 12
Ranking and Rank-Rating

Graham Cleaver

12.1 Introduction

12.1.1 Overview

In sensory evaluation we rarely have the objective to characterize a single product in isolation. In general, we need to position each product in relation to the other products in the set and quantify their positions relative to one another on the chosen sensory attributes. The most common practice here, conventional descriptive analysis, is to present each product to the assessors one at a time according to a standardized protocol. The assessors are trained to rate each product on the selected attributes as absolute scores, often in relation to reference points on the scale as appropriate. This standardizes the process across the assessors and promotes reproducibility of scoring.

It is clear, though, that human perception is more comparative in nature (Lawless & Heymann 2010). Consequently, we perform better when assessing products in relation to one another rather than in absolute terms and where there is less reliance on the memory of products assessed previously. Even those of us without perfect pitch, for example, could identify which of two consecutive notes played on a piano was pitched higher, but would struggle as the time interval between the notes was increased and other notes were interspersed.

The motivation for ranking methodology is to exploit this natural mode of comparative perception and to present the products together so that they can be assessed against each other. In this way, the assessors can assess all of the products and, where appropriate, reassess them as required, so that they can place them in increasing order in terms of perceived intensity on the chosen sensory attribute, or some hedonic measure in the case of a consumer study. The benefits of allowing the assessors to freely reassess products will depend on the product category but have been clearly demonstrated (Kim & O'Mahony 1998) in terms of reducing discrimination errors and increasing the ability to discriminate between the products.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

As a potential additional step in the assessment process, it may be appropriate to specify that the assessors should also rate the products on a standard scale. This is known as rank-rating.

12.1.2 Aim

The aim of ranking (and rank-rating) as a sensory methodology is to compare a set of products in terms of their perceived intensities on a set of defined sensory attributes. Compared to conventional descriptive analysis methodologies (e.g. QDA), the primary aim of ranking is to obtain the most sensitive comparison of the relative attribute intensities across the products, rather than mean scores along a predefined intensity scale. The extension of rank-rating is intended, to some extent, to bridge that gap.

12.1.3 Process

The process, described in some detail in section 12.3, is designed to be simple: for each attribute separately in turn, the assessor samples and resamples all of the products and then places them in increasing order of perceived intensity on that attribute. Whilst the procedure is inherently simple, it is not suited to all areas of application and guidelines for consideration are provided in section 12.7.

12.2 History and Background

Ranking has been used as a test methodology from the early days of psychophysics (Thurstone 1931) and applied as a core sensory methodology since that time. Along with the rise in application of standardized sensory methods in academia and industry, particularly from the 1950s onwards, ranking has been most widely applied in the context of foods research and development. However, the potential constraints inherent in the ranking process, discussed in section 12.7, mean that ranking methods can also be particularly suited to some non-foods applications such as ranking the smooth feel of swatches of fabric.

Traditional methods in sensory and consumer research have largely been based on methodologies that require a set of people to assess a range of products and give each a score based on a chosen scale, for example a set of ordered categories or a continuous analogue scale of fixed length. For controlled sensory testing with trained assessors using different variants of descriptive analysis, steps are taken in the recruitment and training so that the assessors can do this as consistently as possible. For studies with untrained consumers, this standardization process is neither appropriate nor practical and the strategy adopted can be to restrict the evaluation to liking only or to limit the inclusion of additional scoring of sensory attributes. The chief advantage of these scoring-based methodologies comes through the modelling of the resulting data and the ability, through appropriate experimental design and mixed-model analysis of variance, to separate out and quantify sources of variation and provide effective comparisons between products.

However, reliance on the assessor's memory required to give repeatable absolute scores of intensity has meant that there has been continuing interest in alternatives. Also, any analysis based on calculating mean values of product scores carries an implicit assumption that the scale points are equally spaced in perceptual terms. Ranking makes no such assumptions and so, in combination with its potential for increased sensitivity, it can be very appealing as a method to consider for application whenever appropriate.

It should be noted that in its simplest form, ranking of two products only, the method is referred to as 'paired comparison' or as a 'two-alternative forced choice test (2AFC)'. The term 'ranking' is normally applied when ranking three or more products, but clearly there are parallels with these two-product variants in terms of applicability and interpretation.

12.3 Methodology

One of the potential advantages of a ranking methodology is that the process of training of assessors is relatively short compared to that required for the various forms of descriptive analysis. These descriptive analysis methodologies require different levels of training in consistent use of the rating scale and usually in relation to one or more reference points along the scale.

Essentially, assessors performing a ranking methodology need to understand the task and have a common agreed interpretation of the sensory attributes involved.

12.3.1 Selection of Assessors

For any sensory or consumer ranking study, the primary selection criteria apply as for other types of test methodology, taking care to exclude assessors with health-related conditions or hypersensitivity due to allergic reaction to the products on test.

For an internal sensory panel, the assessors should have an appropriate level of sensory acuity. For a foods panel, for example, screening tests may be used based on the ability to recognize basic tastes and accurately compare model solutions with varying levels of, for example, sucrose, sodium chloride or citric acid.

For a consumer study with naive subjects, no selection based on sensory acuity would normally be carried out, with the main objective being to recruit subjects after screening on health issues and according to a predefined quota based on, for example, gender, age and current product usage.

12.3.2 Training of Assessors

To explain the process of ranking to untrained assessors for the first time, it may help to run a training exercise with simple stimuli to ensure that they understand the ranking process. There is a chance to be creative here in terms of possible training exercises and some examples are shown below.

12.3.2.1 Step 1

For example, shapes of varying size (Figure 12.1) could be prepared as separate physical stimuli on pieces of card and labelled in the same way as products would be. A form would be provided with instructions for the ranking task to place them in increasing order of size.

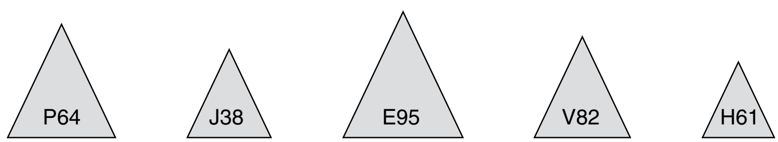


Figure 12.1 Shape stimuli for training.

12.3.2.2 Step 2

The next stage could be to provide objects of the same size and varying in weight (Figure 12.2), the task being to assess each in turn, and then reassess them as required until the assessor can place them in order of increasing weight from left to right.

Identical containers with varying contents could be used for this purpose. This weight example could also be used to illustrate the difference between two forms of the ranking task. As described above, with free re-evaluation of the stimuli as often as required, the assessors should be able to rank the objects in increasing order of weight even when the differences are quite small. In the other form of a ranking task where each stimulus has to be assessed once only in a defined order, many of the judgements required to place them in rank order will be based on holding a memory of the weights of multiple objects and the test will be much less sensitive. This could lead to discussion on the different number of stimuli that would be appropriate with each form of the ranking process.

Figure 12.2 Weight stimuli for training.

12.3.2.3 Step 3

The next step could be a ranking exercise based on stimuli appropriate for the planned ranking evaluation. For a tactile ranking task, ranking of surfaces with varying degrees of roughness could be used, for example graded sandpaper. For a study on foods or drinks then model solutions for selected basic tastes, with increasing concentrations in water of sucrose for sweetness, sodium chloride for saltiness, etc., could be used (Figure 12.3). The task here for the assessors is to sample each of the solutions, initially in a predefined order, and then to retaste them until they can place them in increasing order according to the designated attribute.

Figure 12.3 Model solutions for training.

In all of the above training examples, there is a correct ordering of the products and so the rankings provided by the assessors can be compared against this predefined order as a check on how well they are performing. Each of these tasks can be made progressively more challenging by reducing the spacing between adjacent stimuli.

12.3.3 Attribute Selection and Definition

The selection of appropriate attributes for a ranking study is critical, particularly as the number of attributes has a direct impact on possible sensory fatigue and time/resource required to complete the study. Very often, the key attributes will be known based on the specific objectives of the study or as established drivers of consumer liking. Where this is not the case, some preliminary qualitative exercise using an appropriate product set as stimuli could be designed to elicit and group potential attributes and create a prioritized selection of a suitably small number, typically six or less. As with descriptive analysis, it is essential that each of the attributes is clearly defined and the interpretation is understood by each of the assessors. Particular attention must be given to any complex attributes. Further details here are given in section 12.7.1.

12.3.4 Selection of Products

As with any sensory study, the selection of the products to be included is critical. This is very much linked to the overall aim of the study, the hypotheses to be tested and the choice of experimental design. Whilst the aim of the study itself may be to quantify the size of the sensory differences between the products, it helps to have some view on the likely size of difference during the design process through exploiting existing product knowledge or carrying out some preliminary testing. If the products are all very similar to one another then the planned study may be too small to pick up any differences as significant and lead to inconclusive results. Conversely, if the differences are very large then a formal sensory test may not really be required and the resources better used elsewhere.

Where the aim of the study is to quantify and model the relationship between perceived sensory intensity of a given attribute and an appropriate instrumental measure (e.g. sensory in-mouth thickness versus measured viscosity, based on varying levels of thickening agent) then guiding the choice of products towards equal perceptual spacing on thickness will increase the efficiency of the design. In this case, a linear relationship may be expected between sensory thickness and log (viscosity) and this may be used to help determine the optimum product set.

With a series of ranking studies carried out over time, unless they have been designed according to a coherent experimental design, it will not be possible to carry out an overall analysis of the combined data. Inclusion of a fixed reference product in every study would, however, help put the results in a common context and enable a qualitative review of the results across the studies.

12.3.5 Presentation of Products

The products as presented to the assessors for ranking are labelled with a code to enable the products to be identified by the panel leader but disguise the true identity from the assessors. With ranking, it is particularly important that these codes themselves do not imply any prior ordering of the products which might influence how the assessor would place them in order.

For example, single digit numbers 1, 2, 3, 4, 5… should never be used for this purpose. Similarly, single letters should be avoided because of the alphabetical ordering. A frequently used system of coding is three-digit codes (e.g. 547, 912, 652, etc.) to disguise the products. Another option, designed to distract attention from either numerical or alphabetical ordering, is to choose alphanumeric codes, for example one letter and a two-digit number as in section 12.3.2.

Where the same assessors are required to rank the same set of products on different occasions, a different set of codes is required for each ranking session to avoid any problems with the assessors remembering their previous ordering of the products. Ideally, separate codes would be used for each attribute to be ranked to ensure independence of the rankings but this is unlikely to be practical.

For clarity of interpretation, these codes are not used for presentation and reporting purposes, and meaningful codes or product names should be used instead.

12.3.6 Ranking Direction

This deserves special mention as it can be a source of confusion at both the assessment stage and in the reporting/presentation of the findings. It is quite common to think of the product that is ranked first as the one which is highest in intensity. However, if the ranks are assigned numbers, say 1 to 6, then it may be natural to associate the highest number with the product which is highest in intensity. In the sensory community, both 'rank 1 = lowest' and 'rank 1 = highest' are commonly used, and examples of both are used in the case studies in section 12.8.

The potential confusion can be avoided by:

- making the ranking direction absolutely clear on the form or data entry screen, and the physical ordering of the products from left to right during the ranking process
- reinforcing this during assessor training
- being absolutely clear about how the data are recorded in terms of assigning 1 to the lowest or highest ranked product, and developing a standard practice for this for a given study type
- clarifying on every presentation chart and analysis output table how the numbering of the ranks corresponds to increasing attribute intensity.

12.3.7 Ties

When ranking a set of products on a given attribute, an assessor may feel that they are unable to separate out two or more products into a defined order. In general, it is best to avoid ties in ranking and to use a forced-choice procedure to specify that the assessors make their best effort to give each of the products a separate ranking even when they are not too sure. This is mainly to avoid the possibility that an assessor, particularly one who is less experienced, might assign a tied rank as a way of making the task easier. Where it is felt that tied ranks should be allowed then products tying for second and third places, for example, can each be assigned a rank of 2.5 in the data and the analysis adjusted accordingly.

With rank-rating, one option is to insist that products are given separate ranks but allow the assessor to give equal ratings to reflect that they were very close in perceived intensity.

12.3.8 Assessment Protocol

An example of a questionnaire for a ranking task based on tasting a selection of products is shown in Figure 12.4.

Very much depending on the product category and the constraints of testing, there are many variations on the methodology and these will need to be taken into account in the instructions for the assessors.

- If sensory fatigue or carry-over from one product to the next are limiting factors then it may be necessary that each product is assessed only once in the defined order. It may be that ranking is less suited to this situation and conventional descriptive analysis methodology is preferred – see section12.7.3.

You are provided with the following samples of orange juice labelled as:

H64	L31	P19	M75	T28

Please taste them initially in this order and then taste and re-taste them as you wish until you can place them in front of you from left to right in order of increasing sweetness. Between each tasting, please cleanse your palate by taking a sip of the water provided. When you have done this, write the sample codes in the boxes below:

Lowest	Rank order of sweetness ➔			Highest

Figure 12.4 Example questionnaire for ranking task.

- It may be necessary to ensure that assessors cleanse their palate between each product. For example, products which leave a coating in the mouth require palate cleansing whereas cleaner-tasting products may not. Ranking of non-food items such as smoothness of hair switches does not require a cleansing step between evaluations, unless the switch was depositing something onto the fingers. Each new study design needs consideration based on the specific constraints, though working towards establishing best practice over time based on experience.
- In some test situations or cultures, it may be suspected that providing a specific assessment order is taken as a cue or starting point for how the products should be ranked. In that case, the methodology can be adapted so that the products are placed randomly in front of the assessor, not in a straight line, with instructions to assess and reassess each product but without specifying the order of first assessment.
- It may be that the products are too large or cumbersome for assessors to move them around in order to finalize their rankings. Or it may be thought advisable to impose some structure on the ranking process, for example when using children as assessors. In these situations, it may help to instruct the assessors to first pick the product that they would rate as highest on the chosen attribute, and then, from the remaining products, pick out the next highest, and so on.
- In a consumer study with a single task based on ranking products in terms of liking, it may be very informative to capture each consumer's reasons for the chosen rank ordering. Whilst it may not be possible to ask for this additional information on every product assessed, the most practical option may be to request that the consumer makes comments only on the products ranked highest and lowest.

12.3.9 Rank-Rating Methodology

One of the possible limitations of ranking methodology is that it does not allow the assessor to express a measure of the size of difference in intensity between products being ranked and ranking alone does not provide overall scores for each product that can be positioned along a predetermined reference scale.

In fact, we can derive a quantitative measure for each product along an attribute intensity dimension which in turn gives an overall measure of their separation from one another, but these measures on their own do not indicate how high or low the products score. All of the products could be relatively 'low' in intensity or 'high' in intensity – we cannot tell from the ranking results alone. Including a reference product can help in providing a context for the rankings but adaptation of the methodology, as described here, can also be considered.

Traditional descriptive sensory analysis with monadic presentation is designed to provide absolute scores for product comparisons but relies on the assessors' ability to make judgements of attribute intensity based on memory and specifically their memory of the perceived intensity of products included in the training process and in relation to intensities corresponding to reference points along the scale.

To exploit the potential advantages of ranking in terms of reducing dependence on memory for the assessor to reliably position products against one another and in combination with a task that involves scoring of intensity, the rank-rating procedure (Kim & O'Mahony 1998) was introduced. The aim here is to combine the extra sensitivity of ranking from side-by-side assessment of products and free retasting (where possible) together with a rating task to provide the product scores along a predetermined scale.

For rank-rating, the assessors are asked to rank the products in increasing order of attribute intensity as before, and then to rate each product on an intensity scale. This process works best if the assessors are able to physically move the products around in front of them during the ranking process until the final ordering is achieved.

In principle, any type of scale can be used. For example, it could be an ordered unlabelled category scale (Figure 12.5) or it could be a continuous line scale marked at the extremes to indicate 'none' on the left and 'extremely high' on the right.

For a paper-based exercise, it may be convenient to construct a large version of a rating scale that is positioned in front of each assessor. The assessor can then position each product against the chosen score, moving each of them around as required until the task is complete.

Alternatively, if the evaluation is carried out with on-screen presentation of the task and direct computer entry of the response then the data entry screen would be created via the chosen sensory management software. The interface would be designed with a representation of the scale and the ability to move

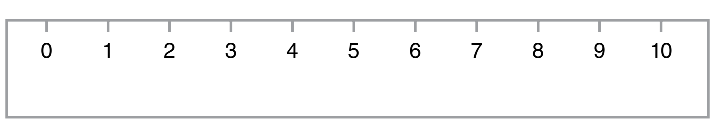


Figure 12.5 Example of rank-rating intensity scale.

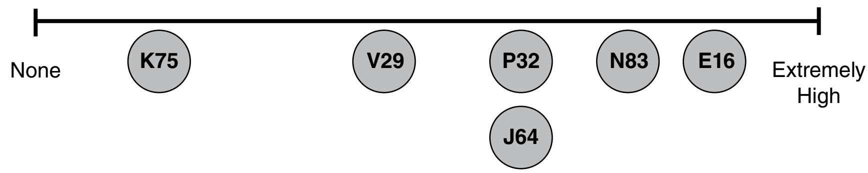


Figure 12.6 Rank-rating intensity scale with positioning of products.

icons, one for each product, around during the evaluation phase until the final set of positions is achieved (Figure 12.6).

Whatever form of scale is used, it is important that the assessors are familiar with the scale and the definition of any reference points or category descriptors.

To consider the sensory features of rank-rating, it is convenient to take the piano analogy from section 12.1.1 further. During the ranking phase, the assessor is tasked, without sight of the keyboard, to place a series of notes in increasing order of pitch and can request that the notes are replayed as required and in any sequence until a final ordering is achieved.

For the rating task, the keyboard serves as the rating scale so that the set of stimuli rated as positions along the scale reflect the subject's view of the pitch levels of the notes and the spacing between them. If the subject has been trained and is able, within the limits of perceptual variation, to position a note correctly against the right piano key, without the luxury of being able to hit the key to check, then the rating procedure is providing absolute scores that could be compared from one exercise to the next. If, on the other hand, the subject treats the keyboard in the rating phase simply as a scale from low to high pitch then the chosen positions and separation between them will have a relative interpretation along the low to high continuum, of one selected note relative to another in the set.

Extensive training, as in standard descriptive analysis, would be required to achieve absolute scoring in order that the subject can match the perception of a stimulus against the memory of reference points on the scale. With rank-rating, the additional challenge would be to ignore the influences of the other stimuli that it is being ranked against while doing so.

Pragmatically, one may take the view that the extra sensitivity provided by the ranking phase in terms of more reliable relative positioning will benefit the subsequent scoring, but caution needs to be applied when comparing the results from two different rank-rating studies in absolute terms.

Rank-rating is a flexible methodology and this analogy may help to highlight the choice between two extremes described below.

The scale could be simply structured as in Figures 12.5 and 12.6 and little additional training would be required for the assessors other than familiarization with the scale. Rank-rating was developed originally for this type of testing and most applications have used this format. This has the merit of keeping the task simple, in the spirit of ranking generally, minimizing errors in scoring by allowing side-by-side assessment and resampling, and focusing the assessors' attention on spreading the product scores over the scale to reflect the relative spacing between them. As well as in-laboratory sensory testing, this format would probably be chosen for studies with untrained consumers and hedonic assessment. It should be clear, though, that this test format is appropriate for one-off studies but not where there is a requirement to make comparisons between products in different studies in terms of absolute scores. Some limited comparison could be made in terms of differences relative to a common fixed reference product.

Alternatively, the assessors could be trained in the same way as for a sensory descriptive analysis methodology with the expectation that they would all score products on the scale in the same way and enable comparison of products across studies in absolute terms. Some caution needs to be applied here, as follows.

- This does require several training sessions with exposure to a range of products in order that a common scoring process is established (instilling a memory of 'notes along the scale').
- Compared to standard descriptive analysis, where each product is scored independently, the motivation for including the prior ranking phase is that placing the products in order first can help the assessor to position the products relative to one another on the scale. The aim here is that the resulting set of scores along the intensity scale will best match the assessor's perception of the set of products as a whole. This does mean, however, that the scoring of a given product will potentially be influenced by the particular set of products against which it is being compared. Despite the best intentions of the prior training, sensory contrast effects (Lawless & Heymann 2010) come into play, with relative scoring heavily influencing aspiration for scoring in absolute terms. This has particular relevance for incomplete block studies where the set of products presented together will vary from one session to another and potentially from one assessor to another within the same session.
- Again in relation to standard descriptive analysis where many sensory attributes can be scored with each product presented separately, it should be clear that rank-rating is suited to situations where there are only a small number of attributes to be assessed.

12.4 Experimental Design

12.4.1 Complete Block Design

When the order of first assessment is to be specified, the designated ordering should be determined from a balanced randomized design. When each subject ranks all products, it is convenient to use a Latin square for this design and a variant of this design is a Williams Latin square (Wakeling & MacFie 1995) which will ensure balance in terms of order and the number of times each product directly follows each of the other products.

Table 12.1 Example of design balanced for order effect and carry-over.

Assessor	Product no: order of first assessment					
	Order 1	Order 2	Order 3	Order 4	Order 5	Order 6
1	2	5	4	6	1	3
2	5	6	2	3	4	1
3	3	1	6	4	5	2
4	6	3	5	1	2	4
5	1	4	3	2	6	5
6	4	2	1	5	3	6

An example is shown in Table 12.1 for six products and six assessors, with each product assessed first in each order position once and each product following directly after each of the other products by exactly one.

Replicates of this design, rerandomizing the codes 1–6, can be used to create a design for six products and 12, 18, 24 assessors and so on. With 15 assessors, say, then two complete replicates and one half-replicate of this design can be used. Note that this is the same design process that would be used in a conventional descriptive analysis study. With free resampling (retasting in the case of food), the order of first assessment should be less critical than in sequential monadic studies.

With ranking studies implemented through sensory management software (for example, Compusense®, Fizz®), the specification of the first assessment order would be implemented as part of the design process, and a new randomization used for each attribute to be ranked to ensure independence of the rankings.

12.4.2 Incomplete Block Design

Ranking-based methods are traditionally used where the assessor can reliably rank all of the products against one another in a single task. Where the total number of products exceeds the number that can be ranked in a single session then an incomplete block design can be considered, provided that an appropriate experimental design is used and the necessary analysis software is available. A simple example is shown in Table 12.2, with seven products in all but with each assessor ranking four products. This is an example of an 'incomplete block' test design.

In this particular case, a standard balanced incomplete block design (BIB) is available for sets of seven assessors, each of whom assesses four out of the complete set of seven products. Although lists of available designs are available (Wakeling & Buck 2001), it is more convenient now to use specialist design of experiments software (for example, JMP®, XLSTAT®) based on the total number of products, number of assessors and number of products that can be ranked within the same task (block size).

Table 12.2 Example of a balanced incomplete block design.

Assessor	Products to be assessed			
	Order 1	Order 2	Order 3	Order 4
1	7	2	4	6
2	3	1	6	4
3	5	6	3	2
4	1	4	2	5
5	4	5	7	3
6	2	3	1	7
7	6	7	5	1

This design (Table 12.2) has the properties that:

- each of the seven products is assessed the same number of times (four)
- each of the seven products is assessed alongside each of the other six products (i.e. occurs in the same set of four products) the same number of times (twice)
- each product occurs in each order position exactly once.

With this set of desirable features, randomized replicates of this design can be used to construct a design for the study as a whole. In the table data for case study 2 (section 12.8.2), two replicates of the design are used for 14 assessors in each of two sessions.

With incomplete designs, it is always desirable to balance the design as much as possible but in practice, for a given number of products and block size, in most cases it will not be possible to achieve perfect balance. If complete balance is not achievable then a partially balanced completed block design may be used. We would always want to try to ensure that each product is assessed the same number of times overall, but the number of times each product pair occurs together in the same set may vary somewhat.

The example here is designed to show the principle of incomplete blocks. In practice, larger designs with several sets of rankings per assessor are required to produce a robust overall analysis.

Use of incomplete block designs extends the range of ranking-based methodologies and also into applications with more strategic studies with larger numbers of products, but it should be clear that more specialist analysis methods are required, as illustrated in section 12.8.2. It may also be advisable to simulate some data before running the study using the proposed design and check that the data can be analysed using the planned analysis method.

12.4.3 Size of Study and Replication

As with sensory descriptive analysis methods, it is advisable to build up knowledge over time of the sensitivity of a given panel to detect typical product differences as statistically significant and create rules of thumb for required replication in future studies. As a starting point for discussion and to provide guidance in a new application area where historical data are not available, we can say that:

- a ranking study for a complete block design should be more sensitive to detect true differences compared to a standard descriptive analysis methodology on the same product set, so typical study designs with 12 assessors and two or three replicates of each ranking by each assessor should suffice to detect typical differences deemed important to the researcher. A similar replication would be required for a rank-rating study
- for practical reasons across all descriptive analysis methods, an overall replication of, say, 36 reps per product is achieved by a design with, for example, 12 assessors performing the same task three times on separate occasions. Provided the number of assessors is not too small, so that the performance of the panel can be unduly influenced by how one or two particular assessors are performing, or that the results become very specific to the particular chosen set of assessors, then this practice can work well. In addition, the replicated nature of the design opens up more possibilities for checks on assessor performance, illustrated in section 12.8.1.1
- with consumer studies, the focus may be more on understanding the differences between consumers in the pattern of liking across the products rather than simply averaging over consumers, so the number of consumers, usually performing each task only once, will be much larger, typically 100–200
- with incomplete block designs, the attention should be directed to the number of times each product pair occurs within the same block. Typically, this should be at least 12 for an effective design.

12.5 Analysis of Ranking Data

In this section, the analysis of ranking data by three different methods is illustrated using a small data set with 12 assessors each ranking four products (Table 12.3). Somewhat larger studies are included in the case studies in section 12.8.

The assessor ID here is the unique code for that assessor used for every study in which he/she participates. This is the code which will appear in any output relating to assessor performance to identify the assessor to the sensory panel leader but avoiding use of actual names or initials.

Three analysis methods for ranking data are presented in this section and these are briefly summarized here.

- The Friedman test provides an overall test of whether there are significant differences between the products and, if so, which pairwise comparisons of products differ significantly.
- The R-Index analysis does not give an overall test of significance between products, but focuses on the pairwise comparison between products and expresses the difference in a more perceptually relevant and meaningful way.
- The parametric analysis gives an overall test of differences between products and a measure of relative intensity for each product along a continuous

Table 12.3 Sample set of ranking data (4 = high).

Assessor	Assessor ID	Prod A K87	Prod B F43	Prod C H91	Prod D W26
1	A3652	4	2	3	1
2	A2154	3	1	2	4
3	A5321	1	3	4	2
4	A7529	3	2	4	1
5	A9231	4	2	3	1
6	A7492	4	1	3	2
7	A6170	3	2	4	1
8	A3187	2	3	4	1
9	A9072	4	2	3	1
10	A2850	3	2	4	1
11	A1170	4	1	3	2
12	A5702	3	2	4	1
Sum Of Ranks		38	23	41	18

underlying scale. This measure also provides pairwise differences between products which can be expressed in a statistic similar to the R-Index. This analysis is particularly suited to studies with large numbers of products and can be used with incomplete block designs. It is, however, more complicated to implement in terms of requiring specialized software.

It is a useful first step in the analysis to present graphically the results of the number of times each product was ranked in each position, here as a stacked histogram. For this diagram (Figure 12.7), the products are shown arranged in order based on the sum of ranks in the table above. This provides a visual indication of clear differences between the products and the expectation that the analysis should reveal statistically significant differences between them.

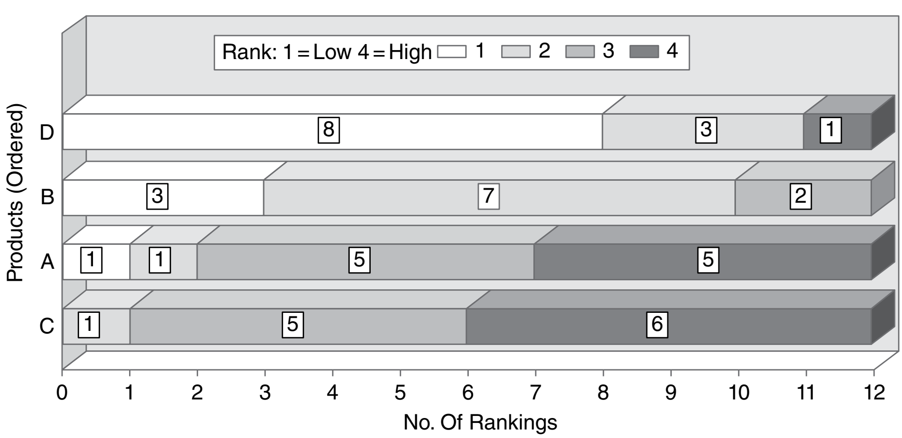


Figure 12.7 Stacked bar chart of ranks for each product.

12.5.1 Friedman Analysis

The most commonly applied statistical test for ranking data is the Friedman two-way analysis of variance which is widely described in the sensory and statistical literature (Siegel & Castellan 1998). The calculations are straightforward and can be performed by hand,as illustrated below, but the procedure is included in many statistical packages, for example XLSTAT®, and this will be the likely preferred method of analysis.

The first step for a manual analysis is to perform an overall significance test of whether (a) the products are different from one another ('alternative hypothesis') or (b) they are all the same ('null hypothesis').

First, we calculate the sum of the squares of the rank totals from Table 12.3.

$$\sum R^2 = \left(38^2 + 23^2 + 41^2 + 18^2\right) = 3978$$

And then substitute this value in the formula:

$$T = 12\sum R^2/(bk(k+1)) - 3b(k+1)$$

where k = number of products (4) and b = number of blocks (in this case 12 assessors). For this data set, $T = 12 \times 3978 / \left(12 \times 4 \times 5\right) - 3 \times 12 \times 5 = 18.9$

For small data sets, this value can be compared against the tabulated values in Appendix 12.1 to provide the overall test of significant difference between the products. Referring to this table, we see that value of T = 11.1 would correspond to significance at P = 0.01, so we may conclude that the products are significantly different.

For larger data sets, the T value may be treated as a chi-square statistic with k-1 degrees of freedom and the significance checked either by reference to tables or direct calculation using, for example, the CHIDIST function in Excel®.

If this overall test proves statistically significant, as here, then the products can be compared against one another as the second step in the analysis process.

If the chosen significance level for these comparisons in 5%, then the 'least significant difference' is calculated using this formula:

$$LSD = t\sqrt{\left(bk\left(k+1\right)/6\right)}$$

where t is the Student's t statistic with (b-1)(k-1) degrees of freedom. For this data set, $11 \times 3 = 33$ df, t = 2.035, so $LSD = 2.035\sqrt{((12 \times 4 \times 5)/6)} = 12.87$.

Pairs of products with sums of ranks which differ by more than this amount are significantly different at the 5% level.

In this example, we would conclude that products D and B did not differ significantly at the 5% level, and similarly products A and C, but the other pairwise comparisons were statistically significant, as summarized in Table 12.4, with products underlined where there is no significant difference between them.

Table 12.4 Rank totals for each product and significance test between products.

Rank totals for each product **Increasing intensity →**				**Least significant difference**
D	B	A	C	
18	23	38	41	12.87
....		....		

As with all statistical tests involving pairwise comparisons between multiple products, there is a possibility of some differences appearing as significant by chance as a consequence of the large number of significance tests being performed. To protect against this, some adjustment can be made in the pairwise significance tests but this is best done using statistical software rather than hand calculations. For example, XLSTAT® uses the Nemenyi procedure (Nemenyi 1963) to perform these multiple comparison tests.

It should be noted here that an alternative to the Friedman test calculations is to perform a standard analysis of variance on the rank data (Hobbs 2009) treating the data as though they were continuous scores.

In practice, this analysis of variance will give similar results to the Friedman test of the same data, for two reasons.

- The Friedman test is essentially comparing the products in terms of the mean ranks, as a standard analysis of variance would do.
- Analysis of variance is fairly robust against departure from the assumption of normality associated with ranked data rather than continuous scores.

In fact, one standard practice for analysing very non-normal data, for example skewed data with extreme outliers, is to convert the data to ranks within each block and apply the analysis of variance to the ranked data. In ranking studies the primary data are already in rank form.

One advantage is that software packages offer a variety of multiple comparison procedures for standard analysis of variance whereas the choice of options for a Friedman analysis will be more limited. It must be stressed, however, that this analysis of variance option only applies to complete block ranking studies; with incomplete blocks, a ranking of any particular product in a block is specific to the products that it has been ranked against.

12.5.2 R-Index Analysis

The R-Index is a measure of perceptual difference between two products that was developed in the field of psychophysics and has a basis in signal detection theory. It is particularly suited to the situation where there is a high level of replication of each task by each assessor. In psychophysics studies, it is not uncommon, for example, for each ranking task to be replicated 10 times by each assessor. Applied in the context of ranking (O’Mahony 1992), the R-Index is a measure of the per-

Table 12.5 Sample set of data reformatted by ranking position.

Rep	Low.....Ranking.....High			
	1st	2nd	3rd	4th
1	D	B	C	A
2	B	C	A	D
3	A	D	B	C
4	D	B	A	C
5	D	B	C	A
6	B	D	C	A
7	D	B	A	C
8	D	A	B	C
9	D	B	C	A
10	D	B	A	C
11	B	D	C	A
12	D	B	A	C

ceptual difference calculated between each pair of products in turn as the probability that one product would be ranked higher than the other. The calculated R-Index varies between 0% and 100% with 50% representing equality.

As an example, consider the results of 12 separate rankings of four products (Table 12.5) which uses the same data as in Table 12.4 but structured by ranking position.

There are two types of R-Index (O'Mahony 1992) which are described in detail in this influential paper in the context of ranking. The first is the 'John Brown' R-Index R_{JB} (Brown 1974) which, taking the example of product A versus product B, is calculated simply as the proportion of times that product A is ranked higher than product B, in this case 10 out of 12 so $R_{JB} = 83.3\%$. Note that product C is ranked higher than product B in all 12 replicates so that $R_{JB} = 100\%$ for the C versus B comparison.

The statistical significance of the difference between each pair of products can be determined by using a binomial test (Siegel & Castellan 1998) of whether R_{JB} is significantly different from 50. For A versus B, 10 out of 12 gives $P = 0.038$ with a two-sided significance test so we can conclude that the difference is significantly different ($P < 0.05$).

We can express the results as a two-way table (Table 12.6) in terms of the percentage of comparisons where Product 1 is ranked higher than Product 2, taking the opportunity to order the products for clarity.

The second type of R-Index, R_{MAT}, is based on the theoretical comparisons of each of the ranking positions of the first product against each of the ranking positions of the second product.

Focusing on the R_{MAT} calculation for A versus B, this is based on the 144 (12×12) theoretical comparisons of the 12 reps of A against the 12 reps of B.

Table 12.6 Pairwise R_{JB} values and statistical significance.

Product 1	Product 2			
	D	B	A	C
D	–	25.0 ns	16.7*	8.3 **
B	75.0 ns	–	16.7*	0.0 ***
A	83.3 *	83.3 *	–	50.0 ns
C	91.7 **	100.0 ***	50.0 ns	–

Note that some of these comparisons, 12 in this example, are made within the same ranking set but most, 132 in this example, are made between ranking positions in different sets.

In each of these 144 comparisons, if the ranking position for A is higher than the ranking position for B then this is scored as '1'. If A is lower than B this is scored as '0'. If the ranking position is the same then this is scored as '½'. R_{MAT} is then calculated as the total of these scores expressed as a percentage of the number of theoretical comparisons. This is most easily calculated when the results are summarized as count of the number of times each product appears in each ranking position (Table 12.7).

Table 12.7 Ranking distribution for each product.

Product	Low...Ranking ...High			
	1st	2nd	3rd	4th
A	1	1	5	5
B	3	7	2	0
C	0	1	5	6
D	8	3	0	1
Total	12	12	12	12

For these data, the calculation for A versus B is:

$$\text{Total score} = 1\text{x}(3\text{x}½) + 1\text{x}(3 + 7\text{x}½) + 5\text{x}(3 + 7 + 2\text{x}½) + 5\text{x}(3 + 7 + 2)$$

$$\text{Total score} = 1.5 + 6.5 + 55 + 60 = 123$$

$$R_{MAT} = 100\text{x}(123/144) = 85.4\%$$

The significance of each pairwise comparison can be determined by using a Mann-Whitney U test with correction for ties (Siegel & Castellan 1998) since the R-Index is calculated here in the same way as the U statistic.

Table 12.8 Pairwise R_{MAT} values and statistical significance.

Product	Product			
	D	B	A	C
D	–	30.2 ns	12.5 **	7.3 ***
B	70.0 ns	–	14.6 **	7.3 ***
A	87.5 **	85.4 **	–	43.8 ns
C	92.7 ***	92.7 ***	56.3 ns	–

Table 12.8 shows pairwise R_{MAT} values and significance tests.

Note that the two types of R-Index do not give identical results and it important to recognize the difference between the two. The simple R_{JB} statistic for A versus B is determined solely from the relative positions of A and B within each ranking set and is independent of how the other products C and D are ranked, either against one another or in relation to A and B.

The R_{MAT} statistic is based on ranking positions, rather than relative rankings within a set, and so the calculation for A versus B is influenced by the ranking positions of C and D across the data set.

The R_{MAT} statistic has been more widely used in the sensory community than the R_{JB} statistic but both deserve consideration in the context of the analysis of ranking studies. In choosing between the two, the question comes down to whether the R-Index is strengthened or weakened by including comparisons between ranking positions across different ranking sets. Where these ranking sets come from the same assessor and the aim is to calculate an R_{MAT} value for each assessor before combining them, as in applications in the psychophysics community, the practice has stronger alignment with signal detection theory (O'Mahony 1992) and is well-founded. Where the ranking sets come from different assessors who are potentially different from one another in terms of underlying perceptual ordering of the products, the practice is more questionable and the simpler R_{JB} statistic may be preferred. One other consideration here may be that the R_{JB} statistic can be applied to incomplete block studies (see section 12.4.2), whereas the R_{MAT} statistic cannot.

As noted earlier, it is clear that both types of R-Index analysis focus on the pairwise comparison of products and do not provide an overall test of differences between the products.

12.5.3 Parametric Analysis

A flexible and efficient way of analysing ranking data is to fit a parametric model to the data (Levitt 1975) which estimates a parameter for each of the products such that the estimated values represent relative positions of the products along an underlying intensity scale of the particular sensory attribute. There is a parallel here with the analysis of paired comparisons where a set of products are compared two at a time in a designed experiment and where the task for the participant is to specify which is preferred or is higher on a specified attribute. The data are conventionally analysed with a Bradley–Terry model (David 1988) to provide an overall ordering and positioning of the products along a preference or attribute continuum. Since paired comparisons can be viewed as a special case of ranking in incomplete blocks, namely with block size 2, the parametric analysis for ranking data is a generalization of the Bradley–Terry type of modelling.

Whereas software is readily available for the previous two forms of analysis or it could be easily programmed in Excel, a more sophisticated analysis package is required for this type of analysis. SAS®, PROC PHREG (proportional hazards regression) can be used for this purpose (Allison & Christakis 1994), setting up the data set with one line per ranking and coding each product as a combination of dummy (0/1) variables, i.e. when Product = 1 P1 = 1 and P2–P4 are all set as 0, as illustrated in Table 12.9 for the first two rankings.

Table 12.9 Data format for PHREG analysis.

Assessor	Product	Dummy variables				Rank 1 = high
		P1	P2	P3	P4	
1	A	1	0	0	0	1
1	B	0	1	0	0	3
1	C	0	0	1	0	2
1	D	0	0	0	1	4
2	A	1	0	0	0	2
2	B	0	1	0	0	4
2	C	0	0	1	0	3
2	D	0	0	0	1	1

In fact, with k products the model only requires that k-1 dummy variables are used with any one of the products omitted. The parameter for the omitted product is set to 0 and conventionally the parameters are then mean-centred on 0, with the result that the final estimates are the same whatever parameterization is used. With the data set created in this form, the analysis can be run with the following code:

```
PROC PHREG DATA=dataset;
  STRATA ASSESSOR;
  MODEL RANK=P1-P4;
  ESTIMATE 'A vs B' P1 1 P2-1;
RUN;
```

Similar ESTIMATE lines are included for other pairwise product comparisons. The analysis first provides an overall test for differences between the products as a likelihood ratio statistic, in this case 23.18, which, treated as a chi-square test with 3 df, is significant ($P < 0.0001$). The output parameter estimates for each product and the significance of the pairwise comparisons at the chosen significance level (typically 5%) can be represented as shown in Figure 12.8, positioned along a line corresponding to increasing attribute intensity and with products which are not statistically significant underlined by the same line.

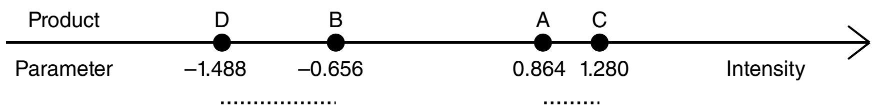


Figure 12.8 Summary of analysis from parametric model.

The SAS code is included here to illustrate how to fit this type of model to rank data but other statistical packages could be used to fit the same model.

An early example of the application of this type of model (Levitt 1973) shows how it can be applied in the food industry with an incomplete blocks study.

The scaling of the final intensity scores may seem somewhat arbitrary, but a feature of this analysis is that the estimated parameters on this scale for each product can be transformed, for each product pair, into estimates that one product would be ranked higher than the other. If the estimated parameters for products i and j are p_i and p_j then the estimated probability that product i is ranked higher than product j is $\exp(p_i)/((\exp(p_i)+\exp(p_j))$. The pairwise values are shown in Table 12.10.

Table 12.10 Pairwise probabilities that product (row) is ranked higher than product (column).

Product	Product			
	D	**B**	**A**	**C**
D	–	0.30	0.09	0.06
B	0.70	–	0.18	0.13
A	0.91	0.82	–	0.40
C	0.94	0.87	0.60	–

In sensory studies, and particularly those with a large number of products, often there is an objective to correlate the sensory results with those from other measures on the same set of products, such as instrumental measurements or consumer liking. The attribute intensity values from the parametric analysis can be used for this purpose. With complete block studies, it will be found that the parameter estimates will correlate strongly with simple mean ranking scores and not so much is gained. The practice has real benefit when incomplete block designs have been used to enable the inclusion of a larger number of products in a single study and mean ranking scores cannot be used.

Although requiring more specialized software, parametric analysis of ranking data has a number of appealing features.

- It provides an overall test of differences between the products as well as pairwise significance tests between them, as for the Friedman test.
- The estimated parameters can be converted to a table of pairwise estimates of the probability that product x would be ranked higher than product y, to correspond to the type of analysis produced by the R-Index approach.
- The model can be extended to account for and test for differences due to additional structure in the data, for example gender and age of the participants (Allison & Christakis 1994).
- Tied ranking can be easily accommodated in the model, at least in the SAS PHREG implementation, simply by using equal values for the tied observations and specifying TIES = EXACT in the model.
- This method of handling of ties opens up possibilities for application to other types of methodologies. For example, a consumer study could be designed so that eight product options are presented together and the task for each participant is to specify which three he/she would choose and to rank those three in order of preference. For the analysis model, this ranking of the chosen three options could be input together with a tied ranking of 4 for the products that were not chosen.

12.6 Analysis of Rank-Rating Data

Conventionally, rank-rating data are analysed by the same type of analysis of variance used for descriptive analysis (described in Chapter 5). However, particular attention needs to be given here to the assumption of independence of the scores and the associated error of each score in the analysis of variance. In conventional descriptive analysis with sequential monadic presentation of the products, the process is designed to encourage the assessors to be able to give scores which reflect the absolute intensity of the specific product they are evaluating, and scoring in relation to their memory of the intensity of reference products and corresponding marks on the scale. The extent to which they can do this reliably and provide consistent scores of absolute intensity will be limited by their memory of the intensity of those reference products.

In contrast, with rank-rating the process itself is designed to be comparative so the ranking and scoring of a particular product will be dependent on the set of products that it is being compared against. This brings clear advantages in the ranking stage, removing the dependence on memory of products assessed some time previously in determining the relative position of products to one another, but, as noted in section 12.3.9, it raises the question of whether the subsequent rating can be interpreted as an absolute measure of intensity.

This is one example of a context effect present in all types of sensory assessment and illustrated clearly (Lawless & Heymann 2010) in a study comparing soups with varying levels of salt addition. When a product with a medium level of salt was assessed against two other products with lower levels, it was scored much higher than when it was assessed against two products with higher levels of salt.

This effect may be seen even with monadic presentation of products but is likely to be enhanced by the rank-rating process and presenting all of the products together. The extent to which this effect could be controlled for rank-rating by extensive prior training of assessors would need to be checked before treating the scores as absolute measures of perceived intensity.

Interpretation of the nature of the scores therefore should be the prime consideration in interpreting the output of the analysis of rank-rating studies. The mean scores can be used for comparison across the products included within a given study but caution is necessary when comparing the mean scores of products across separate studies since this would require an assumption of absolute scoring.

12.7 Applications

12.7.1 Test Situations Well Suited to Ranking

Where ranking is an appropriate methodology, it has a number of clear potential advantages over alternative methodologies.

- It is a more intuitive task and consequently less rigorous training for the assessors is required compared to descriptive analysis.
- Consequently, there will be a quicker turn-round in delivery of test results.
- Particularly where the product differences are small, the test is potentially more sensitive than descriptive analysis and so lower replication may be required to achieve the same power to detect differences.
- Ranking is independent of scale whereas descriptive analysis, in calculating mean values, relies on an implicit assumption of equal perceptual spacing between the scale points.

It is useful here to highlight the features of different test situations that will be particularly suited to using a ranking methodology and, conversely, applications for which it is not suited.

12.7.1.1 Free Reassessment (Free Retasting)

Ranking will perform best in the situation where the products to be ranked can be assessed and reassessed repeatedly (free retasting in the case of food, with or without palate cleansing, as appropriate) until the subject can place them in rank order. For some product types, this will not be possible, for example a confectionery product that needs to be sucked to elicit the specific attribute of interest.

12.7.1.2 Ability to Move Products Around

The process of coming to a final ordering of the products will be greatly facilitated if the assessors can move the products physically in front of them until they can place them in rank order.

12.7.1.3 Unidimensional Attribute

Most attributes, like ‘saltiness’ of potato crisps for example, may be considered to be clearly defined along a single dimension of underlying intensity and should present no problems to the assessors during the explanation of the task and attribute definition. Care needs to be taken, though, in the case of more complex attributes. For example, suppose the project objective was to compare a range of desserts in terms of ‘creaminess’. The desserts could vary in terms of creamy appearance, creamy flavour or creamy texture. The product that was highest on one of these may not be highest on another. Without clear definition of the particular aspect of creaminess of interest, the danger is that the assessors will use their own criteria for the ranking, leading to confusing results.

For consumer studies, we expect that ‘liking’ will be a complex attribute but a key objective of the study. In addition, therefore, to a top-line assessment of aggregate overall liking, one objective may be to explore this complexity through modelling individual patterns of preference and identifying underlying consumer segmentation.

12.7.1.4 Limited Number of Attributes

Each attribute to be ranked requires a separate evaluation of each of the products in the set and, in the case of food, consumption of samples from each of these to enable the ranking. Clearly, the number of attributes to be ranked is likely to be limited here by constraints on time and sensory fatigue. Spreading the attributes over more than one session may help to reduce sensory fatigue but the process is likely to be time and resource intensive. It should also be noted that a fresh sample of each product may be required for each attribute ranking and to ensure that the rankings of consecutive attributes are independent of one another. Consequently, considerably greater quantities of each product may be required than for a comparable descriptive analysis methodology with monadic presentation and all attributes scored on the same sample of each product.

12.7.1.5 Limited Number of Products

In general, ranking works best in the situation where the number of products to be compared is within the number that can be ranked by the assessors in each ranking evaluation. This is known as a ‘complete block’ design. Typically, up to six products may be ranked in a single evaluation though this maximum will vary according to the specific constraints of the product type.

It is possible to use a ranking methodology with large numbers of products to be compared and using a statistical design whereby each time the assessors rank a subset of the products. These are called 'incomplete block' designs – an example is shown in section 12.8.2.

12.7.1.6 Stand-Alone Exercise

Ranking is usually used in the situation where the main aim is to provide a sensitive comparison between the particular set of products included in that study, and not to compare those products against external measures or combine the results with other studies. If different ranking studies are to be combined to create an overall ordering of a larger set of products then the studies should be designed with this in mind. This would not be possible just by having a single reference product common to all studies, for example. Instead, a statistical design should be used to permit sufficient overlap between the studies to enable an efficient combined analysis. In this way, it would be possible for a well-designed set of ranking studies to provide a sensitive overall comparison of the complete set of products.

12.7.2 Test Situations Not So Well Suited to Ranking

It is important to recognize the limitations of ranking and that it is not suited to all test situations. Some of these are implied by failure to meet the criteria in the previous section, but other considerations are listed below.

12.7.2.1 Sensory Fatigue

Sensory fatigue limits the number of products that can be assessed or restricts the process of free retasting in the case of food; consequently this will put constraints on the test design. For example, where the products differ in terms of strength and character of aroma then the acuity of the senses may soon be dulled through adaptation.

12.7.2.2 Amount of Product Consumed

As with all sensory studies on food and drinks, the total amount of product consumed in a single session and over sessions on the same day can be a primary consideration. The products may contain specific ingredients for which there is a maximum level cleared for daily intake and ethical clearance for the study must be obtained by designing the study within that constraint. Ranking of foods on multiple attributes will generally require consumption of a greater quantity of each product than a comparable descriptive analysis procedure.

12.7.2.3 Carry-over

Particularly in the case of food, there may be some carry-over effects from one product to the next. For example, the product may leave a residue coating in the mouth that would influence the perception of the next product to be tasted. This may necessitate palate cleansing between each assessment as would be standard practice in this situation in a descriptive analysis methodology. This restriction may limit the free retasting process in a ranking exercise.

12.7.2.4 Strong Aromas

As well as the adaptation issue mentioned above, strong aromas from foods or personal products may create confusion between products when presented together as a set for a ranking exercise, rather than separately for conventional descriptive analysis.

12.7.2.5 Product Changes During Assessment

It is important to remember that the product itself may change during the time taken to complete the ranking. A set of hot beverages will tend to cool down during assessment and a set of ice creams, for example, will tend to warm up and even start to melt. The temperature at which a product is assessed can influence its perception so it is important to keep this under control. Insulated serving containers can help here, but it will be important to time a range of evaluation exercises and monitor the changes in product temperature as part of the process of checking whether ranking is a viable test methodology for the project.

12.7.2.6 Undue Influence of Appearance

Products may differ clearly on an appearance characteristic, for example ranging from light to dark, which may cause confusion in terms of the specified attribute focused on some other aspect of the products. In this situation, it may be possible to perform the test under coloured lighting to disguise the appearance differences.

12.7.2.7 Need to Compare the Products on Many Attributes

Each ranking exercise of a designated attribute requires testing a complete set of products and if this has to be repeated many times then this can have serious implications for practicality of the methodology. In particular, for evaluation of foods/drinks, the total amount that could be consumed in a single session will be a major consideration, together with likely sensory/task fatigue.

12.7.3 Choice of Test Methodology: Ranking versus Rank-Rating versus Alternative

Given the pros and cons of ranking described here, it will be important to decide whether ranking is the best choice of methodology for a given project objective and product type. If the concerns are not too great, it may be that the extra sensitivity of the ranking procedure will outweigh the limitations. For example, it may be felt that a certain amount of carry-over from one product to the next, even with palate cleansing, can be tolerated and still provide a sensitive comparison between the products, provided steps are taken to minimize any other adverse influences. The most likely barrier to use of ranking in the case of food is the total amount of product consumed in a given session, which may have an absolute limit based on safety guidelines or ethical considerations.

If there are concerns that ranking is not appropriate then the most likely alternative is some form of descriptive analysis with products presented monadically.

If a ranking-based methodology is practical then the next decision is whether the extended rank-rating procedure should be used rather than ranking alone. Ranking will provide measures of the relative positions of the products along an underlying intensity scale and the separation between the products but, in the absence of any prior information about the products, no measure of where they sit along the 'none' to 'very high' continuum of intensity. If one of the products is a known reference then the positioning of products against this reference may be sufficient.

If mean product scores along a scale are required then the decision is between using the rank-rating methodology or reverting to a standard methodology with sequential monadic presentation.

If the mean product scores must have an absolute interpretation, either in terms of comparing results across different studies or using incomplete block design within a given study, then a standard methodology with sequential monadic presentation is likely to be the safer choice. Conversely, if the product mean scores need only be interpreted in a relative way, accepting that a mean score for one product will be influenced by the products against which it has been compared, then rank-rating is a viable option and will probably benefit from the increased sensitivity from the simultaneous comparison of the products.

The essential elements of the decision process are captured in Figure 12.9.

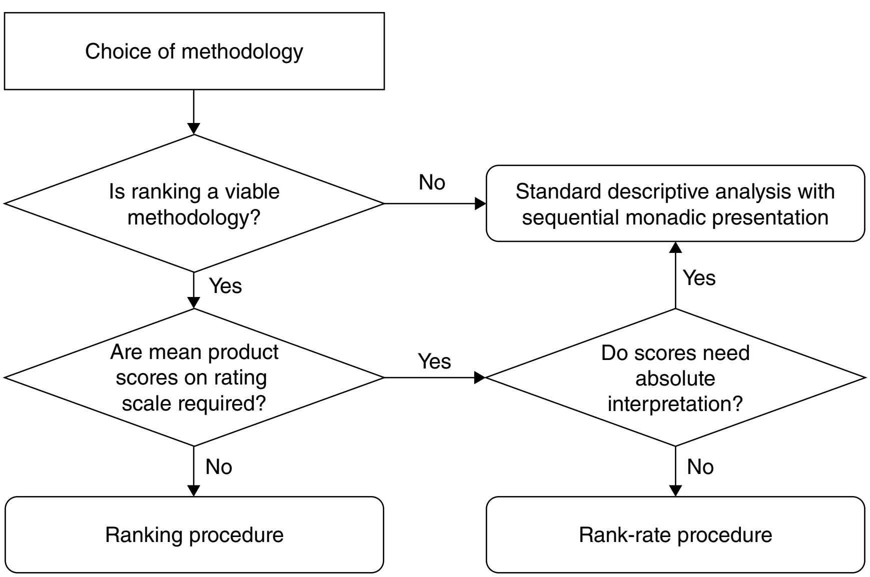


Figure 12.9 Decision chart for choice of methodology.

A particular note should be made here for studies with naive participants, for example consumer studies carried out in a central location (an in-home consumer study involving ranking of products is unlikely to be practical). Whereas assessors participating in an in-lab sensory study are likely to be familiar with using rating scales, for naive consumers this may be the first time that they have been tasked with rating products on a scale and they may find that they need to have assessed several products presented monadically before they can make sense of the rating scale and give liking scores to the subsequent products that have the intended absolute interpretation in terms of position along the scale. In this situation, presenting all the products together and ranking them first in terms of liking may be an easier task to explain and help the participants to give more reliable liking scores as a second stage. A study specifically designed to compare separate groups of consumers each rating the same set of dessert products (Cleaver & Wedel 2001) but with different liking elicitation methods showed improved results from the rank-rate procedure compared to conventional sequential monadic presentation.

12.8 Case Studies

Three case studies are presented here to illustrate different features of ranking and rank-rating studies in terms of design, analysis and output. Each is based on simulated data but representative in size and nature of typical sensory studies and with the data presented in full so that the analyses can be recreated by following the analysis processes described in this chapter.

12.8.1 Case Study 1: Complete Block Example

The data in Table 12.11 are an example of a study comparing various savoury snack products where each of 12 assessors ranks all six products on increasing levels of perceived crispness and repeats the same task in three separate sessions. The attribute 'crispness' was clearly defined to the assessors and there were no particular problems associated with ranking six products, so ranking was deemed to be an appropriate methodology to provide a sensitive comparison of the products.

Note that different codes are used in each session to disguise the identity of the products. Note also that 'Assessor' refers to the numbering within this study whereas 'Assessor ID' is the unique code for that assessor that applies across all studies in which she/he participates. This is particularly important for output on assessor performance as illustrated in this case study, and where individual assessor performance needs to be monitored over time. In this case, the Assessor ID also identifies the gender of each assessor.

Table 12.11 Case Study 1: Data for complete block example (1 = high).

Session	Assessor	Assessor ID	Prod 1 J72	Prod 2 V19	Prod 3 M27	Prod 4 F38	Prod 5 K54	Prod 6 C95
1	1	F395	4	5	3	2	6	1
	2	F276	3	4	2	1	6	5
	3	F742	6	5	2	1	4	3
	4	F265	6	5	1	2	4	3
	5	F193	5	4	1	3	6	2
	6	M281	4	2	5	1	6	3
	7	M327	3	4	6	2	5	1
	8	M812	6	3	2	1	4	5
	9	F921	4	5	1	2	6	3
	10	F432	4	5	2	1	6	3
	11	M625	3	5	2	1	6	4
	12	F547	5	3	4	2	6	1
Session	**Assessor**	**Assessor ID**	**Prod 1 B75**	**Prod 2 E89**	**Prod 3 U48**	**Prod 4 H26**	**Prod 5 D63**	**Prod 6 N13**
2	1	F395	4	3	2	1	5	6
	2	F276	4	5	3	2	6	1
	3	F742	3	4	6	1	2	5
	4	F265	4	6	2	3	5	1
	5	F193	4	6	2	1	5	3
	6	M281	6	2	1	4	3	5
	7	M327	1	5	2	4	6	3
	8	M812	4	5	2	3	6	1
	9	F921	4	5	2	1	6	3
	10	F432	2	6	4	1	5	3
	11	M625	3	5	2	1	6	4
	12	F547	2	3	1	5	6	4
Session	**Assessor**	**Assessor ID**	**Prod 1 K32**	**Prod 2 G41**	**Prod 3 Y59**	**Prod 4 Q12**	**Prod 5 P70**	**Prod 6 L25**
3	1	F395	2	4	5	6	3	1
	2	F276	5	6	4	1	3	2
	3	F742	5	4	3	1	6	2
	4	F265	5	6	3	2	4	1
	5	F193	6	4	1	2	5	3
	6	M281	3	6	1	2	5	4
	7	M327	1	5	3	2	6	4
	8	M812	5	4	1	2	6	3
	9	F921	3	6	1	2	5	4
	10	F432	4	3	5	2	6	1
	11	M625	5	6	3	1	4	2
	12	F547	6	5	1	2	4	3
			144	164	91	71	183	103

Although the results from the different analysis procedures are included here, this case study is used to specifically illustrate:

- how the replication in a complete block study can be used to check assessor performance
- a comparison of different analyses using the two variants of the R-Index.

12.8.1.1 Assessor Performance

As with descriptive analysis, it is good practice to carry out some checks on assessor performance for individual studies and, ideally, to monitor performance over time. This is particularly important during the training phase when, as highlighted earlier, there is often a correct ordering of the products or stimuli against which the actual rankings can be compared. It is important to remember that all checks on assessor performance, whatever the methodology, are heavily dependent on the particular set of products being compared. We can expect assessors to discriminate well between products with large sensory differences but the same assessors, presented with products that are much closer in sensory space, will not be able to achieve the same level of discrimination and so performance measures will be lower. With this in mind, assessor performance is most effectively monitored by quantifying each assessor's performance in relation to the other assessors in each study and accumulating the results over a series of studies to identify consistent patterns or trends in performance over time in relative terms.

It should be noted that the simple checks on assessor performance described here are most effective when the ranking involves at least five products and, specifically for the check on internal consistency, that each ranking task is repeated on more than one occasion.

This data set in Table 12.11, with replication of the task on different occasions, is used to demonstrate how simple performance measures can be used when there is no prior ordering of the products. Taking assessor F276 as an example, the rankings for the individual sessions, the mean rankings for this assessor and the mean rankings over all assessors are set out in Table 12.12.

Table 12.12 Case Study 1: Individual and mean rankings for assessor F276.

Assessor F276	Ranking (1 = high)					
	Prod 1	**Prod 2**	**Prod 3**	**Prod 4**	**Prod 5**	**Prod 6**
Session 1	3	4	2	1	6	5
Session 2	4	5	3	2	6	1
Session 3	5	6	4	1	3	2
F276: Mean	4.00	5.00	3.00	1.33	5.00	2.67
All Assessors	4.00	4.56	2.53	1.97	5.08	2.86

There are two simple measures that can be applied here to provide a summary of performance of this assessor.

- *Internal consistency*. When an assessor is required to perform the same task on three separate occasions, one might optimistically assume that the same set of rankings would result each time. If the products were all very different from one another on the designated attribute, this might be the case, but sensory studies are usually carried out where the products are potentially confusable and some inconsistency is to be expected. We can use the Kendall coefficient of concordance (Siegel & Castellan 1998) as a measure of agreement amongst the three sets of rankings. The resulting statistic ranges from 0 (no agreement at all between the three sets of rankings) to 1 (perfect agreement). In this case, as suggested by inspection of the ranks in each session, the level of agreement is moderate and the concordance value W = 0.594.
- *Agreement with the panel as a whole*. A complementary test of assessor performance is to check how each assessor's rankings (mean over three replicates in this case) correlate with the mean rankings over all assessors. The Spearman rank correlation (Siegel & Castellan 1998) between these two sets of mean rankings is 0.928 for assessor F276. A modification of this approach is to correlate the mean rankings for the selected assessor with the mean rankings based on data excluding that assessor.

The measures of performance for each individual assessor are shown in Table 12.13.

This exercise is particularly useful in the training phase for ranking to help identify assessors whose ranks are less consistent than others or possibly have a different interpretation of the designated attribute. Over a series of live ranking

Table 12.13 Case Study 1: Measures of performance for each assessor.

Assessor	Assessor ID	Consistency	Agreement
		Kendall concordance	Spearman correlation
1	F395	0.149	0.783
2	F276	0.594	0.928
3	F742	0.492	0.714
4	F265	0.835	0.600
5	F193	0.835	0.886
6	M281	0.289	0.812
7	M327	0.619	0.551
8	M812	0.670	0.886
9	F921	0.924	0.943
10	F432	0.695	0.829
11	M625	0.822	0.986
12	F547	0.416	0.771
Median		0.644	0.820

studies, it can also be useful to monitor performance over time using these statistics. Assessors who are repeatedly inconsistent or not in accord with the rest of the panel may be in need of retraining.

It should be stressed, though, that within any given live ranking study, assessors should not be removed from the analysis on the basis of these statistics which are based on limited replication. An exception here could be if a particular assessor showed both (a) a high level of internal consistency and (b) a high negative correlation for agreement, which together may be indicative of that assessor recording the rankings in the wrong intended direction and that this had not been picked up in the training phase.

Another way of exploring agreement between assessors is to create a biplot, using standard statistical software such as JMP or XLSTAT, based on the mean product rankings for each assessor (Figure 12.10). This analysis is exactly the same as running an internal vector preference map based on consumer liking scores for a set of products. Note that in cases, as here, where rank = 1 represents the products ranked highest, the direction of the rankings needs to be reversed before this analysis in order that the direction of the vectors conveys increasing intensity of the attribute.

In this representation, there is one vector for each assessor and the projections from each product onto that vector (draw a line from that product to meet the vector at right angles) should correspond to the rank ordering of the products by that assessor. For example, focusing on assessor F276, the

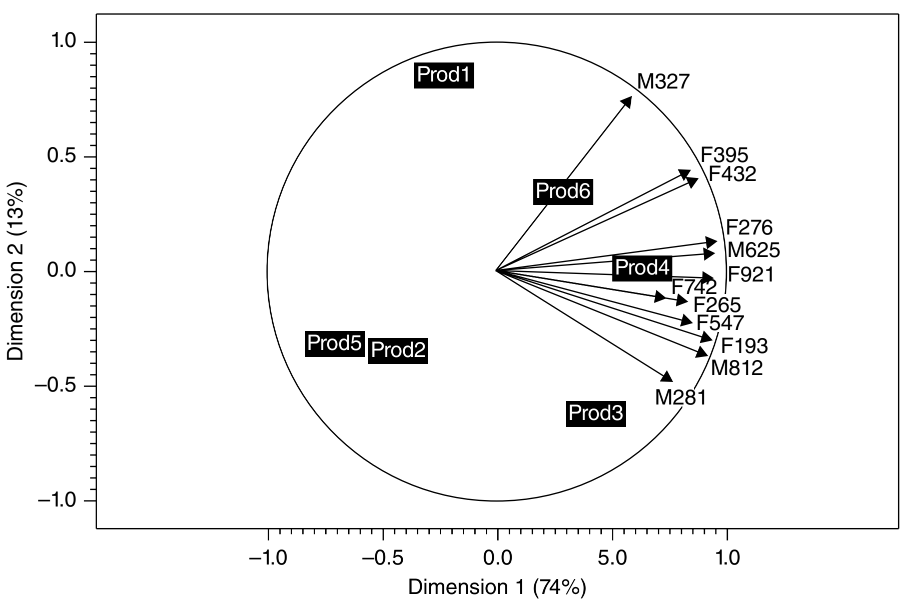


Figure 12.10 Case Study 1: Biplot representation of mean rankings for each assessor.

projections are in the order: Prod4 (Highest), Prod3 = Prod6, Prod1, Prod2, Prod5 (Lowest), which is a close representation of the ordering in Table 12.12. By contrast, the product projections for assessor M327 are rather different. In this form of biplot, the length of the vector is a measure of how well the orientation of that vector represents the data for that assessor. The aim of this analysis is to provide a visual way of interpreting differences between the assessors. In this example, the results suggest that the differences between the assessors are driven mostly by disagreement over the relative positioning of products 1 and 3 in the ranking.

Where there is good agreement across the assessors, the percentage of variability accounted for by the first dimension, in this case 74%, will be high and this can serve as one metric of panel performance and which should be related to the overall measure of panel agreement described above.

12.8.1.2 Friedman Analysis

The standard simple analysis of ranking studies is the Friedman analysis of variance and following the analysis procedure from 12.5.1 gives T = 77.75 which treated as a chi-square statistic with 5 df shows that the overall difference between the products was highly significant ($P < 0.0001$).

Proceeding to pairwise significance tests between the products gives a least significant difference (LSD) at the 5% level of 31.3. Alternatively, if the differences between all pairs of products are of interest, a multiple comparison test can be used to take account of the large number of possible pairs. This is best implemented with sensory software such as XLSTAT which uses the Nemenyi procedure for multiple comparisons, giving the results in Table 12.14 where products which are not significantly different are underlined.

Table 12.14 Case Study 1: Friedman Analysis Multiple comparison test using the Nemenyi procedure.

Rank totals for each product Increasing intensity →					
Prod 5	Prod 2	Prod 1	Prod 6	Prod 3	Prod 4
183	164	144	103	91	71

12.8.1.3 R-Index Analysis

This example is used to illustrate how the two types of R-Index can be used to analyse ranking studies.

Firstly, the simple R_{JB} statistic for each of the pairwise product comparisons is determined from the number of times, within each ranking set, that the first product is ranked higher than the second. For example, from Table 12.11 Product 3 is ranked higher than Product 2 in 30 out of the 36 ranking sets, so $R_{JB} = 83.3\%$. Testing this for significance gives $P = 0.0007$. Calculating these individual R_{JB} values and summarizing the significance tests with *** = $P < 0.001$, ** = $P < 0.01$, * = $P < 0.05$, the results are shown in Table 12.15.

Table 12.15 Case Study 1: R_{JB} Analysis Pairwise R-Index R_{JB} values.

Product	Pairwise R-Index (R_{JB}) values					
	Product 5	**Product 2**	**Product 1**	**Product 6**	**Product 3**	**Product 4**
Product 5	.	33.3 ns	8.3 **	0.0 ***	0.0 ***	0.0 ***
Product 2	66.7 ns	.	33.3 ns	0.0 ***	0.0 ***	0.0 ***
Product 1	91.7 **	66.7 ns	.	25.0 ns	0.0 ***	8.3 **
Product 6	100.0 ***	100.0 ***	75.0 ns	.	25.0 ns	8.3 **
Product 3	100.0 ***	100.0 ***	100.0 ***	75.0 ns	.	16.7 *
Product 4	100.0 ***	100.0 ***	91.7 **	91.7 **	83.3 *	.

For analysis of the same data using the R_{MAT} version of the R-Index, two types of analysis are possible. The first method is to pool all of the data over all assessors into a table for each of the possible pairwise comparisons, illustrated in Table 12.16 for Product 1 versus Product 2.

Table 12.16 Case Study 1: R_{MAT} Analysis Pooled data across all assessors.

Product	Ranking (1 = highest)						Total
	6th	**5th**	**4th**	**3rd**	**2nd**	**1st**	
Product 1	2	3	7	11	7	6	36
Product 2	0	2	5	8	13	8	36

Following the calculations in section 12.5.2:

$$R_{MAT} = 100 \times \left(2x(0) + 3x(0 + 2x½) + 7x(0 + 2 + 5x½) + 11x(0 + 2 + 5 + 8x½)\right.$$
$$\left. + 7x(0 + 2 + 5 + 8 + 13x½) + 6x(0 + 2 + 5 + 8 + 13 + 8x½)\right) / (36 * 36) = 38.4\%$$

Analysing the same table with a Mann-Whitney U test with a correction for ties (Siegel & Castellan 1998) gives $P = 0.0414$. The results for all possible pairs could be set out in the same format as Table 12.15 using the alternative R-Index values and significance tests.

Table 12.17 Case Study 1: R_{MAT} Analysis R-Index R_{MAT} values calculated separately for each assessor.

Pairwise comparison		Assessors arranged in increasing order of rank correlation with R-Index mean												R-Index mean	Wilcoxon test
Prod A	Prod B	7 M327	6 M281	3 F742	1 F395	12 F547	4 F265	8 M812	5 F193	11 M625	10 F432	9 F921	2 F276		
2	4	5.6	33.3	0.0	33.3	27.8	0.0	5.6	0.0	0.0	0.0	0.0	0.0	8.8	***
5	**6**	0.0	33.3	38.9	27.8	5.6	0.0	11.1	0.0	11.1	0.0	0.0	11.1	11.6	***
1	**4**	77.8	16.7	0.0	38.9	27.8	0.0	0.0	0.0	0.0	5.6	0.0	0.0	13.9	**
2	**3**	33.3	22.2	33.3	33.3	22.2	0.0	0.0	0.0	0.0	27.8	0.0	5.6	14.8	***
2	**6**	5.6	66.7	27.8	33.3	33.3	0.0	33.3	0.0	0.0	11.1	0.0	16.7	19.0	**
1	**3**	83.3	22.2	33.3	50.0	11.1	0.0	0.0	0.0	11.1	61.1	0.0	22.2	24.5	*
1	**6**	72.2	44.4	22.2	33.3	22.2	0.0	16.7	0.0	44.4	22.2	33.3	27.8	28.2	**
3	**4**	33.3	55.6	0.0	38.9	77.8	61.1	61.1	72.2	0.0	5.6	66.7	5.6	39.8	ns
3	**6**	38.9	72.2	44.4	33.3	61.1	38.9	72.2	94.4	77.8	22.2	100.0	38.9	57.9	ns
1	**2**	100.0	27.8	38.9	66.7	38.9	72.2	22.2	38.9	88.9	77.8	100.0	77.8	62.5	ns
2	**5**	88.9	72.2	44.4	66.7	88.9	5.6	83.3	72.2	55.6	72.2	66.7	55.6	64.4	ns
4	**6**	50.0	83.3	100.0	38.9	44.4	27.8	66.7	72.2	100.0	77.8	100.0	72.2	69.4	*
1	**5**	100.0	55.6	38.9	77.8	66.7	27.8	61.1	61.1	88.9	100.0	100.0	72.2	70.8	*
3	**5**	77.8	83.3	55.6	77.8	94.4	100.0	100.0	100.0	100.0	94.4	100.0	83.3	88.9	***
4	5	100.0	88.9	100.0	72.2	88.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	95.8	***

With replication of the same ranking task by each assessor in different sessions, it is possible to calculate an R-Index for each assessor in turn for each of the pairwise product comparisons, and then perform a significance test of whether the overall R-Index differs from the chance level of 50%. Three replications of each product pair by each assessor is low for this type of analysis but it is included here for comparison and to illustrate the differences between assessors, rather than pooling the data over assessors at the outset.

Following the same R-Index calculations as before for each product pair, but separately for each assessor, the R_{MAT} values are set out in Table 12.17, taking the opportunity to order the results in terms of:

- rows (product pairs) by increasing value of the mean R_{MAT} value over assessors, and
- columns (assessors) by the rank correlation (over the product pairs) between that assessor's R_{MAT} value and that averaged over all assessors, so that assessors who are most in agreement with the panel as a whole are towards the right of the table.

The significance test applied here is designed to test whether the mean R-Index for each product pair differs from 50% (corresponding to no difference at all) in relation to the variation between assessors. In each case, an R-Index greater than 50% denotes that the product in the 'Prod A' column was perceived as higher in intensity than the product in the 'Prod B' column, and vice versa for an R-Index less than 50%. A Student's t test could be used for this significance test but with 12 assessors providing information on the assessor variation, the non-parametric Wilcoxon matched-pairs signed-ranks test (Siegel & Castellan 1998) was used here instead.

12.8.1.4 Parametric Analysis

Following the analysis procedure set out in section 12.5.3 and setting the strata to be a combination of assessor and session, running the equivalent SAS code using the PHREG procedure provides parameter estimates for each product which are set out in Figure 12.11, with products that do not differ significantly at the 5% level underlined. As before, the parameter estimates can be converted to estimates, for each product pair, of the probabilities that one product is ranked higher than the other, as shown in Table 12.18.

Figure 12.11 Case Study 1: Results from parametric analysis.

Table 12.18 Case Study 1: Parametric Analysis Pairwise probabilities that product (row) is ranked higher than product (column).

Product	Product 5	Product 2	Product 1	Product 6	Product 3	Product 4
Product 5	.	0.38	0.29	0.14	0.12	0.07
Product 2	0.62	.	0.40	0.22	0.18	0.11
Product 1	0.71	0.60	.	0.29	0.24	0.15
Product 6	0.86	0.78	0.71	.	0.44	0.31
Product 3	0.88	0.82	0.76	0.56	.	0.36
Product 4	0.93	0.89	0.85	0.69	0.64	.

12.8.2 Case Study 2: Incomplete Block Example

In this study there are seven products to be compared in terms of sweetness of a dessert product but it was felt that assessors could not reliably compare all seven products in a single ranking session. It was decided instead that the largest number of products that could be ranked reliably was four, with appropriate palate cleansing between each tasting. The 'incomplete block' test design used here is created by replicating and randomizing four times the basic design shown earlier in section 12.4.2.

The resulting ranking data are shown in Table 12.19.

Setting the data out in this way already gives a visual impression of the product differences.

- Where Product 6 is included in the product set, it is nearly always ranked highest or second-highest.
- Products 2 and 5 are nearly always ranked lowest or second-lowest.
- Product 7 is nearly always ranked second or third in each set of four products and so we would expect it to appear towards the middle of the estimated intensities for the complete set of seven products in the parametric analysis.

Though not illustrated here, it is possible to apply a variant of the Friedman analysis to incomplete block studies using the Skillings–Mack procedure (Skillings & Mack 1981). Software implementation is available through, for example, XLSTAT, SAS, Stata® and R.

A simple analysis is possible using the R_{JB} variant of the R-Index analysis, which, for each product pair, is determined from those blocks in which both products appear. In this example which is a balanced incomplete block design, each product pair occurs together in exactly 12 blocks so the R_{JB} value is simply the percentage of those comparisons in which the first product was ranked higher than the other. Table 12.20 shows the results from each pair with significance from an exact binomial test.

For a parametric analysis, setting up the data in the same way as in the example in section 12.5.3 and applying the same type of model gives the set of parameter estimates and pairwise multiple comparisons shown in Figure 12.12, with products that do not differ significantly at the 5% level underlined.

Table 12.19 Case Study 2: Incomplete block data (1 = high).

Session	Assessor	Assessor ID	Prod 1 H67	Prod 2 K82	Prod 3 Q70	Prod 4 V23	Prod 5 S92	Prod 6 C41	Prod 7 F17
1	1	F167		4		3		2	1
	2	F752	4		3	2		1	
	3	F387		3	2		4	1	
	4	F960	4	3		1	2		
	5	F759			1	2	4		3
	6	F490	1	4	2				3
	7	F117	3				4	1	2
	8	F201	4		1	2			3
	9	F862				1	4	2	3
	10	F939	3	4		2		1	
	11	F350	1	3			4		2
	12	F555		4	2			1	3
	13	F802	3		2		4	1	
	14	F209		3	1	2	4		
Session	**Assessor**	**Assessor ID**	**Prod 1 N60**	**Prod 2 G39**	**Prod 3 E13**	**Prod 4 L79**	**Prod 5 W21**	**Prod 6 D87**	**Prod 7 M52**
2	1	F167			1	2	4		3
	2	F752		4			3	1	2
	3	F387	3			2	4	1	
	4	F960	4	3		1			2
	5	F759	4		2			1	3
	6	F490		4	1	3		2	
	7	F117	2	3	1		4		
	8	F201		3	1		4		2
	9	F862		4		2	3	1	
	10	F939	3	4				1	2
	11	F350	2			1	4		3
	12	F555	3		2		4	1	
	13	F802			4	3		1	2
	14	F209	3	4	1	2			
Session	**Assessor**	**Assessor ID**	**Prod 1 T29**	**Prod 2 U81**	**Prod 3 J59**	**Prod 4 X91**	**Prod 5 B14**	**Prod 6 P32**	**Prod 7 Y74**
3	1	F167	4			1		2	3
	2	F752	3	2			4	1	
	3	F387	3		1		4		2
	4	F960		4	1			3	2
	5	F759			2	3	4	1	
	6	F490		3		1	4		2
	7	F117	3	4	1	2			
	8	F201		4	1	3			2
	9	F862	3	2			4		1
	10	F939			1		4	2	3
	11	F350	3	4	2			1	
	12	F555		4		2	3	1	
	13	F802	3		1	2	4		
	14	F209	4			1		2	3

Using all the data in a single analysis, the comparisons between the products will be more sensitive than that from the simple R_{JB} analysis. As before, the parameter estimates can be converted into pairwise probabilities that one product will be ranked higher than the other (Table 12.21).

Table 12.20 Case Study 2: R-Index R_{JB} for each pair of products.

Product	Product 5	Product 2	Product 1	Product 7	Product 4	Product 3	Product 6
Product 5	.	33.3 ns	8.3 **	0.0 ***	0.0 ***	0.0 ***	0.0 ***
Product 2	66.7 ns	.	33.3 ns	0.0 ***	0.0 ***	0.0 ***	0.0 ***
Product 1	91.7 **	66.7 ns	.	25.0 ns	0.0 ***	8.3 **	0.0 ***
Product 7	100.0 ***	100.0 ***	75.0 ns	.	25.0 ns	8.3 **	16.7 *
Product 4	100.0 ***	100.0 ***	100.0 ***	75.0 ns	.	16.7 *	25.0 ns
Product 3	100.0 ***	100.0 ***	91.7 **	91.7 **	83.3 *	.	25.0 ns
Product 6	100.0 ***	100.0 ***	100.0 ***	83.3 *	75.0 ns	75.0 ns	.

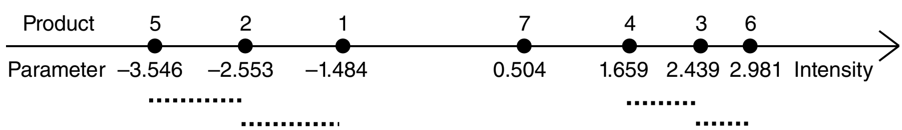


Figure 12.12 Case Study 2: Results from parametric analysis.

Table 12.21 Case Study 2: Parametric Analysis Pairwise probabilities that product (row) is ranked higher than product (column).

Product	Product 5	Product 2	Product 1	Product 7	Product 4	Product 3	Product 6
Product 5	.	0.27	0.11	0.02	0.01	0.00	0.00
Product 2	0.73	.	0.26	0.04	0.01	0.01	0.00
Product 1	0.89	0.74	.	0.12	0.04	0.02	0.01
Product 7	0.98	0.96	0.88	.	0.24	0.13	0.08
Product 4	0.99	0.99	0.96	0.76	.	0.31	0.21
Product 3	1.00	0.99	0.98	0.87	0.69	.	0.37
Product 6	1.00	1.00	0.99	0.92	0.79	0.63	.

Table 12.22 Case Study 3: Ranking and ratings.

Session	Assessor	Assessor ID	Rank (1 = lowest)					Score (0–10)				
			1 J64	2 P32	3 N83	4 E16	5 V29	1 J64	2 P32	3 N83	4 E16	5 V29
1	1	M649	2	3	4	5	1	2	4	5	10	1
1	2	M794	2	3	4	5	1	5	6	7	10	4
1	3	M805	2	3	5	4	1	6	7	10	10	3
1	4	F329	2	3	5	4	1	2	6	9	8	1
1	5	F670	1	4	3	5	2	3	6	5	8	4
1	6	M186	2	3	5	4	1	2	4	10	7	1
1	7	F280	2	3	5	4	1	1	2	8	6	1
1	8	F734	2	3	4	5	1	2	3	5	8	1
1	9	F270	1	3	5	4	2	3	6	10	7	4
1	10	F448	3	2	4	5	1	4	4	6	8	1
Session	**Assessor**	**Assessor ID**	**Rank (1 = Lowest)**					**Score (0 – 10)**				
			1 T54	**2 B97**	**3 F47**	**4 L72**	**5 S89**	**1 T54**	**2 B97**	**3 F47**	**4 L72**	**5 S89**
2	1	M649	2	3	4	5	1	3	7	8	2	3
2	2	M794	4	2	3	5	1	4	5	9	2	4
2	3	M805	2	3	5	4	1	8	10	9	3	8
2	4	F329	3	2	5	4	1	4	10	6	1	4
2	5	F670	2	3	4	5	1	6	8	8	2	6
2	6	M186	2	3	5	4	1	5	10	7	2	5
2	7	F280	2	3	5	4	1	3	8	6	1	3
2	8	F734	2	3	4	5	1	4	6	8	1	4
2	9	F270	2	3	4	5	1	6	9	10	1	6
2	10	F448	2	3	4	5	1	6	6	8	3	6

12.8.3 Case Study 3: Rank-Rating Example

As an example of rank-rating data, Table 12.22 shows the results of 10 assessors each evaluating the same set of five products on two occasions. If the assessors are assigning numerical ranks to the products in the ranking phase, it is sensible to assign 1 = Lowest and 5 = Highest to avoid confusion with the direction of scoring in the rating phase.

Clearly, the ranking data can be analysed in the same way as before. The pragmatic way of analysing the scoring data is to treat the data as though they were conventional descriptive analysis data and use an appropriate mixed-model analysis of variance and choice of correction for multiple comparisons.

As an example, using PROC MIXED in SAS gives an overall test of significance for product differences of $F(4,36) = 54.97$ and $P < 0.0001$. Using the Tukey–Kramer method for adjusting the significance of the pairwise comparisons to allow for multiple comparisons, the results are summarised in Table 12.23, with products that do not differ significantly at the 5% level underlined.

Table 12.23 Case Study 3: Mean scores and pairwise multiple comparisons.

Mean score for each product Increasing intensity →				
Prod 5	Prod 1	Prod 2	Prod 3	Prod 4
1.95	3.55	4.85	7.70	8.05
	....	....	....	....

12.9 Summary

Ranking-based methods do have wide potential application in sensory evaluation, especially where a sensitive comparison between products is required across a limited number of attributes. Care must be taken, though, in deciding whether ranking can be used for a given test situation and guidelines are provided in this chapter to help make that decision.

Although primarily used in the situation where an assessor is able to rank all products against one another in a single task, the technique can also be applied when only a ranking of subsets of the products is possible, extending the area of application into studies with large numbers of products to compare. This does require careful consideration of experimental design and access to more sophisticated software however.

Three methods of analysing ranking data are presented in this chapter: the standard Friedman test, two variants of the R-Index analysis and a parametric analysis which has potential for wider use in the sensory community. Some simple methods of evaluating assessor performance are also included.

Extending the methodology by using rank-rating provides a pragmatic way of obtaining mean intensity scores for each product along the chosen rating scale.

12.10 Future Directions

Ranking-based methods will always be a core element of the sensory analysis toolbox, especially where a very sensitive comparison between a set of products is required over a limited set of attributes. The ability to model the resulting data in new ways, in combination with appropriate experimental design, does open up more possibilities for wider application with larger numbers of products and through use of incomplete block designs. There is scope for a systematic evaluation of incomplete designs in the context of ranking, through a combination of a large-scale simulation exercise and confirmatory studies, to provide practical guidance on the minimum block size and replication required to provide robust comparison of a given number of products.

The parametric analysis models open up more possibilities for creative design of new test methodologies. For example, a 'choose and rank' procedure could be devised for a specific hedonic attribute, such as 'smooth feel', or simply overall liking. The consumers are asked to select a (specified or unspecified) number of products after sampling each of them based on which they would choose. They are then asked to rank on that attribute just the products they have chosen, in order to keep the task simple. Giving an equal tied rank to the products not chosen will enable an analysis across all the products but focus attention on comparison of the well-liked products.

As an alternative to the parametric model presented here, a Thurstonian analysis of ranking data is also possible, requiring specialist software (Ennis & Ennis 2013), with the specific feature that the estimated product effects are presented in terms of Thurstonian d′ values, providing some consistency with size of product difference estimated from other test methodologies.

Some simple ways are presented here to assess and compare how panellists have performed a given ranking task. As a single study is likely to be quite limited in size, these performance measures will be quite variable. This process therefore becomes much more powerful when the performance measures from each study are calculated automatically and part of the analysis process, appended to a live database, and then plotted out over time. This provides an ongoing check of how each panellist is performing in relation to the others, and picks up any trends in underperformance that need to be addressed by retraining or deselection from the panel. With the increased use of sensory management software in the design and conduct of sensory studies and the subsequent databasing of the primary and secondary data, this type of application can be automated to make much better use of sensory data over time, rather than treating each sensory study as a stand-alone exercise.

Appendix 12.1: Critical Values for the Friedman Test

No. of products (k)	No. of blocks (b)	Significance level			
		P < 0.10	P < 0.05	P < 0.025	P < 0.01
3	3	6.000	6.000	–	–
	4	6.000	6.500	8.000	8.000
	5	5.200	6.400	7.600	8.400
	6	5.333	7.000	8.333	9.000
	7	5.429	7.143	7.714	8.857
	8	5.250	6.250	7.750	9.000
	9	5.556	6.222	8.000	8.667
	10	5.000	6.200	7.800	9.600
	11	4.909	6.545	7.818	9.455
	12	5.167	6.500	8.000	9.500
	13	4.769	6.000	7.538	9.385
	14	5.143	6.143	7.429	9.000
	15	4.933	6.400	7.600	8.933
4	2	6.000	6.000	–	–
	3	6.600	7.400	8.200	9.000
	4	6.300	7.800	8.400	9.600
	5	6.360	7.800	8.760	9.960
	6	6.400	7.600	8.800	10.200
	7	6.429	7.800	9.000	10.371
	8	6.300	7.650	9.000	10.500
	9	6.467	7.800	9.133	10.867
	10	6.360	7.800	9.120	10.800
	11	6.382	7.909	9.327	11.073
	12	6.400	7.900	9.200	11.100
	13	6.415	7.985	7.369	11.123
	14	6.343	7.886	9.343	11.143
	15	6.440	8.040	9.400	11.240
5	2	7.200	7.600	8.000	8.000
	3	7.467	8.533	9.600	10.133
	4	7.600	8.800	9.800	11.200
	5	7.680	8.960	10.240	11.680
	6	7.733	9.067	10.400	11.867
	7	7.771	9.143	10.514	12.114
	8	7.800	9.300	10.600	12.300
	9	7.733	9.244	10.667	12.444
	10	7.760	9.280	10.720	12.480
6	2	8.286	9.143	9.429	9.714
	3	8.714	9.857	10.810	11.762
	4	9.000	10.286	11.429	12.714
	5	9.000	10.486	11.743	13.229
	6	9.048	10.571	12.000	13.619
	7	9.122	10.674	12.061	13.857
	8	9.143	10.714	12.214	14.000
	9	9.127	10.778	12.302	14.143
	10	9.143	10.800	12.343	14.229

Source: Martin et al. (1993).

References

Allison, P.D. & Christakis N.A. (1994) Logit models for sets of ranked items. *Sociological Mathodology*, **24**, 199–228

Brown, J. (1974) Recognition assessed by rating and ranking. *British Journal of Psychology*, **65**, 13–22.

Cleaver, G.J. & Wedel M.(2001) Identifying random-scoring respondents in sensory research using finite mixture regression models. *Food Quality & Preference*, **12**, 373–384.

David, H.A. (1988) *The Method of Paired Comparisons*. London: Hodder Arnold.

Ennis, D.M. & Ennis. J.M. (2013) A Thurstonian ranking model with rank-induced dependencies. *Journal of Classification*, **30**, 124–147.

Hobbs, G. (2009) Using SAS® for Nonparametric Statistics. SAS® Conference Proceedings: Western Users of SAS® Software, San José, California.

Kim, K.O. & O'Mahony, M. (1998) A new approach to category scales of Intensity I: Traditional vs Rank-Rating. *Journal of Sensory Studies*, **13**, 241–249.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evaluation of Food*. New York: Springer.

Levitt, D.J. (1973) Ranking in incomplete blocks: the use of the technique and the interpretation of the results. *Journal of the Science of Food and Agriculture*, **24**, 739–745.

Levitt. D.J. (1975) Ranking in incomplete or complete blocks: a parametric analysis. *Biometrics*, **31**, 745–748.

Martin, L., Leblanc, R. & Toan N.K. (1993) Tables for the Friedman rank test. *Canadian Journal of Statistics*, **21**, 39–43.

Nemenyi, P.B. (1963) Distribution-free multiple comparisons. PhD thesis, Princeton University.

O'Mahony, M. (1992) Understanding discrimination tests: a user-friendly treatment of response bias, rating and ranking R-Index and their relationship to signal detection. *Journal of Sensory Studies*, **7**, 1–47.

Siegel, S. & Castellan, N.J.(1998) *Nonparametric Statistics for the Behavioral Sciences*. New York: McGraw-Hill.

Skillings, J.H. & Mack, G.A. (1981) On the use of a Friedman-type statistic in balanced and unbalanced block designs. *Technometrics*, **23**(2), 171–177.

Thurstone, L.L. (1931) Rank order as a psychophysical measure. *Journal of Experimental Psychology*, **14**, 187–201.

Wakeling, I.N. & Buck, D. (2001) Balanced incomplete block designs useful for consumer experimentation. *Food Quality & Preference*, **12**, 265–268.

Wakeling, I.N. & MacFie, H.J.H. (1995) Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality & Preference*, **6**(4), 299–308.

CHAPTER 13

Free Choice Profiling

Pieter H. Punter

13.1 Introduction

Free choice profiling (FCP) is a form of quantitative descriptive analysis (QDA). In QDA trained assessors or consumers rate different products on a predefined set of sensory attributes (the vocabulary). In trained or expert panels, every assessor has been trained to understand and agree on the meaning of the attributes used, often with the use of reference standards (Lawless & Heymann 2010; Meilgaard et al. 2007; Stone & Sidel 2004). The perceived attribute intensities can be rated on line or category scales. Training can take from several days to several months, depending upon the amount of sophistication required. When used with consumers, there is no training phase. In this case, the attributes used should be 'consumer' terms. Since every assessor rates the same set of attributes, the perceived intensities can be averaged over assessors and the resulting product profiles can be analysed statistically (both univariate or multivariate) and they can be represented visually as spider plots.

13.2 Methodology of Free Choice Profiling

13.2.1 Principle

In FCP, no common vocabulary is used, as each assessor is asked to use their own attributes. These attributes can be generated by the assessor or selected from a predefined list and assessors can use as few or as many terms as they need to describe the sensory characteristics of the products to be judged (see section 13.1.1.2). The only requirement for the assessors is that they use these terms consistently when judging the different products. They should use the same attributes for all the products, that is, they have to define the list of attributes beforehand as well as the meaning of each attribute in their own words

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

before rating a product. Their definition of the attributes should not change during the assessments. Hence, sweetness should always represent the same notion for every product they judge. As in conventional profiling, the products are presented one after another, the tastings are done individually and each attribute is rated on a scale expressing the perceived intensity.

Compared to classic profiling with trained panels, FCP is much faster and requires less preparation. Since each assessor uses their own attributes, there is no need to train them in the meaning of the different attributes. However, they should have sufficient experience with the product category to be able to characterize the similarities and differences between the products. The lack of training also implies that the data contain more noise and that they are less accurate (Narain et al. 2003). In that respect, FCP is definitely not a substitute for classic QDA. Only when the FCP assessors have been trained individually or when they have a lot of expertise and experience with the product category can FCP be considered as a substitute for QDA. But in that case, there is relatively little gained.

Although the training argument is important, FCP is not really a substitute for classic profiling. However, there are a number of interesting applications of this technique. One example is the way a producer of yeast uses FCP. Its salespeople travel around the globe and visit vineyards. They will taste different wines with the wine makers and each will use a different vocabulary (not only in meaning but also in language). The data are entered in the laptop and processed on the spot. Ten minutes later they can discuss the results and compare the attributes they used to characterize the wines. The quality of the results can be checked by taking replicas. In this way, people from different countries who speak different languages can compare and discuss different products.

Other examples of the applications of FCP are the assessment of the behaviour of pigs in their natural environment (Wemelsfelder et al. 2001), the evaluation of health in dairy cattle, the evaluation of behaviour of wild elephants and the assessment of emotional expressions in dogs (Walker et al. 2010). Again, in these applications the quality of the individual assessors defines the quality of the results. When the product set is very different, training or experience is not essential but when the products become more similar, the assessors must be more accurate. Although we cannot train assessors for FCP, the more familiar they are with the products, the more accurate they will be.

In the past 15 years, several alternative methods for 'quick' profiling have been introduced. They have in common that training is not needed and participants can use their own attributes or that no attributes are needed at all. Examples are projective mapping/Napping® (Pagès 2005b; Pages et al. 2010; Risvik et al. 1994) and flash profiling (Dairou & Sieffermann 2002; Delarue & Sieffermann 2004). FCP can be seen as a precursor to those analyses.

13.2.2 Vocabulary

13.2.2.1 Generation of the Vocabulary

In FCP, each assessor will generate their own attribute set for characterization of the products. This is not always easy; assessors can have difficulty in generating sensory terms spontaneously. A more structured way to generate sensory terms is by using a repertory grid method (Kelly 1955; Martinez Michel et al. 2011). The repertory grid method is a way to elicit descriptive terms from assessors by means of a series of comparisons among pairs or triads of products. The assessors are presented with two or three products and asked to describe (dis)similarities between them. These descriptions are then listed and used as attributes for the assessment of the different products. The resulting data can then be analysed with generalized Procrustes analysis (GPA).

In the attribute creation phase, we can instruct the assessors to select attributes from a larger set and ask them to define the meaning. In the case of product experts, they will use relevant terms and know the meaning.

13.2.2.2 Analysis and Interpretation

Interpretation of the different terms used can also be a difficult if not impossible task. For example, a product could be described as ‘garbage’, ‘neighbour’, ‘swimming pool’ and ‘chlorinated’ by four different assessors, leaving the experimenter without a clue about their meaning. Whilst in practice this example is not very realistic, this situation can be observed when wine or perfume experts describe products using their individual vocabularies.

The interpretation of the results of FCP data is also much more complex and time consuming because of the idiosyncratic nature of the vocabulary. Instead of 20–30 predefined attributes for the characterization of each product, as in conventional profiling, there are often 5–30 attributes for each assessor and they are all different. With 15 assessors, there will be a couple of hundred different attributes. When the profiling test is carried out with 100 or more consumers, there will be a couple of thousand different attributes.

In practice, naive assessors tend to select very similar terms. In a free choice beer tasting with 80 consumers, around 3000 different terms were created but it turned out that they used only about 30 different names (they used different terms for bitter, sour, sweet, hoppy, etc. but the meanings of the different terms were very similar; there were relatively few unique terms which only meant something to the specific consumer). Still, the original set consisted of several thousand attributes and it was only afterwards that the overlap became apparent.

A related drawback is that there has to be a separate attribute list for each assessor. Both computerized and paper and pencil execution of the test require a lot more preparation.

13.2.3 Notation

Let's denote n as the number of products tested and K as the number of panellists (Ss). In QDA, all panellists describe the products on the same list of attributes, so let's denote m as the number of attributes. In FCP, each assessor creates his/her own list of attributes so let's call m_k the number of attributes used by the consumer k to describe the products.

13.2.4 Data Gathered

Figure 13.1 shows an example of a classic profiling data set and Figure 13.2 shows a FCP data set.

In FCP, each assessor uses their own words. In that situation, panellist k will be associated to a set of m_k attributes. Hence, the data are organized according to Figure 13.2.

In the case of FCP, each table contains different attributes. As the attributes differ in their meaning, one could eventually consider adding chemical or instrumental data as an extra table to the data set. Hence, the additional matrix (let us call this one Chemical_{K+1}) would represent chemical or instrumental measurements for the same set of products in the same way as the description of the products by consumer k. In that case, the additional table is treated as an extra assessor and the chemical or instrumental measures as the attributes and the user can link the sensory attributes to the chemical or

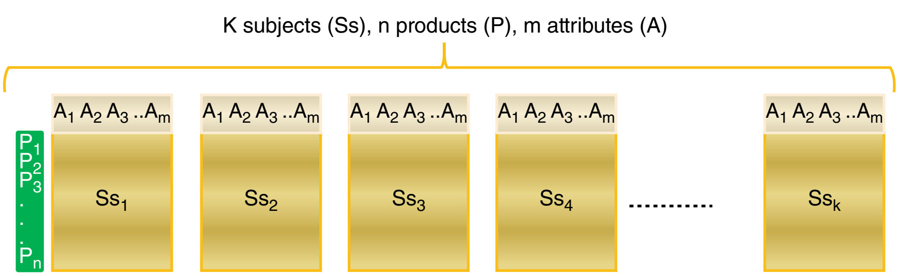


Figure 13.1 Data set from classic QDA profiling.

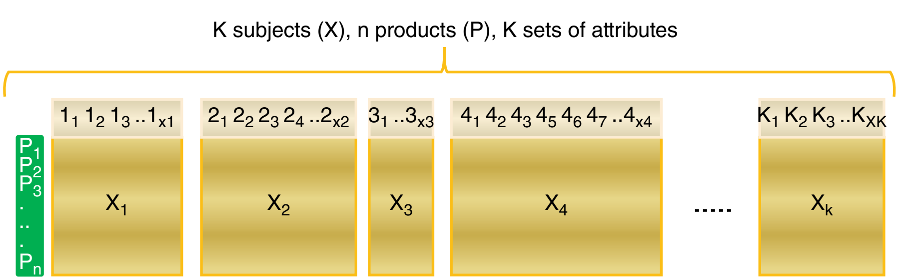


Figure 13.2 Data set from free choice profiling.

instrumental attributes. The only requirement in FCP is that every set (matrix) rates the same set of products.

13.2.5 Analysis

The data collected with QDA (see Figure 13.1) can be analysed in the classic way. For each product, the averaged attribute ratings can be computed, we can use analysis of variance to see if products differ from each other on the different attributes and we can compute the product space by means of principal component analysis (PCA). The data collected with FCP (see Figure 13.2) cannot be analysed in a normal way as each subtable is different (both in size and in kind of attributes). Since each assessor has been using different attributes, there is no way to average the data. In fact, FCP is another example of multi-block data, and several statistical analyses are now available; among them, we can cite GPA (Gower 1975), Indscal (Caroll & Chang 1970) or multiple factor analysis (MFA) (Escofier & Pagès 1994; Pagès & Husson 2001).

Historically, it was the availability of GPA that led to the creation of free choice profiling in the 1980s. A review of the GPA method is presented in section 13.3.

13.2.6 Derivative of FCP: The Flash Profile

Flash profiling (FP) is an alternative sensory analysis technique adapted from FCP to understand the sensory positioning of products (Dairou & Sieffermann 2002; Delarue & Sieffermann 2004). Untrained assessors select their own terms to describe and evaluate a set of products simultaneously, and then rank the products for each attribute that they created. They are asked to focus on descriptive terms, not hedonic terms. In contrast to FCP, assessors do not have to rate the intensity on each attribute but they must rank the products for each selected term. All products are presented simultaneously, in contrast to FCP where the products are presented one after another.

Here again, as each assessor uses his/her own descriptive terms to evaluate the products, we are in a situation of multi-block analysis. Hence, the individual sensory maps can be analysed using GPA or other methods to create a consensus configuration.

Although this method allows for 'quick' profiles, it is not suited for every type of product. Hot products like coffee, soups or products which have strong adaptation effects are less appropriate.

13.3 Generalized Procrustes Analysis

Free choice profiling was 'invented' in the 1980s by Tony Williams and Gillian Arnold as a quick alternative to classic profiling with trained panellists and became possible thanks to a statistical technique, Procrustes analysis.

13.3.1 History

Procrustes was a character in Greek myth. An innkeeper who plied his trade in Attica, he put his victims on an iron bed. If they were longer than the bed, he cut off their feet. If they were shorter, he stretched them (Figure 13.3).

The Procrustes analysis does something similar to data as it uses three different transformations (translation, isotropic scaling and rotation/reflexion) to make individual configurations as similar as possible to each other.

Procrustes analysis was originally developed to compare two different solutions from factor analysis (one-sided orthogonal Procrustes rotation) (Hurley & Cattell 1962; ten Berge 1977). Schönemann (1968) and Schönemann and Carroll (1970) extended it to a two-sided orthogonal Procrustes rotation with scaling and Kristof and Wingersky (1971) extended it to more than two configurations.

Gower (1975) introduced a GPA with a scaling factor and analysis of variance (ANOVA). Williams and Langron (1983) realized this technique made it possible to match and average the product configurations from individuals who use their own (different) attributes into a common product space.

13.3.2 The Birth of FCP

Thanks to GPA, it was not necessary to train people to use the different attributes in the same way. Statistics would take care of the differences in meaning and the training phase could be skipped. The procedure was named 'free choice' profiling because every assessor was free to select their own attributes.

In the period 1983–1985, GPA was introduced in sensory evaluation for the analysis of free choice profiling data (Arnold & Williams 1986; Langron 1983; Williams & Langron 1983).

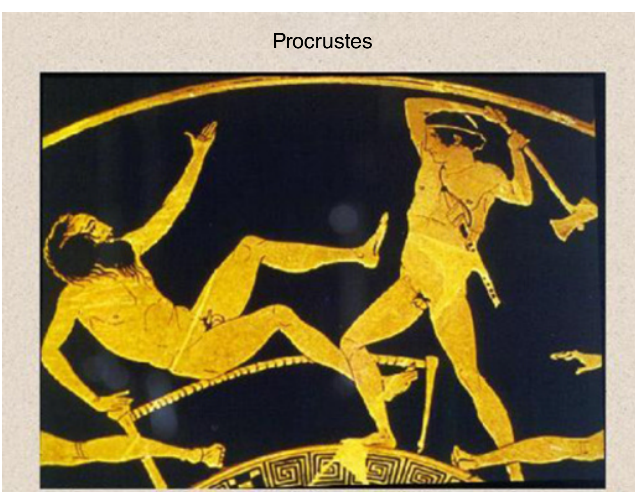


Figure 13.3 The Procrustes transformation in action.

13.3.3 Methodology

13.3.3.1 Scope and Application

The aim of the free choice profiling test is to obtain ratings from a fixed set of products (preferably 5–15) from different individuals. Each individual assessor can invent and use as many attributes as he/she needs to describe the sensory characteristics of a set of products. The number of assessors can vary between three and hundreds but for practical reasons it is recommended to restrict the number to 20–30 at most.

For each assessor, an individual product space is generated on the basis of the attribute ratings and the individual product spaces are summarized into a group average space. In this group average space, the individual attribute ratings are projected.

13.3.3.2 Principle of the Test

Each assessor will be presented with a representative set of products and asked to generate a list of attributes to characterize the product set. There are different ways to obtain the attributes but in the end, each assessor will have a list of relevant attributes and a score sheet. Next, they will be presented with a number of products (in a sequential monadic order taking care of order and carry-over effects) and requested to rate each product on the different attributes. The ratings can be category scales, line scales or magnitude estimates.

This results in K matrices (one for each assessor) with n products and m_k attributes (see Figure 13.2). The procedure is identical to the procedure for classic profiling, except that each assessor uses his/her own attribute list.

13.3.3.3 Statistical Analysis

The FCP data are multi-block data which can be analysed with GPA. For a mathematical explanation of GPA the following references are available.

- Gower (1975): GPA
- Dijksterhuis and Punter (1990): the interpretation of GPA and allied methods
- Dijksterhuis (1997): multivariate data analysis in sensory and consumer science
- Naes et al. (2010): statistics for sensory and consumer science

The principle of GPA is quite simple. Since each assessor is using different attributes, the data cannot be averaged. Instead, each individual matrix is used to compute an n-dimensional product space. These individual product spaces are averaged into an average group space.

The analysis can be defined as follows.

- Consider K configurations of P products in n^K-dimensional spaces (for each assessor k, we have a n^k-dimensional product space).
- Represent these K configurations in a common space (called consensus space) while minimizing the goodness of fit criterion with the aid of only three transformations:

- translation (move the centroids of each configuration to a common origin)
- isotropic scaling (shrink or stretch each configuration isotropically to make them as similar as possible)
- rotation/reflection (turn or flip the configurations).

Algorithmically, GPA is an iterative procedure. It starts with a common space (it can be the PCA space, the space related to the first panellist or the results of the MFA performed on the individual tables in the case of Procrustes MFA) on which all individual configurations are matched. A new consensus space is then calculated and the procedure starts again, until a maximum number of iterations is reached or until the gain in goodness of fit is lower than a certain level.

The following steps are taken:

- GPA performs the transformations on each set
- individual configurations are averaged when they are as similar as possible
- the resulting high-dimensional space is reduced by means of PCA to a lower dimensionality
- the total variance in the data is partitioned over sets, objects or dimensions.

Example: let's suppose we have four assessors (consumer 1, 2, 3 and 4), who rated *P* products (rows) on different number of attributes (columns). The data matrices are shown in Figure 13.4. Since they are of different size, columns of zero's can be added to equalize them ('zero padding'; Figure 13.5).

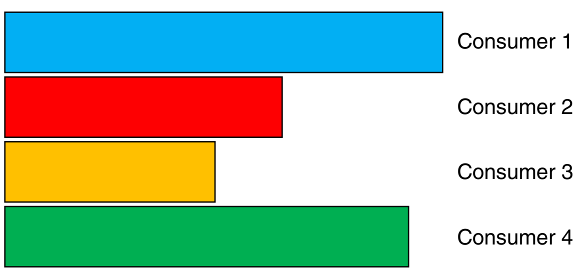


Figure 13.4 FCP data matrix for four consumers with different number of attributes.

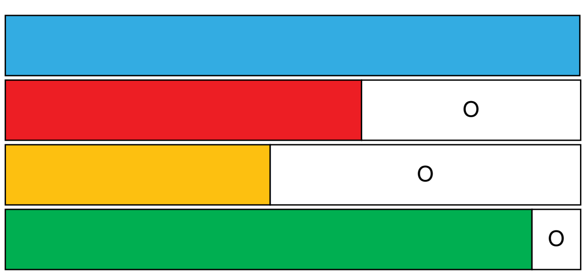


Figure 13.5 Zero padding of data matrix to make them of equal size.

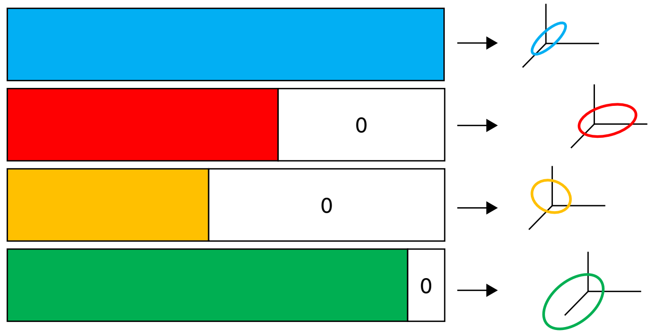


Figure 13.6 Configurations for the individual consumers.

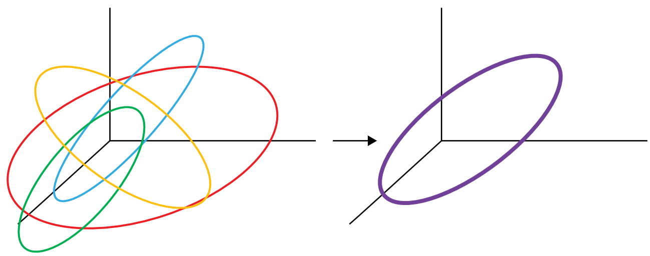


Figure 13.7 Creating the averaged consensus space on the basis of the individual configurations.

The next step is to compute n^k-dimensional product spaces for each assessor (Figure 13.6). This results in four product spaces, represented by the four clouds. Next, we apply the three Procrustes transformations (translation, isotropic scaling and rotation/reflexion) to these product spaces in order to make them as similar as possible. Note that these transformations are isotropic, which means that the individual clouds have the same shape before and after transformation. The transformation only helps for the adjustment of the spaces. This is different from the procedure used in MFA where the clouds can be stretched more in some dimensions than in others. A more systematic comparison of MFA and GPA has been proposed by Pagès (2005a).

For each assessor, a weight is computed which indicates how much their data had to be shrunk or stretched to fit the consensus space. The final result can then be summarized in a consensus configuration (the bold circle on the right of Figure 13.7).

Next, the individual spaces and the individual attributes (for example, attribute k from assessor j) can be projected into the consensus space. In this way, you can see ‘consensus’ in different attributes by looking at their respective projections in the space.

13.3.4 Extra Statistical Calculations: Procrustes Analysis of Variance (PANOVA)

The total amount of variance in the data is rescaled to 100% and partitioned over assessors (sets), objects and dimensions. In formula:

$$V_T = V_C + V_W$$
$$V_C = V_I + V_O$$

in which:

V_T = total variance
V_C = consensus variance
V_W = within or error variance
V_I = variance explained by the first Q dimensions of the consensus space
V_O = unexplained variance (the part associated with the higher dimensions)

This can be represented schematically (Figure 13.8).

In contrast with PCA, the amount of variance explained in itself does not give an indication for the significance or fit of the final solution. So to check for the significance of the result, a test is performed to check whether the percentage consensus variance could be obtained in random situations. To do so, random situations are created by randomly permuting the rows of each individual configuration. Hence, we position ourselves in a situation where consumer would return the exact same configurations, but after redistributing randomly the labels of the products on the space. A consensus space between these random configurations is then defined, and the percentage consensus variance is measured. This process is

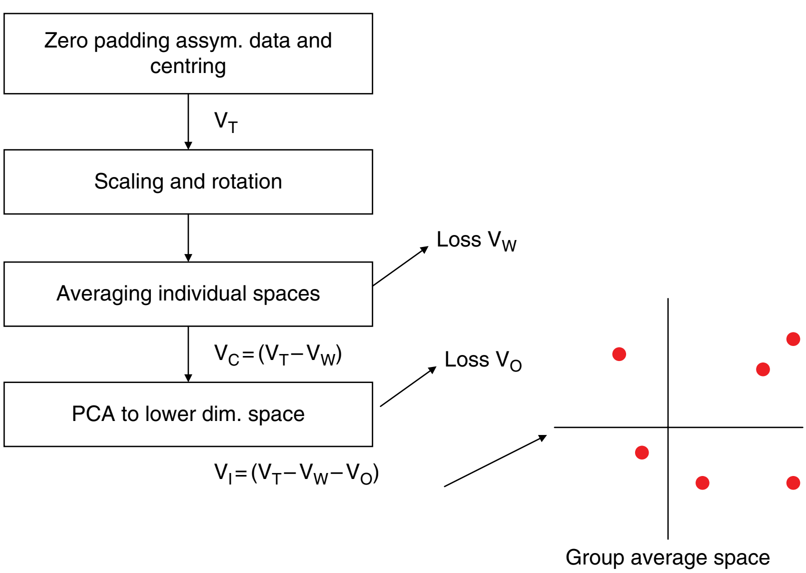


Figure 13.8 Schematic representation of the PANOVA.

repeated a large number of times (in practice, 200 times), and the distribution of the percentage consensus variance under H_0 (random situation) is measured. The real percentage consensus variance is positioned on that distribution and is compared to the 95th and 99th percentile of the percentage consensus variance.

When the actual percentage explained variance is larger than the highest value found in the permuted sets, we can conclude that the solution is significant.

13.3.5 GPA and Computers

In 1987, Arnold published a GPA macro in Genstat for the analysis of FCP data and in 1989 Schlich published an SAS macro.

The software for GPA was not easily available for most sensory practitioners until 1988 when the first version of Procrustes-PC was introduced by OP&P (Procruster-PC v1.0, OP&P, Utrecht). Although the first release was not very fast (analysis of a dataset with 12 assessors and six products would take 4–6 hours), it made it possible for most practitioners to run GPA on a PC. Today, GPA is implemented in Senstools.Net (http://senstools.com/, OP&P), XLSTAT (www.xlstat.com, Addinsoft), SensoMineR (the R-package, http://sensominer.free.fr) and several other statistical software packages.

13.4 Case Study

The example is taken from the R-package SensoMineR.

13.4.1 Material

Twelve luxury perfumes* were evaluated by six experts (students from a French perfume school). Each expert used their own vocabulary (Table 13.1). The individual vocabularies were created in focus groups; each expert chose their own attributes. In the final test, the perfumes were presented sequential monadically in balanced order in one 60-minute session. For J'Adore, two different variants were included: an eau de toilette and an eau de parfum. Although they are not identical, they can be regarded as replicas. Perfumes were rated on 100 mm line scales with anchors (weak–strong) on 10 mm and 90 mm.

13.4.2 Results

13.4.2.1 Consensus Space

The resulting consensus space (GPA group average space) is shown in Figure 13.9.

The first dimension differentiates J'Adore and Pleasures from Shalimar and Angel. The two J'Adore variants are very close together in this space, showing a good repeatability of the assessors. The second dimension differentiates Aromatics Elixir and Chanel No. 5 from Angel and Lolita Lempicka.

* All product names mentioned in this chapter are tradenames, trademarks or registered trademarks of their respective owners.

Table 13.1 Attribute lists for the six assessors.

Products	Attributes					
	assessor 1	assessor 2	assessor 3	assessor 4	assessor 5	assessor 6
Angel	Flower	Oriental	Flowerish	Flowerish	Flower	Fruity
Aromatics Elixir	Woody	Pleasant	Green	Pleasant	Green	Musk
Chanel No. 5	Fruity	Flowerish	Aldehyde	Amber	Fruity	Green
Cinema	Vanilla	Fruity	Pleasant	Green	Woody	Orange flower
Coco Mme	Fruity	Woody	Vanilla	Fruity	Musk	Rose
J'Adore EP	Spicy	Luxurious	Woody	Chypre	Pleasant	Fruity
J'Adore ET	Green	Disgusting	Musk	Light		Patchouli
L'Instant	Animal					Amber
Lolita Lempicka	Pleasant					
Pleasures	Leather					
Pure Poison	Aldehyde					
Shalimar	Powder					

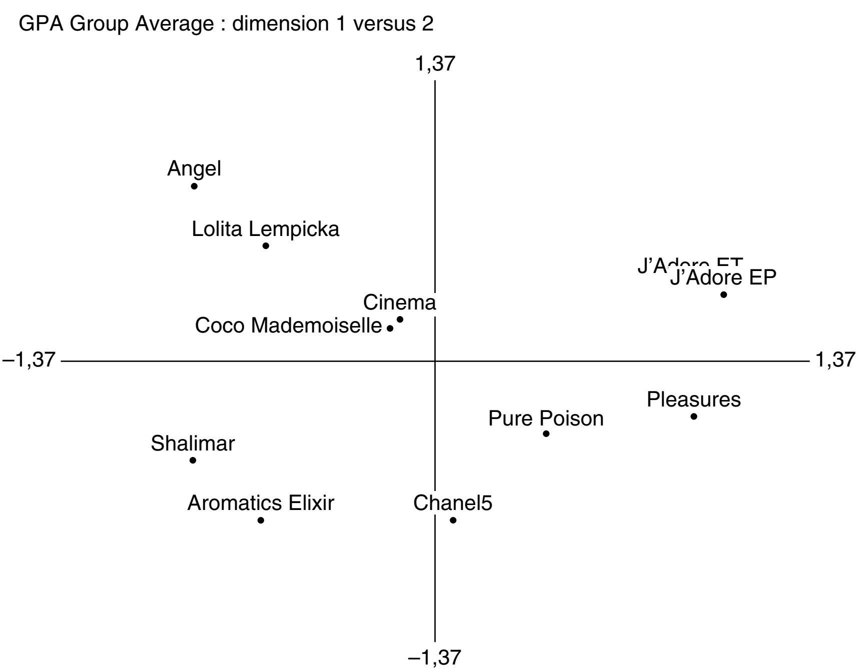


Figure 13.9 Consensus space (dimension 1 accounts for 32.8% and dimension 2 for 18.2% of the variance).

13.4.2.2 Significance of the Consensus Space (Permutation Test)

The main result from the analysis is the consensus space. But before going further in the interpretation of the results, we have to make certain that the results (in terms of the amount of variance explained by the data) are meaningful or significant. In contrast to the classic PCA, the percentage of variance accounted for is not a useful measure to assess the fit of the GPA model. To find out if the outcome of the GPA is meaningful and not random, a permutation test is carried out. The actual result is compared with the results based on many random computations. Several hundred analyses are run on the same data set with randomly permuted product labels. This procedure does not change the numerical values but removes the 'meaning' from the data. The percentage of variance accounted for in the real data set is compared with the distribution of the variances from the permuted data sets. When the percentage in the real data set is higher than the 95th or 99th percentile in the distribution of the permuted data sets, we can conclude that the actual result is not random but significant.

The permutation test for the perfume data shows that the total variance accounted for in the real data set is 73%. The 95th percentile from the permuted data set is 64% and the 99th percentile is 66%. The conclusion is that the result is not random.

13.4.2.3 PANOVA Results and Individual Weights

Next, we inspect the PANOVA. The total variance in the data is rescaled to 100% and partitioned over dimensions, products and assessors. Table 13.2 shows the PANOVA results for the first three dimensions.

Table 13.3 shows the distribution of the explained variance for the 13 products by dimension. This table shows us that the first dimension is characterized by J'Adore, Pleasures, Angel and Shalimar. The second dimension is characterized by Angel, Aromatics Elixir and Chanel No. 5 and the third dimension is characterized by L'Instant, Cinema and Aromatics Elixir.

Table 13.4 shows the distribution of the explained variance by assessor and dimension. This table shows us that assessors 1, 2 and 4 show the highest discrimination in the first dimension (they account for most of the variance). Assessors 1 and 2 also have the highest explained variance in dimension 2.

Next, we look at the individual weights (Figure 13.10). Here, it appears that assessor 1 and assessor 3 have a slightly different point of view compared to the others (different scaling behaviour). Assessor 1 had to be shrunk and assessor 5 had to be stretched in order to fit in the consensus space. However, the differences are quite small.

Table 13.2 PANOVA results per dimension (%).

	Explained	Residual	Total
Dimension 1	33	5	38
Dimension 2	18	4	22
Dimension 3	9	4	13
Total	60	13	73

Table 13.3 Explained variance by product and dimension.

	Dimension 1	Dimension 2	Dimension 3
Angel	4.63	4.33	0.76
Aromatics Elixir	2.42	3.63	1.96
Chanel No. 5	0.03	3.61	0.01
Cinema	0.10	0.24	2.28
Coco Mme	0.16	0.15	0.06
J'Adore EP	6.69	0.64	0.40
J'Adore ET	5.37	0.84	0.05
L'Instant	0.07	0.24	2.52
Lolita Lempicka	2.29	1.88	0.18
Pleasures	5.36	0.44	0.23
Pure Poison	0.98	0.77	0.29
Shalimar	4.70	1.41	0.29
Total explained	32.8	18.2	9.0

Table 13.4 Explained variance by assessor and dimension.

	Dimension 1	Dimension 2	Dimension 3
Assessor 1	7.56	4.09	0.93
Assessor 2	7.59	4.11	0.82
Assessor 3	4.10	2.89	1.97
Assessor 4	8.49	1.43	0.56
Assessor 5	2.50	2.19	2.57
Assessor 6	2.59	3.47	2.18
Total explained	32.8	18.2	9.0

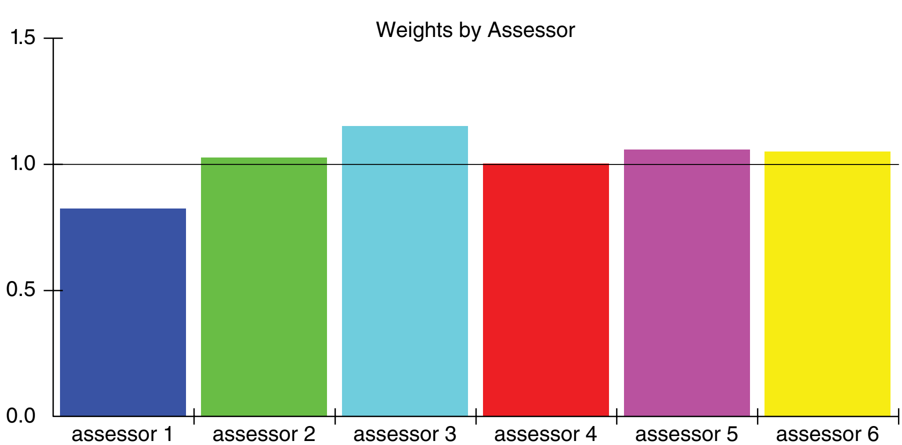


Figure 13.10 Individual weights.

13.4.2.4 Interpretation of the Consensus Space

Since each assessor used their own attributes, we cannot plot average attributes ratings in the product space, we can only inspect the attributes from each individual. The next figures show the attributes for assessor 1 and 5 (Figures 13.11 and 13.12).

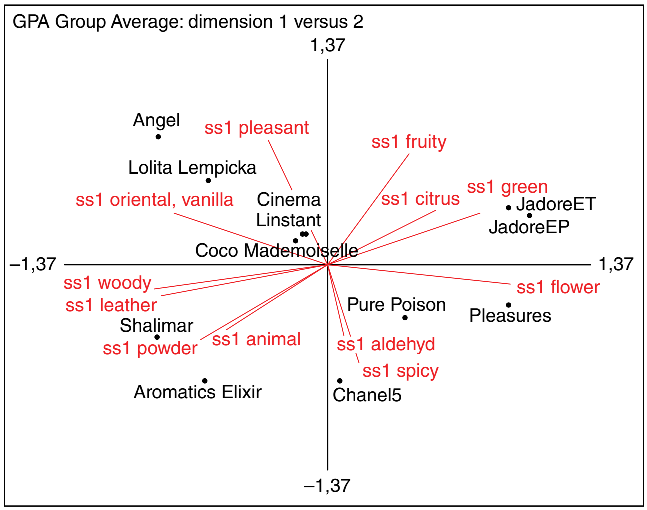


Figure 13.11 Consensus space with the attributes of assessor 1 (ss1).

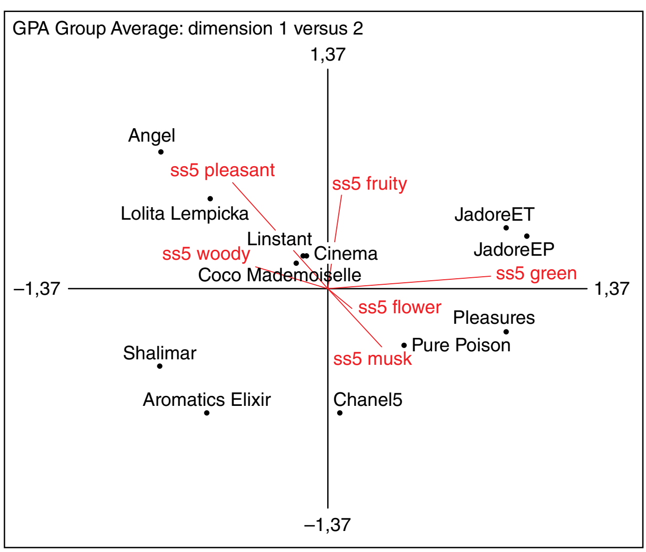


Figure 13.12 Consensus space with the attributes of assessor 5 (ss5).

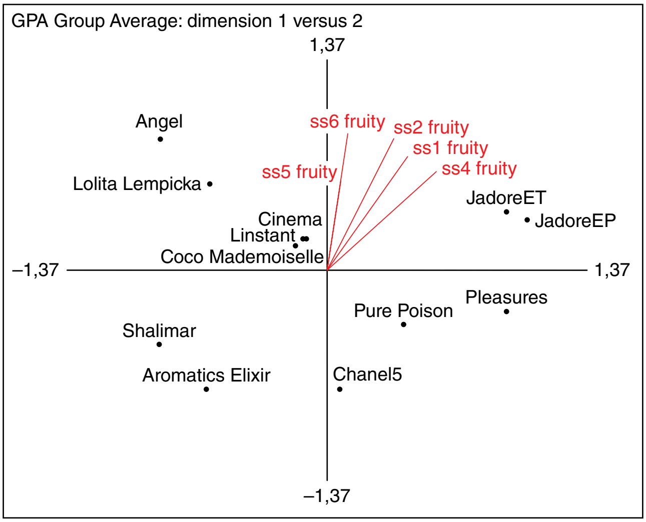


Figure 13.13 Consensus space with the attribute 'fruity' for different assessors.

Assessor 1 characterizes J'Adore as citrus and green, Chanel No. 5 as aldehyde and spicy, Shalimar as leather and woody and Angel and Lolita Lempicka are pleasant and oriental.

Assessor 5 characterizes J'Adore also as citrus and green. Chanel No. 5, Pure Poison and Pleasures are characterized as musk and flower, Coco Mme and Shalimar are woody and Angel and Lolita Lempicka are pleasant.

Since the assessors also use similar terms, we can check if they give the same meaning to these terms (or whether they describe the same products with these terms). Figure 13.13 shows the results for 'fruity'. As can be seen, the different assessors use the attribute 'fruity' in a very similar way; the perfumes in the top quadrants (Angel, Lolita Lempicka and J'Adore) are fruity.

Figure 13.14 shows the results for 'musk'. There is less agreement for this attribute, although all three assessors agree that the perfumes on the right side of the space are characterized by musk but some would point to J'Adore while others would point to Pure Poison or Pleasures.

13.5 Options/Practical Considerations

Free choice profiling delivers:

- a perceptual space which shows the relationship between the products in terms of sensory perception
- an interpretation of this space in terms of the descriptors used by the different assessors
- an insight in the way different assessors describe the underlying dimensions of the product space.

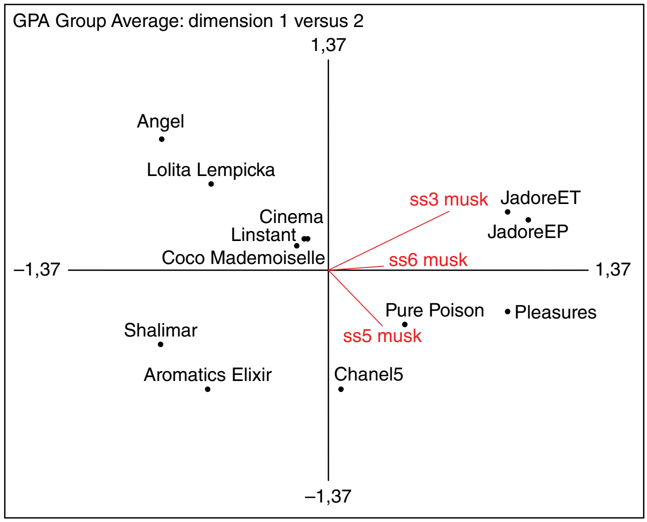


Figure 13.14 Consensus space with the attribute 'musk' for different assessors.

In spite of the promises in the 1980s, FCP has not become an alternative to classic profiling. The main reason is the complexity of the interpretation of results (averaged results are not possible; interpretation is more cumbersome as you have to check assessor by assessor). However, there are a number of situations in which the method is indispensable, for instance when comparing results from different countries (small groups of assessors who rate a number of products in their own language), wine tastings in the field or tastings with a small group of chefs or experts.

Although FCP proved not to be very practical as a replacement for trained panels, the statistical technique for the analysis (GPA) is extremely useful for conventional profiling data. It will show the individual results and the different uses of the scales and we can inspect the agreement between the assessors on an attribute-by-attribute basis.

One very useful application is during training of a sensory panel. By treating the results of the panellists as free choice data, their behaviour can be monitored on an attribute-by-attribute basis. Compared to the ANOVA approach or other panelcheck methods, GPA gives a more visual result.

13.6 Summary

This chapter describes the origin of generalized Procrustes analysis (GPA) and the application of the technique in free choice profiling. In free choice profiling, each assessor can use their own vocabulary to characterize a set of products. Although FCP is not really a substitute for QDA, analysing QDA data with GPA can be very helpful because it shows how each individual is using the attributes. FCP is very useful for situations with a small number of subjects who speak different languages. The GPA technique is also very useful to compare the product spaces from multivariate analyses (for instance, the product space of expert and consumers).

13.7 Future Considerations

Free choice profiling is not restricted to sensory applications, it can be applied in every situation where the same set of objects is described by different sets of attributes. Examples are the comparison of instrumental and chemical measurements, comparison of panels from different countries or analysis of animal behaviour. FCP is not used that often in sensory analysis or consumer research because of the difficulties in interpretation of the large number of attributes.

Free choice profiling has not become an alternative to classic profiling but the application of GPA has made it possible to treat classic profiling data as free choice. When assessors use the same vocabulary, there is no assurance that they attach the same meaning to the attributes. Even trained assessors can differ in their interpretation of the attributes. Treating the data as free choice makes it possible to compare individual response behaviour and to assess agreement between assessors for each of the attributes used.

References

Arnold, G.M. (1987) A Generalized Procrustes macro for sensory analysis. *Genstat Newsletter*, **18**, 61–80.

Arnold, G.M. & Williams, A.A. (1986) The use of generalized Procrustes techniques in sensory analysis. In: Piggot, J.R. (ed.) *Statistical Procedures in Food Research*. Amsterdam: Elsevier, pp. 233–253.

Caroll, J.D. & Chang, J.J. (1970) Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart–Young" decomposition. *Psychometrica*, **35**, 283–319.

Dairou. V. & Sieffermann. J.M. (2002) A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, **67**, 826–834.

Delarue, J. & Sieffermann, J.M. (2004) Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality & Preference*, **15**, 383–392.

Dijksterhuis, G.B. (1997) *Multivariate Analysis in Sensory and Consumer Science*. Washington, DC: Food and Nutrition Press.

Dijksterhuis, G.B. & Punter, P.H. (1990) Interpreting Generalized Procrustes Analysis 'Analysis of variance' tables. *Food Quality & Preference*, **2**, 255–265.

Escofier, B. & Pagès, J. (1994) Multi factor analysis (AFMULT package). *Computational Statistics & Data Analysis*, **18**, 121–140.

Gower, J.C. (1975) Generalized Procrustes Analysis. *Psychometrica*, **40**, 33–51.

Hurly, J.R. & Cattell, R.B. (1962) The Procrustes program: producing direct rotation to test a hypothesized factor structure. *Behavioural Science*, **7**, 258–262.

Kelly, G. (1955) *The Psychology of Personal Constructs*. New York: Norton.

Kristof, W. & Wingersky, B. (1971) Generalization of the Orthogonal Procrustes Rotation Procedure for More than Two Matrices. Proceedings of the 79th Annual Convention of the American Psychological Association, p. 89.

Langron, S.P. (1983) The application of Procrustes statistics to sensory profiling. In: Williams, A.A. & Atkin, R.K. (eds) *Sensory Quality in Foods and Beverages*. Chichester: Ellis Horwood, pp. 89–95.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evaluation of Food*, 2nd edn. New York: Springer.

Martínez Michel, L., Punter, P.H. & Wismer, W.V. (2011) Perceptual attributes of poultry and other meat products: A repertory grid application. *Meat Science*, **87**, 349–355.

Meilgaard, M.C., Civille, G.V. & Carr, T.B. (2007) *Sensory Evaluation Techniques*, 4th edn. London: Taylor and Francis.

Naes, T., Brockhoff, P.B. & Tomic, O. (2010) *Statistics for Sensory and Consumer Science*. Chichester: John Wiley and Sons.

Narain, C., Paterson, A. & Reid, E. (2003) Free choice and conventional profiling of commercial black filter coffees to explore consumer perceptions of character. *Food Quality & Preference*, **15**, 31–41.

Pagès, J. (2005a) Analyse factorielle multiple et analyse procrusteenne. *Revue de Statistique Appliquée*, **4**, 61–86.

Pagès, J. (2005b) Collection and analysis of perceived product inter-distances using multiple factor analysis: application to the study of 10 white wines from the Loire Vallery. *Food Quality & Preference*, **16**, 642–649.

Pagès, J. & Husson, F. (2001) Inter-laboratory comparison of sensory profiles: methodology and results. *Food Quality & Preference*, **12**, 297–307.

Pagès, J., Cadoret, M. & Lê, S. (2010) The sorted napping: a new holistic approach in sensory evaluation. *Journal of Sensory Studies*, **25**, 637–658.

Risvik, E., McEwan, J.A., Colwill, J.S., Rogers, R. & Lyon, D.H. (1994) Projective mapping: a tool for sensory analysis and consumer research. *Food Quality & Preference*, **5**, 263–269.

Schlich, P. (1989) A SAS/IML Program for Generalized Procrustes Analysis. Proceedings of the SAS European User Group International conference, Germany.

Schönemann, P.H. (1968) On two sided orthogonal Procrustes problems. *Psychometrica*, **33**, 19–33.

Schönemann, P.H. & Carroll, R.M. (1970) Fitting one matrix to another under choice of a central dilation and a rigid motion. *Psychometrica*, **35**, 245–255.

Stone, H. & Sidel, J.L (2004) *Sensory Evaluation Practices*, 3rd edn. Amsterdam: Academic Press.

ten Berge, J.F.M. (1977) Orthogonal Procrustes rotation for two or more matrices. *Psychometrica*, **42**, 267–276.

Walker, J., Dale, A., Waran, N., Clarke, N., Farnworth, M. & Wemelsfelder, F. (2010) The assessment of emotional expression in dogs using a free choice profiling methodology. *Animal Welfare*, **19**, 75–84.

Wemelsfelder, F., Hunter, T.E.A., Mendl, M.T. & Lawrence, A.B. (2001) Assessing the 'whole animal': a free choice profiling approach. *Animal Behaviour*, **62**, 209–220.

Williams, A.A. & Langron, S.P. (1983) A new approach to sensory profile analysis. In: Piggot, J.R. (ed.) *Flavour of Distilled Beverages*. Chichester: Ellis Horwood, pp. 219–224.

CHAPTER 14

Flash Profile Method

Wender L.P. Bredie, Jing Liu, Christian Dehlholm and Hildegarde Heymann

14.1 Introduction

Descriptive sensory analysis has been a very successful approach in determining relevant sensory differences between products. Methods like quantitative descriptive analysis (QDA) (Stone et al. 1974) and the more elaborate Spectrum™ Method as well as generic variations on these methods have gained popularity in product development and research, when specific and precise descriptions of sensory properties are needed (Meilgaard et al. 2006). However, such descriptive methods are costly and time-consuming and not needed for all situations. The flash profile method (Sieffermann 2000) is one of the attractive alternatives for acquiring rapid sensory evaluations on products.

In conventional descriptive sensory analysis, extensive training of the panel will in many cases prevent inconsistencies in the use of the sensory vocabulary. Hereto, different approaches have been adopted. One could focus on a high degree of training on specific sensory attributes and scales with reference materials as in the Spectrum Method (Drake & Civille 2002; Meilgaard et al. 2006) or utilize panel discussions with references for sensory attributes to reach cognitive clarity on the sensory vocabulary between the panel members as in QDA. Conventional descriptive approaches often follow combinations of Spectrum Method and QDA in order to adapt to specific research goals. All of these methods clearly invest time in training of the sensory subjects in order to be able to understand the meaning of the subsequent sensory profiles and to enable the correlation of results to, for instance, instrumental analysis.

The flash profile method deviates from conventional profiling methods by giving a high degree of freedom to the subjects in using their own sensory vocabularies. Thus, the method needs less time for subjects to reach a consensus vocabulary and training on the use of intensity scales. As a consequence, the flash profile method relies strongly on subsequent data analysis to achieve

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

a panel consensus space for interpretation. Whereas several studies (Dairou & Sieffermann 2002; Delarue & Sieffermann 2004; Dehlholm et al. 2012a) have shown that the product space is fairly robust when using the flash profile method, interpretation of the sensory space may be more difficult. In this chapter we will give an overview of the flash profile method and place it in a more general context of methods dealing with individual vocabularies in descriptive sensory analysis.

14.2 Theoretical Framework

The theoretical arguments for using individualized vocabularies as included in the flash profile method can be founded in Kelly's personal construct theory (Kelly 1955). This theory was based on the idea that the construction system used by a person to perceive, understand, predict and control the surrounding world is composed of a finite number of dichotomous constructs. Kelly originally developed the repertory grid method (RGM) (1955) in the field of psychology in order to establish individual constructs for objects. The main idea of this structured interrogation technique is to get subjects to define their own constructs (attributes) by describing the ways in which elements and their associated meanings vary. The first application of this method in sensory science occurred in the early 1980s. Olson (1981) introduced the idea of using this technique to elicit the characteristics of food products. In this study, subjects were asked to define a scale to measure the amount of each construct perceived in the objects. Thus, each individual proposes his own constructs and scales for evaluating the objects. Applications of the RGM have been extensive in the fields of cognitive science (Feixas et al. 2010), food science in order to generate sensory product attributes (Berg & Rumsey 2006) and to study consumer conceptualizations in market research (Frost & Braine 1967; Swahn et al. 2010; Veinand et al. 2011; Yan et al. 2011).

The flash profile method promotes an individual approach in eliciting a language to rank product perceptions but is much less rigid than the RGM in the selection of product attributes. In the flash profile method, the panel consensus is determined subsequently and it is not necessarily clear what the attributes from the individual panel members mean in the consensus space. As can be seen in Table 14.1, understanding of the true meaning of a sensory attribute can be difficult and confounded with different distinctions.

In the ideal case, the panel members use a term in relation to their product evaluation in the same way and a true consensus is obtained. But in other cases, discrepancies may occur, including correspondence, where panel members use a different term for the same percept; conflict, where panellists use the same term for a different percept; and contrast, where panellists use a different term for different percepts. Gaines and Shaw (1993) highlighted the importance of

Table 14.1 Different ways in which the sensory vocabulary of a panel can relate to the distinctions between the sensory stimuli.

		Sensory terminology	
		Same	**Different**
Stimulus distinctions	Same	**Consensus** People use sensory attributes for distinctions in the same way	**Correspondence** People use different sensory attributes for the same distinction
	Different	**Conflict** People use same sensory attributes for different distinctions	**Contrast** People differ in sensory attributes and distinctions

Source: adapted from Lorho (2010)

recognizing each of these situations to better measure the level of commonality and idiosyncrasy between individual conceptual systems. In conventional profile methodology, group sessions with the products available are very important because they stimulate discussion and motivate panellists to explain their own perceptions to the panel members. But even in this approach a true panel consensus is not guaranteed.

Free choice profiling (FCP) (Williams & Arnold 1984; Williams & Langron 1984) emphasized the use of individual vocabularies. This method gained popularity in the 1980s and 1990s, and was used with analytical sensory panels but mostly to generate product-relevant vocabularies with consumers. FCP has lost popularity as a descriptive analysis method mainly because of the difficulty in providing a rigid background for interpreting the descriptive sensory space. More details on this method are given in Chapter 13.

The flash profile method can be seen as a more recent variant of the free choice profiling method. The flash profile combines the individual attribute elicitation of the FCP method with comparative ranking of samples on these attributes. The comparison of all products at the same time, with the possibility to go back to previous assessments, is claimed to remove the need for product familiarization and training with the attributes. The process of elicitation and training is thereby drastically reduced. By making use of expert subjects, a flash profile can be obtained over a very few sessions.

No reports have been published on how the flash profile method compares to free choice profiling and other individual vocabulary methods such as the RGM. However, since the flash profile relies on experienced or expert subjects, it may be expected that these subjects have already a considerable frame of reference for naming sensory perceptions. The flash profile method has been compared with conventional descriptive sensory profiling on products that are typical for an industry situation of brand comparisons (Delarue & Sieffermann 2004). The study concluded that the flash profile could discriminate between products as well as the conventional profiling method. However, the semantic structure was more complex for the flash profile method. In another study on fruit jams (Dairou & Sieffermann 2002), where the product differences were rather obvious, a similar conclusion was reached. The flash profile was significantly faster than conventional profiling, fairly robust and thus lowered the cost of the evaluation. Other studies have shown similar findings (Blancher et al. 2007; Dehlholm et al. 2012a).

14.3 Overview of the Flash Profile Method

14.3.1 Aim

The flash profile aims to provide a rapid and reliable sensory profile of a set of products according to their major sensory differences. The method relies on individual vocabularies and comparative assessments to shorten the time needed to perform the analysis.

14.3.2 Panel

The flash profile was originally developed to be carried out by experienced panellists. The panel could be sensory experts and/or professionals who have well-established sensory experiences with the products. For the number of subjects, the original flash profile method used eight panel members (Dairou & Sieffermann 2002) but later applications of the method reported 6–12 panel members to be sufficient (Albert et al. 2011; Delarue & Sieffermann 2004; Dehlholm et al. 2012a), whereas Delarue (2014) stated that a panel of 4–5 members is the minimum to yield a stable configuration on sensory product differences.

Recent studies have also explored the application of the flash profile method with consumer panels. In these studies, 24–50 consumers were considered sufficient (Moussaoui & Varela 2010; Varela & Ares 2012, 2014). Depending on the objectives of the study, the number of consumers could even range to 200 (Ballay et al. 2006). Consumer panels, however, require special explanation on the use of the method and instruction on using product attributes. The interpretation of the attribute space from consumer panels may also be difficult due to the extensive and unclear use of sensory terminology.

14.3.3 Process

A schematic overview of the main steps in the flash profile method is given in Figure 14.1. The sessions typically take place in computerized sensory booths without interaction with the other panel members. A whole product set is presented simultaneously to the panel after they are given a brief outline of the test procedure.

In the first session, each subject creates his/her own provisional list of attributes when confronted with the series of stimuli/products. The subjects are asked to list the sensory characteristics that best describe the differences between the

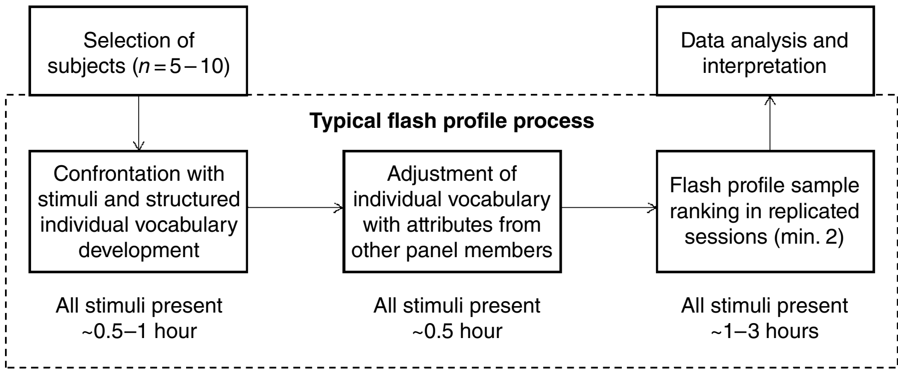


Figure 14.1 Schematic overview of the steps in the original flash profile method.

samples and are instructed to avoid the use of hedonic terms. To assist in vocabulary development, the subjects are instructed to group terms according to sense modality such as appearance, sound, texture, taste, etc. They can re-examine the products as often as they wish and take as much time as they need. Subjects are instructed to pause between exposures to the stimuli/products to minimize carry-over effects such as adaptation. The attribute generation process lasts about 30–45 min and involves assessment of all the samples. The time spent will vary depending on the number and kind of products to be evaluated.

In a subsequent session, a pooled list of all generated attributes is provided to the subjects. They are allowed to modify their own list accordingly to make sure that all relevant dimensions are included by checking the respective samples. A break will be necessary for the panel leader to update the individual vocabularies.

In the final sessions the subjects rank the samples according to their own vocabulary one attribute at the time in order of the perceived level of intensity. The original flash profile method suggests three replicate sessions (Dairou & Sieffermann 2002). This repetition step has also been carried out in FCP studies to enable evaluation of panel performance (Costell et al. 1995; Deliza et al. 2005). But the number of replicates may vary depending on the type of stimuli/products. The number of samples to be evaluated with the flash profile method also varies and will determine the number of replicates to be used. The method allows for ties, that is, allocating the same rank number to samples that have the same perceived intensity.

14.3.4 Alternative Approaches

It has been noted that the flash profile can result in complex vocabularies. In order to shorten and better define the lists of attributes, an alternative to the flash profile method was proposed by Lorho (2010). He combined elements from FCP, RGM and flash profile in one method called individual vocabulary profiling (IVP). The main improvement of the IVP method, compared to the flash profile,

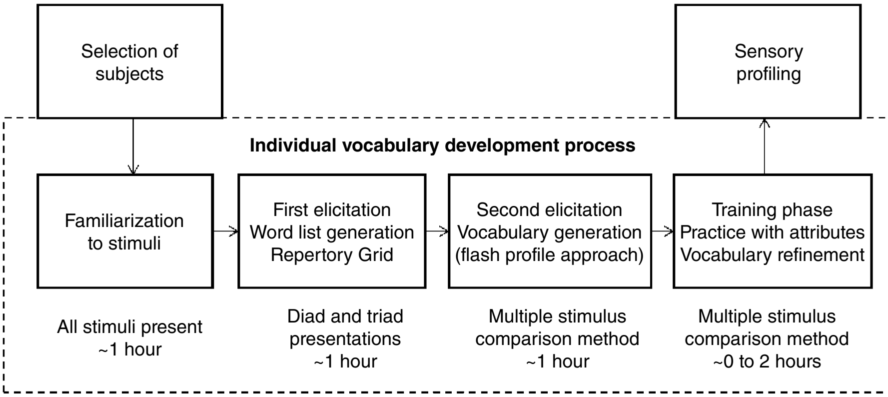


Figure 14.2 Schematic overview of the steps in an extended flash profile method for acoustic panels proposed by Lorho (2010).

was the investment of slightly more time for developing and defining the individual vocabularies. This circumvents the main shortcoming of the flash profile method of the use of attributes with limited interpretative value. Nevertheless, the IVP method still benefits from faster speed, less cost and not being limited to trained panels, compared to traditional descriptive analysis. The IVP approach (Lorho 2010) is schematically presented in Figure 14.2. The steps involved are familiarization with the stimuli, two sessions on individual attribute elicitation and a training phase for vocabulary refinement. The method has been successfully introduced for sound quality evaluations.

Another alternative to the flash profile is ranking descriptive analysis (RDA), proposed by Richter et al. (2010). This method utilized a consensus vocabulary and ranking of samples with a relatively large trained panel. The RDA method involves four steps:

1 selection of the most frequently observed attributes
2 definition of the selected attributes by the panel
3 development of a consensus evaluation procedure with the panel
4 ranking of the samples for all the selected attributes in one replicate session.

In this way, RDA can be considered as a compromise between the flash profile and QDA. Despite the larger number of assessors required, RDA has the advantage of lower costs due to the requirement for fewer sessions than QDA. Similar to the flash profile, a limited number of samples can be included in RDA, because of the ranking exercise. The idea of applying a ranking test instead of scaling in the description of a product has been suggested by Rodrigue et al. (2000).

14.3.5 Practical Considerations

When setting up the flash profile, the experimenter should reflect on a number of considerations mostly related to the time spent on the practical part of the analysis. When a large set of samples is chosen for the analyses, the individual vocabulary development becomes a more time-consuming task. A relevant shortcut can be to present a subset of samples to the subjects. Omitting individual samples is possible, for example where the sample lies within a construct spanned by other samples. This is often the case in experimental studies, while it is more uncertain in sets of market-related samples.

After the first individual vocabulary development session, the individual lists of vocabularies need to be distributed in some way to the whole panel. As it might take some time to copy all lists, a break before the next vocabulary development session is a good idea. After the final vocabulary development, the panel leader must take into account the way in which he/she records the individual assessments on the individual attributes ranks. This practical procedure is different from an ordinary consensus profiling and is dependent on the choice of sensory software. The whole flash profile can, with some effort, be completed within one day, of course depending on the nature of products and the aim of the study. But if one wishes to avoid stressing subjects and also wishes to perform repetitions, the test is recommended to run over several days.

The descriptive result of the flash profile might include a great number of attributes, blurring the results plot considerably relative to one from a conventional descriptive profile. One must allocate sufficient time for analysis and interpretation of flash profile data. The method sets higher demands for the proficiency of the standard analyst. If lemmatization (grouping of similar attributes) is carried out, one must be consistent and report the process. For example, one could group the attributes of the panel according to broad and more refined categories, as one does when building a sensory wheel. This will allow grouping of attributes with a similar meaning. For instance, the term 'fruity' can be linked to subcategories of lemon, banana, pear, etc. Since the flash profile does not generate a panel consensus, the panel leader needs enough knowledge on the products to be able to generate a *post hoc* grouping of sensory attributes by looking at the consensus configuration plots (see section 14.4). For instance, categories like 'fruity', 'banana', 'ester-like' and 'sweet' may be grouped to a 'sweet-fruity' dimension for a particular product, but the panel leader will require sufficient experience and product knowledge to group terms together in a meaningful way.

14.3.6 Statistical Analysis

14.3.6.1 Data Collection

As the evaluation is performed as rank ordering, transferring the ranking positions on the paper sheet into scores is the first step in data analysis. An example of a ranking sheet is presented in Figure 14.3. The product with the lowest intensity is given a score of 1, and then the score is increased according to the ranking order of each product. In this way, the rank numbers are ascending according to the increasing of the sensory intensity (rank position) for the attribute. The rank data are then easily given interpretation in multivariate analysis, since higher rank values correspond to the higher perceived sensory intensities.

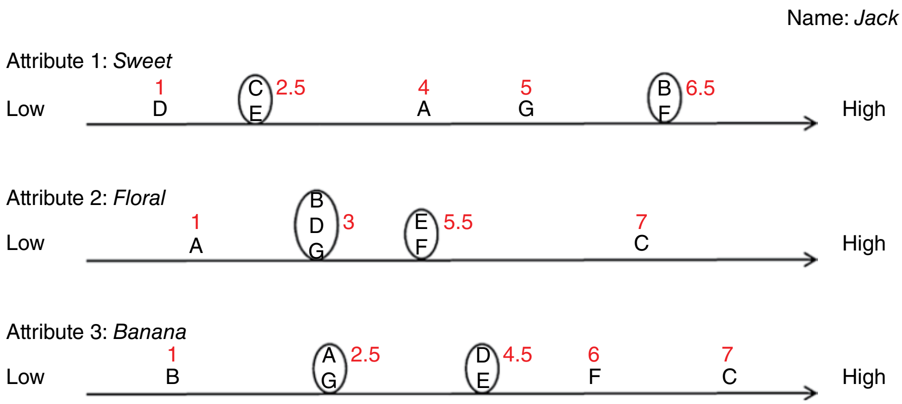


Figure 14.3 Example of a flash profile ranking sheet and data coding (only three attributes are represented).

	Assessor 1			
Product	Sweet	Floral	Banana	…
A	5	1	2.5	
B	6.5	3	1	
C	2.5	7	7	
D	1	3	4.5	
E	2.5	5.5	4.5	
F	6.5	5.5	6	
G	5	3	2.5	
Σ	28	28	28	

Figure 14.4 Example of a data table with rank numbers from a flash profile corresponding to the ranking sheet in Figure 14.3. The rank numbers are presented here for only one replicate.

For products with equal rank positions (tied samples), a mean rank value is given. An example is given in Figure 14.4 showing a data table with rank numbers corresponding to the ranking sheet in Figure 14.3. In this way individual matrices for each assessor (products × attributes) are built. In case rank numbers are assigned manually into data tables, the sums of the rank values for each attribute should be identical (see Figure 14.3).

14.3.6.2 Univariate Analysis

Since the set of attributes is often unique for each assessor in flash profile data, univariate analysis can be used to look at attribute consistency for sensory assessors. The flash profile ranking data could be analysed by non-parametric tests such as the Friedman test. For instance, the Friedman two-way analysis of variance (ANOVA) by ranks could be used to check for consistency in the replications. Those attributes showing a poor statistical reliability could be omitted from

Product	Sweet	
	Replicate 1	Replicate 2
A	4	5
B	6.5	6.5
C	2.5	1.5
D	1	1.5
E	2.5	3
F	6.5	6.5
G	5	4

Figure 14.5 Example of a Spearman's test result when evaluating the panellists' reproducibility between two replicates. The Spearman rank correlation coefficient is: $r = 1 - 6 \times \text{Sum}(V^2) / (N \times (N^2 - 1))$. V is the differences between the ranks of both variables: V = Rank (R1) – Rank (R2); N is the number of products. In this example, $r = 0.9375$.

further multivariate analysis. Alternatively, one-way ANOVA may be used (Lawless & Heymann 2010). One-way ANOVA (product as the main factor) is usually run on the flash profile data for evaluating the panellists' performance. Attributes that are found not to discriminate the products significantly might be excluded from the list of assessors. Spearman's correlation coefficients can also be calculated and grouped for each assessor in order to evaluate the panellist's repeatability (Gkatzionis et al. 2013; Price et al. 2014; Rason et al. 2006). An example of Spearman's correlation test result when evaluating reproducibility between two replicates is shown in Figure 14.5. One could easily run this test in the XLSTAT software. However, the use of univariate analysis such as Spearman's correlation test for the flash profile data also depends on the project aim and nature of the samples, etc. If one prefers to take all assessors' measures into account, all attributes without any exclusion could be used for the following multivariate analysis.

14.3.6.3 Multivariate Analysis

A data analysis method that is particularly well suited for flash profile is generalized Procustes analysis (GPA) (Gower 1975), similarly to the analysis of FCP data. The GPA uses translation, rotation and isotropic scaling of the individual product sets in order to maximize the consensus configuration. It can handle different sizes of response variables on the sample objects for each subject. When an optimal consensus configuration is found, the panel average position of the products can be related to the sensory attributers used by all subjects. Also, for attributes that individuals have in common, the agreement between individuals in the product space can be visualized. This approach is very similar to the analysis of FCP data but is based on ordinal (rank order) data instead of scale data.

Before processing of the data by GPA, the flash profile results are normally averaged over the replicates. The structure of the data matrix will be the same as presented in Figure 14.4, but with the values replaced by the averaged rank numbers over the replicates. The flash profile data thus consist of blocks of variables with different sizes, one block for each assessor. However, the GPA can also be run on the separate data sets of the replicates. The results from the GPA are divided over different dimensions with decreasing importance. As a validation tool, GPA uses a permutation test as diagnostic, by comparing the experimental configuration with a random configuration. Only GPA results with configurations deviating in the permutation test should be given interpretation. The first two dimensions of the GPA are often sufficient to show the consensus configuration for the samples and the correlation space for the attributes.

An alternative way of analysing flash profile data is by multiple factor analysis (MFA) (Dehlholm et al. 2012a). MFA (Escofier & Pagès 1994) is in this case based on PCA as the data are quantitative and allow grouping of the variables and samples. MFA on flash profile data will give a similar result as GPA. However, GPA seems to be better suited for finding the optimal consensus space in individual scaling data. As a consequence, it may give somewhat higher explained variance than MFA. Both GPA and MFA analyses can be done, for instance, in Senstools software (OP&P), XLSTAT (Addinsoft) and R (Ihaka & Gentleman 1996; R Development Core Team 2010) using the functionality of the add-on packages SensoMineR and FactoMineR. The code for running GPA using FactoMineR has been reported by Varela and Ares (2014). The code for running MFA with confidence ellipses, for example showing the 95% probability ellipses around the sample scores, has been reported by Dehlholm et al. (2012b). XLSTAT provides tutorials on how to run GPA and MFA. The program runs as an add-in for Microsoft Excel.

It must be noted that the flash profile rank data are ordinal in nature, meaning that the distances between rank numbers do not necessarily correspond to the differences in perceived intensities between the samples. A sample may be much higher in a sensory intensity than all the other samples, but receive only a rank number that is one integer higher. For the interpretation of the results from the data analysis, one should bear this in mind.

14.3.7 Advantages and Disadvantages

The main benefit of the flash profile is its speed while still achieving a reasonable accuracy. The whole profiling can be completed in a few sessions, which would take much longer in the more elaborate conventional profiling methods, like QDA, Spectrum Method, etc. The simultaneous presentation of products is very efficient because subjects do not need any familiarization phase with samples. From the start, they can generate attributes that discriminate and are relevant for the product. No training phase is needed because no agreement is required between subjects about terms, definitions and evaluation procedures and because the use of ordinal scales is intuitive.

Another advantage of the flash profile is the ease of experimental set-up. It needs significantly fewer sessions than a conventional descriptive profiling.

There are no problems with matching panellist schedules as each session can be arranged individually. All the samples are prepared before the sessions and the subjects have all the samples at the same time. In practice, this means that the experimenter does not need to be present.

A third advantage of the flash profile method in comparison to the conventional profiling methods is the representation of diverse viewpoints by the subjects. They choose their vocabulary according to their own sensitivity and perception. The combination of each subject description reflects different points of view depending on the importance given by the subject to each sensory modality. This combination may enrich the sensory description of the product.

However, in some cases it may be difficult for the panel to evaluate all the samples simultaneously due to fatigue. This is a limitation of the method but in practice, 10 or more samples can be tested, depending on the kind of stimuli/products. For instance, Tarea et al. (2007) reported a flash profile with six trained assessors who successfully examined the sensory texture of 49 samples of pear and apple purées. However, it is worth noting that the subjects declared that the task was really tedious. In this particular case, the assessors used 4–7 attributes and the rankings were found easy because of diversity of the product set.

Other limitations of the flash profile may arise in cases where precise temperature control over the samples is needed or when evaluating the stability of products over time, due to the need for simultaneous sample presentation. Compared to conventional profiling where terminology is not a major concern, the flash profile could also lead to difficulty in interpretation of sensory characteristics, due to a large number of terms and lack of definitions and evaluation procedures. Furthermore, the flash profile terminology may not be easily reused by another panel, which is easier in conventional profiling where the panel is trained with reference standards for sensory attributes.

In cases where sensory attributes of products need to be correlated to instrumental measurements, it is unclear whether to recommend the use of the flash profile. Lassoued et al. (2008) successfully correlated the instrumental texture of bread with the flash profile using MFA to combine the data sets. Also, a recent study showed that flash profile analysis of Stilton blue cheese with different levels of yeast inoculum could be used to correlate to volatile analysis using partial least squares regression (Price et al. 2014). However, such approaches deserve careful consideration. At present, there is not enough evidence to suggest that the flash profile can be used as an alternative for conventional profiling in such correlation studies. In particular, the lack of sensory definitions and focus on the main sensory differences may lead to imprecise and incorrect interpretation. As shown by Dehlholm et al. (2012a) in their study on liver paté, the flash profile provided the best discrimination among fast sensory profiling methods (sorting, Global Napping® and Partial Napping®), but still missed important information on sample differences, which were perceived using conventional profiling.

The training of subjects on a sensory vocabulary is a time-consuming and costly effort, and may not be needed in all situations. Particularly in circumstances where one can afford to have a lower degree of semantic interpretability at a lower cost and higher speed, alternative sensory profiling methods have become a first choice. The flash profile method ranks products in relation to each other for one attribute at a time and is more closely related to conventional profiling methods than perceptual mapping methods. These latter methods, such as Napping and ultra-flash profiling (Pagès 2003, 2005), need more cognitive processing of multiple sensory impressions at the same time, to make a judgement of where to place a product in relation to other products. Such methods are even faster but also more difficult to interpret since the response strategies may be diverse and are much less controlled. Dehlholm et al. (2012a) clearly showed the advantage of the flash profile in comparison to Napping and ultra-flash profile, but argued that structured approaches to Napping, such as Partial Napping, best combined speed and accuracy in rapid sensory profiling. Even though one could claim that such methods are based on more holistic judgements, it is too early to conclude that perceptual mapping techniques provide better profiling data than the flash profile method.

14.3.8 Applications

The flash profile has been applied to many different food products, including jams (Dairou & Sieffermann 2002), dairy products (Delarue & Sieffermann 2004), commercial fruit purées (Tarea et al. 2007), jellies (Blancher et al. 2007), bread (Lassoued et al. 2008), hot beverages (Moussaoui & Varela 2010), lemon iced teas (Veinand et al. 2011), fish nuggets (Albert et al. 2011), liver paté (Dehlholm et al. 2012a), meat (Ramírez-Rivera et al. 2012), carrot snacks (Dueik et al. 2013), blue cheese (Price et al. 2014), chocolates (Rezende et al. 2015) and wine (Liu et al. 2016). Also, it has been successfully applied for non-food products such as hall acoustics (Lokki et al. 2011).

Many applications of the flash profile have been reported and it has become clearer in which situations the method is particularly well suited. In cases where a precise sensory definition of the products is not needed, but flexibility and rapidity of test results are demanded, the flash profile has the advantage above conventional profiling, for instance, when working with products with a short shelf-life, or when fast product insights are in demand in the early stages of product development. But when quicker results are needed, other methods like projective mapping may be preferred.

14.4 Case Studies

14.4.1 Case Study 1: One-Shot Analysis

The suitability and robustness of the flash profile were tested in a sensory evaluation of smoked fresh cheese.

14.4.1.1 Sample

The cheese samples were prepared from kvark (1% fat; 12% protein; Arla Foods, Denmark) which was manually smoked in small portions in a table smoker at 30 °C. Different wood types (beech flakes, common alder and hay), additive (fresh orange peel or not) and smoking times were chosen to span a space for measuring sensory differences (Table 14.2).

Furthermore, for each replicate session of the flash profile, two production batches from the same smoking treatment (3 min) were included (common alder with orange peel (CP3-1 and CP3-2)). This was done to assess the variation due to batch production, since the reliability of the benchtop smoking process was unknown.

14.4.1.2 Method

A trained sensory panel consisting of six subjects (three men and three women) participated in the test. The sensory panel followed a standard flash profile protocol with two replicated sample evaluations.

The flash profile was conducted in three sessions where all of the nine samples were presented. In the first session, the assessors were given an explanation about the procedure. Then, all assessors were asked to list the sensory characteristics that best described the differences among the samples and were instructed to avoid the use of hedonic terms. The process of attributes generation lasted around 30 min. After a short break the assessors were given a pooled attribute list derived from all panel members on the whiteboard. They were allowed to modify their list by adding or deleting the attributes. In the final two sessions, assessors were required to rank the cheese samples according to their own updated attribute list in the sensory booths. They needed to place the code of samples on a line scale anchored at the left side representing low and right side for high. Ties were allowed.

Table 14.2 Selected samples for the flash profile of smoked fresh cheese.

Sample code	Wood type	Smoke additive	Smoking time (minutes)	Batch number
B1	Beech flakes	None	1	1
B3	Beech flakes	None	3	1
BP3	Beech flakes	Orange peel	3	1
C1	Common alder	None	1	1
C3	Common alder	None	3	1
CP3-1	Common alder	Orange peel	3	1
CP3-2	Common alder	Orange peel	3	2
H5	Hay	None	5	1
HP3	Hay	Orange peel	3	1

14.4.1.3 Data Analysis

The score sheets from the assessors were collected and the rank numbers for all nine samples and six assessors were entered in Excel. Rank numbers in the data set were not averaged for the two replicates. Thus each replicate was treated as a unique sample. In this way, the variation due to the replicates in comparison to the variation due to the smoking treatment could be assessed.

The data were analysed by GPA in Senstools using the Procrustes free choice module, since the flash profile data structure is similar to free choice profiling data. The permutation test showed a total variance accounted value of 75%, which was much higher than the 5% upper value of the total variance accounted for of 71%. Therefore, the GPA results contained significant systematic information. Subsequently, the scree plot, showing the consensus and residual variance, was interpreted and two dimensions were selected, since the third dimension only added little extra information.

14.4.1.4 Result and Discussion

In the GPA result, 33% of the variance was explained in the first dimension and the second dimension explained 13%. Figure 14.6 shows the GPA consensus space for the samples. Dim 1 clearly showed that the addition of orange peel gave marked differences in flavour of the beech smoked fresh cheeses (B3 versus

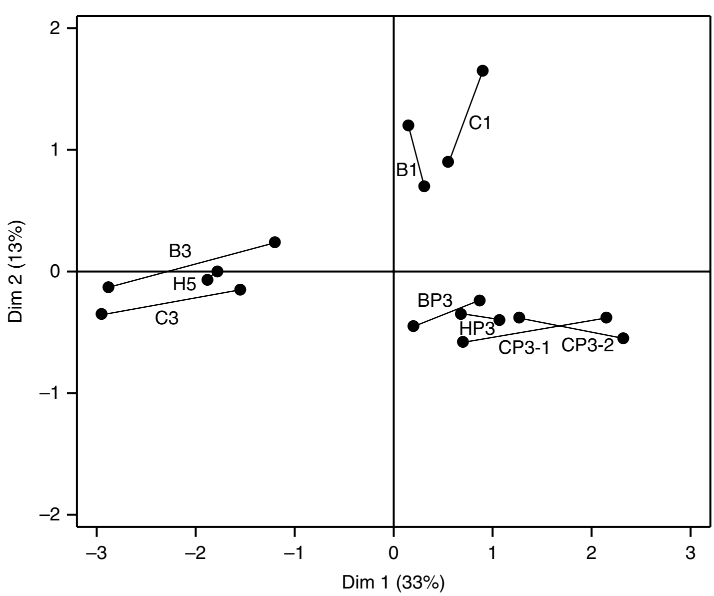


Figure 14.6 GPA consensus product map of the flash profile analysis of smoked fresh cheese. Sample codes: B, beech wood smoke; C, common alder smoke; H, hay smoke; P, orange peel smoke additive; 1, 1 min smoking; 3, 3 min smoking; 5, 5 min smoking; codes 3-1 and 3-2 signify batch replicates for smoking at 3 min.

BP3). Smoking with hay for 5 min gave a similar effect as smoking for 3 min for beech and common alder, which was expected due to the lower smoke condensate onto the cheese. Furthermore, the smoking time of 1 min versus 5 min for beech and common alder smoking was clearly shown on Dim 2. The robustness of the flash profile could be assessed from the variation between the sensory replicates, which appeared to be limiting the discrimination between the H5, B3 and C3 samples as well as the BP3 and HP3 samples. However, the replicate variation was not similar for all samples, indicating that the panel could more easily describe the less complex samples H1 and B1 due to the weak smoke flavour and BP3 and HP3 due to the strong orange peel flavour. However, the batch variation (CP3-1 versus CP3-2) was smaller than the sensory replicate variation. When looking into the attribute spaces, it was clearly observed that the individual vocabularies were extensive. They varied from 19 to 36 sensory attributes. Some interpretation could be given to the individual sensory spaces, for example for the orange peel samples (BP3, CP3 and HP3) 'citrus', 'fruity', 'mandarin', 'apricot', etc. terms were used whereas the most severely smoked cheeses (H5, B3 and C3) were dominated by 'smoked', 'burnt', 'tarmac' attributes, which were absent for the least smoked samples.

Overall, the flash profile analysis could show main differences between the samples, but the sensory replicate variation was relatively large for some of the samples. In this particular case, the sensory panel was expected to perform more accurately if time for training on developing and using the individual sensory vocabularies had been allowed. The case showed that one has to carefully consider the use of fast descriptive analysis in relation to the merits of speed, costs and need for accuracy.

14.4.2 Case Study 2: Small Sensory Differences

The reliability of the flash profile was investigated in a sensory evaluation of a model wine system. In this study, flavours were added to a neutral wine, thus controlling the expected sensory differences between the wine samples. The samples were designed in such a way that the differences were noticeable but relatively small. Therefore, the applicability of the flash profile for samples with small sensory differences was investigated. The case was part of a larger study to investigate the reliability of the flash profile and other rapid methods in comparison to conventional profile analysis (Liu et al. 2016).

14.4.2.1 Sample

A model wine system was designed where three different volatile compounds, benzaldehyde, isopentyl acetate and 2-phenylethanol, were added into a plain white wine (Pinot Blanc, 11.5% v/v) at two different concentrations. In order to investigate the performance of the panels, two blind replicates of benzaldehyde high and 2-phenylethanol low were included. The list and details including odour descriptions are shown in Table 14.3.

Table 14.3 Samples used in the flash profile of a model wine system.

Flavour compounds added	Level of addition (μL/L)	Odour descriptors†
Benzaldehyde High*	8	Almond
Benzaldehyde Low	2	Almond
Isopentyl acetate High	5	Banana
Isopentyl acetate Low	1	Banana
2-Phenylethanol High	100	Rose/floral/spicy
2-Phenylethanol Low*	10	Rose/floral/spicy
Base wine	–	A plain white wine

*The two wines were used as blind replicates.
†Odour descriptors reported in Flavornet (www.flavornet.org/f_kovats.html).
Source: Liu et al. (2016). Reproduced with permission of Elsevier.

14.4.2.2 Method

The flash profile was carried out by a trained sensory panel composed of eight assessors. The flash profile consisted of three sessions in which the nine samples were presented simultaneously using a similar flash profile procedure as described in section 14.4.1.2.

14.4.2.3 Data Analysis

A GPA analysis was carried out on the average data from the two replicates of the flash profile using XLSTAT (version 2014.5.03, Addinsoft). For the analysis, the Gower method was selected and the default permutation test with 300 permutations. The program provides two options for running the GPA, namely Gower and Commandeur. Commandeur is used when missing values occur in the data set.

14.4.2.4 Result and Discussion

A total of 34 flavour attributes were chosen by the panel and each assessor used a vocabulary varying from six to 13 flavour descriptors. Clearly, in this relatively simple product, fewer attributes were used than in case study 1 on smoked cheeses. However, given the three flavour modifications of the wine, the sensory vocabulary was still relatively extensive.

The GPA consensus product map of the flash profile is displayed in Figure 14.7. The first two dimensions accounted for 57% of the explained variance (39% and 17%, dimensions 1 and 2 respectively). The wines with the same added volatile compound were close to each other. The wines with added volatiles in high concentration were located furthest from the centre of the plot and stretched the plot into three different directions. The wines with low concentrations of added volatiles were located closest to the centre and near to the basic wine. Benzaldehyde and isopentyl acetate groups were in the

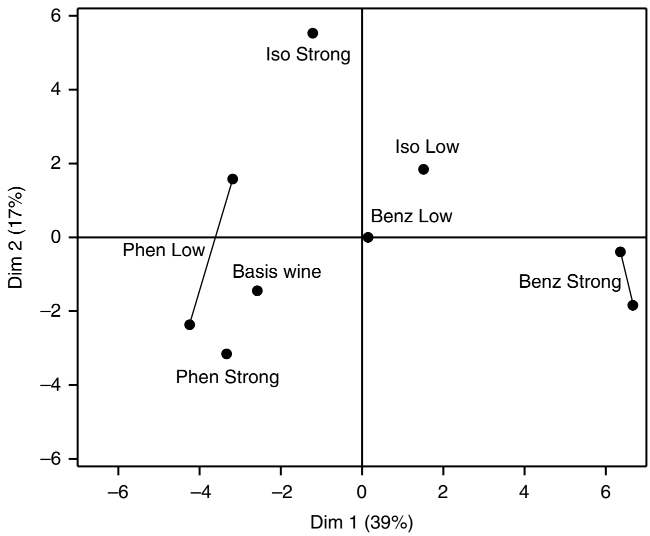


Figure 14.7 GPA consensus product map of the flash profile analysis of wine samples. Benzaldehyde, isopentyl acetate and 2-phenylethanol are abbreviated to Benz, Iso and 2-Phen, respectively. Benz High and 2-Phen Low were each presented twice in each flash profile replicate. The line connecting these sample pairs gives an indication of the reliability of the evaluation of the weak (Phen low) versus the stronger (Benz) attribute intensity.

positive side of dimension 1 and dimension 2, respectively, and the 2-phenylethanol group was in the negative part of dimension 1 and 2. However, all three samples with 2-phenylethanol were close to the basic wine, indicating that the panel had difficulties distinguishing the wines with this compound. This was confirmed in the conventional profile analysis of the wines, where the difference between the wines with added 2-phenylethanol was just significant. The flavour differences of the wines with benzaldehyde and isopentyl acetate additions were more marked (Liu et al. 2016).

The positioning of the two pairs of blind replicates is considered as a way to check the panel performance and examine the accuracy of the methods. The blind duplicates of benzaldehyde high were very close to each other, indicating good reproducibility among the panel. As expected, the blind replicate for the 2-phenylethanol low was judged less reliable. This was probably due to greater confusion in the ranking of the samples with more similar sensory properties (basic wine, 2-phenylethanol low). If the wines were experimental wines where one did not know the sensory difference beforehand, one may have concluded that the wines were different. Hence, when the sensory differences among products are clearly distinguishable, one can use the flash profile method but when the differences are relatively small, confusion in ranking may occur which could lead to spurious conclusions. In this case the flash profile method might not be a good choice.

14.5 Summary

The flash profile is quick to generate a sensory vocabulary, and subsequently each subject uses his/her own vocabulary to rank all samples on their relative differences for the sensory attributes. There is no need to conduct extensive training on the product properties. However, the comparative assessment limits the method to the number of samples that can be tested. Furthermore, since the flash profile relies on subsequent alignment of the products × attributes space, this method leads to a lower degree of precision, especially when the attribute space is complex, for example when determining the flavour of foods. In cases where the attribute space is limited, as for texture or sound, the flash profile has the advantage of generating fast descriptive results on the main sensory differences between products. The ambiguous use of attributes in the flash profile may have an advantage in cases where the generation of multiple attributes is desired, such as for vocabularies for communication with consumers. In comparison to other rapid sensory mapping methods like Napping, results from the flash profile contain quantitative information on the intensity of the attributes. Thus the flash profile method is a good compromise in situations where a rapid respond is needed but enough time is available to run two panel sessions.

14.6 Future Developments

One of the difficulties of the flash profile is the dependence of the results on the capability of individual subjects to name their perceptions. Some assessors may also drop some terms because they may undervalue their own sensory abilities. The strict separation of panel members in order to adopt their individual vocabularies may be made more flexible, allowing more interactions between the panellists. This would lead to more comprehensive sensory vocabularies and still maintain the rapidness of the method. Instead of presenting individually the list of every term at the second session in the flash profile, it could be considered to hold a group session or apply structured interviewing techniques (Lorho 2010). In a joint session, every subject could present their own list of terms and explain the reasons for their choices. This would give significant improvements on the sensory concepts and may also lead to more comprehensive lists of attributes without compromising the individual vocabularies.

References

Albert, A., Varela, P., Salvador, A., Hough, G. & Fiszman, S. (2011) Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality & Preference*, **22**(5), 463–473.

Ballay, S., Warrenburg, S., Sieffermann, J.M., Glazman, L. & Gazano, G. (2006) A New Fragrance Language: Intercultural Knowledge and Emotions. Paper presented at the 24th IFSCC Congress – Integration of Cosmetic Sciences, Osaka, Japan.

Berg, J. & Rumsey, F. (2006) Identification of quality attributes of spatial audio by repertory grid technique. *Journal of the Audio Engineering Society*, **54**(5), 365–379.

Blancher, G., Chollet, S., Kesteloot, R., Nguyen Hoang, D., Cuvelier, G. & Sieffermann, J.M. (2007) French and Vietnamese: how do they describe texture characteristics of the same food? A case study with jellies. *Food Quality & Preference*, **18**, 560–575.

Costell, E., Trujillo, C., Damasio, M.H. & Duran, L. (1995) Texture of sweet orange gels by free-choice profiling. *Journal of Sensory Studies*, **10**, 163–179.

Dairou, V. & Sieffermann, J.M. (2002) A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, **67**(2), 826–834.

Dehlholm, C., Brockhoff, P.B., Meinert, L., Aaslyng, M.D. & Bredie, W.L.P. (2012a) Rapid descriptive sensory methods – comparison and validation of free multiple sorting, partial napping, napping, flash profiling and conventional profiling. *Food Quality & Preference*, **26**, 267–277.

Dehlholm, C., Brockhoff, P.B. & Bredie, W.L.P. (2012b) Confidence ellipses: a variation based on parametric bootstrapping applicable on multiple factor analysis results for rapid graphical evaluation. *Food Quality & Preference*, **26**, 278–280.

Delarue, J. (2014) Flash profile, its evolution and uses in sensory and consumer science. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques*. Sawston: Woodhead Publishing, pp. 121–151.

Delarue, J. & Sieffermann, J.M. (2004) Sensory mapping using flash profile. Comparison with a conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality & Preference*, **15**, 383–392.

Deliza, R., MacFie, H. & Hedderley, D. (2005) The consumer sensory perception of passion-fruit juice using free-choice profiling. *Journal of Sensory Studies*, **20**(1), 17–27.

Drake, M.A. & Civille, G.V. (2002) Flavor lexicons. *Comprehensive Reviews in Food Science and Food Safety*, **2**, 33–40.

Dueik, V., Marzullo, C. & Bouchon, P. (2013) Effect of vacuum inclusion on the quality and the sensory attributes of carrot snacks. *LWT – Food Science and Technology*, **50**(1), 361–365.

Escofier, B. & Pagès, J. (1994) Multiple factor-analysis (Afmult package). *Computational Statistics and Data Analysis*, **18**, 121–140.

Feixas, G., Montebruno, C., Dada, G., del Castillo, M. & Compan, V. (2010) Self construction, cognitive conflicts and polarization in bulimia nervosa. *International Journal of Clinical and Health Psychology*, **10**(3), 445–457.

Frost, W.A.K. & Braine, R.L. (1967) The application of the Repertory Grid technique to problems in market research. *Commentary*, **9**, 161–175.

Gaines, B.R. & Shaw, M.L.G. (1993) Knowledge acquisition tools based on personal construct psychology. *Knowledge Engineering Review*, **8**, 49–85.

Gkatzionis, K., Hewson, L., Hollowood, T., Hort, J., Dodd, C.E.R. & Linforth, R.S.T. (2013) Effect of Yarrowia lipolytica on blue cheese odour development: flash profile sensory evaluation of microbiological models and cheeses. *International Dairy Journal*, **30**, 8–13.

Gower, J.C. (1975) Generalized Procustes analysis. *Psychometrika*, **40**(1), 33–51.

Ihaka, R. & Gentleman, R. (1996) R: a language for data analysis and graphic. *Journal of Computational and Graphical Statistics*, **5**, 299–314.

Kelly, G. (1955) *The Psychology of Personal Constructs*. New York: Norton.

Lassoued, N., Delarue, J., Launay, B. & Michon, C. (2008) Baked product texture: correlations between instrumental and sensory characterization using Flash Profile. *Journal of Cereal Science*, **48**(1), 133–143.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evaluation of Food: Principles and Practices*, 2nd edn. New York: Springer.

Liu, J., Grønbeck, M.S., di Monaco, R., Giacalone, D. & Bredie, W.L.P. (2016) Performance of Flash Profile and Napping with and without training for describing small sensory differences in a model wine. *Food Quality & Preference*, **48**, 41–49.

Lokki, T., Patynen, J., Kuusinen, A., Vertanen, H. & Tervo, S. (2011) Concert hall acoustics assessment with individually elicited attributes. *Journal of the Acoustical Society of America*, **130**(2), 835–849.

Lorho, G. (2010) Perceived quality evaluation and application to sound reproduction over headphones. PhD thesis, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland.

Meilgaard, M., Civille, G.V. & Carr, B.T. (2006) *Sensory Evaluation Techniques*, 4th edn. Boca Raton: CRC Press.

Moussaoui, K.A. & Varela, P. (2010) Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality & Preference*, **21**(8), 1088–1099.

Olson, J.C. (1981) The importance of cognition processes and existing knowledge structures for understanding food acceptance. In: *Criteria of Food Acceptance*. Zurich: Forester Publishing.

Pagès, J. (2003) Direct collection of sensory distances: application to the evaluation of ten white wines of the Loire Valley. *Sciences des Aliments*, **23**, 679–688.

Pagès, J. (2005) Collection and analysis of perceived product inter-distances using multiple factor analysis: application to the study of 10 white wines from the Loire Valley. *Food Quality & Preference*, **16**, 642–649.

Price, E.J., Linforth, R.S., Dodd, C.E. et al. (2014) Study of the influence of yeast inoculum concentration (Yarrowia lipolytica and Kluyveromyces lactis) on blue cheese aroma development using microbiological models. *Food Chemistry*, **145**, 464–472.

R Development Core Team (2010) *R: A Language and Environment for Statistical Computing*. 3-900051-07-0. Vienna: R Foundation for Statistical Computing.

Ramírez-Rivera, E.J., Camacho-Escobar, M.A., Garcia-Lopez, J.C. et al. (2012) Sensory analysis of Creole turkey meat with flash profile method. *Open Journal of Animal Sciences*, **02**(01), 1–10.

Rason, J., Léger, L., Dufour, E. & Lebecque, A. (2006) Relations between the know-how of small-scale facilities and the sensory diversity of traditional dry sausages from the Massif Central in France. *European Food Research and Technology*, **222**(5–6), 580–589.

Rezende, N.V., Benassi, M.T., Vissotto, F.Z., Augusto, P.P.C. & Grossmann, M.V.E. (2015) Mixture design applied for the partial replacement of fat with fibre in sucrose-free chocolates. *LWT – Food Science and Technology*, **62**, 598–604.

Richter, V.B., de Almeida, T.C.A., Prudencio, S.H. & de Toledo Benassi, M. (2010) Proposing a ranking descriptive sensory method. *Food Quality & Preference*, **21**(6), 611–620.

Rodrigue, N., Guillet, M., Fortin, J. & Martin, J.F. (2000) Comparing information obtained from ranking and descriptive tests of four sweet corn products. *Food Quality & Preference*, **11**, 47–54.

Sieffermann, J.M. (2000) Le profil flash: un outil rapide et innovant d'analyse sensorielle descriptive. In: Proceedings of AGORAL 2000, XII rencontres 'L'innovation: de l'idée au success', Montpellier, France, pp. 335–340.

Stone, H., Sidel, J.L., Oliver, S., Woolsey, A. & Singleton, R.C. (1974) Sensory evaluation by quantitative descriptive analysis. *Food Technology*, **28**, 24–34.

Swahn, J., Ostrom, A., Larsson, U. & Gustafsson, I.B. (2010) Sensory and semantic language model for red apples. *Journal of Sensory Studies*, **25**(4), 591–615.

Tarea, S., Cuvelier, G. & Sieffermann, J.M. (2007) Sensory evaluation of the texture of 49 commercial apple and pear purees. *Journal of Food Quality*, **30**, 1121–1131.

Varela, P. & Ares, G. (2012) Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International*, **48**, 893–908.

Varela, P. & Ares, G. (2014) *Novel Techniques in Sensory Characterisation and Consumer Profiling*. Boca Raton: CRC Press.

Veinand, B., Godefroy, C., Adam, C. & Delarue, J. (2011) Highlight of important product characteristics for consumers. Comparison of three sensory descriptive methods performed by consumers. *Food Quality & Preference*, **22**(5), 474–485.

Williams, A.A. & Arnold, G.M. (1984) A new approach to sensory analysis of foods and beverages. In: Adda, J. (ed.) *Progress in Flavour Research*. Proceedings of the 4th Weurman Flavour Research Symposium, Elsevier, Amsterdam, pp. 35–50.

Williams, A.A. & Langron, S.P. (1984) The use of free choice profiling for the examination of commercial ports. *Journal of the Science of Food and Agriculture*, **35**, 558–568.

Yan, W., Chen, C.H. & Chang, W. (2011) A functional-commercial analysis strategy for product conceptualization. *Expert Systems with Applications*, **38**(8), 9879–9887.

CHAPTER 15

Projective Mapping & Sorting Tasks

Dominique Valentin, Sylvie Chollet, Michael Nestrud and Hervé Abdi

15.1 Introduction

Projective mapping and the sorting task are rather old techniques in psychology, dating at least from 1935 for the sorting task (Hulin & Katz 1935) and 1964 and 1983 for projective mapping, but under different names (Coombs 1964; Dunn-Rankin 1983). In contrast, their use is relatively recent in sensory evaluation (1989 and 1990 for the sorting task with Lawless and with MacRae, Howgate & Geelhoed, and 1994 for projective mapping with Risvik et al.) (Chollet et al. 2011, 2014; Valentin et al. 2012). The main purpose of these methods is to obtain similarity measurements between products from participants who, in general, are not required to be trained assessors. These methods are loosely based upon the psychological construct of categorization (Abdi 1986, 1987; Cohen & Lefebvre 2005; Murphy 2002; Rosch 1973, 1978; Rosch & Mervis 1981; Tversky 1977; Valentin & Chanquoy 2012; Wittgenstein 1953), a concept that formalizes the cognitive function of spontaneously organizing the world into meaningful and important natural categories (e.g. animate versus inanimate) (Sha et al. 2015) or categories statistically derived from the correlational structure of objects' properties and features (Tversky 1977). As far as sensory evaluation is concerned, however, the very large literature on categorization and concepts boils down to indicating that categorization requires little effort (i.e. it is mostly automatic) and that it occurs very early in life (probably even at birth if not before) (Valentin & Abdi 2001). Practically, this indicates that tasks based on simple categorization can be performed by untrained participants (including children) (Gombert et al. 1990) as well as by trained assessors and experts (Chollet et al. 2005).

Similarity measurements differ from other descriptive methods used in sensory evaluation because they seek to elicit and analyse global similarities

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp,
Joanne Hort and Tracey Hollowood.

(sometimes called 'holistic') between products rather than analytical descriptions that rely heavily on the ability to translate sensations into words and therefore can overlook hard-to-verbalize product characteristics. Similarity-based methods rely on an overall perceptual (i.e. non-verbal) evaluation of the products. The verbalization of differences – expressed by the participants and therefore dependent upon their subjective scale for perceiving and reporting such differences – between products occurs only in a second step or is altogether omitted. Similarity-based methodologies also have the advantage of bypassing the costly steps of panellist selection and training.

Many techniques exist to measure the similarity among a set of products, including pair or triadic similarity scaling, sorting tasks and projective mapping. Pair or triadic similarity scaling involves direct measurement of the similarity between products. In a pair similarity scaling task, panellists are provided with a pair of products and asked to rate the degree of similarity between the two products on an unstructured scale (e.g. 'please rate how similar these two products are by using a scale from 1 to 7 with "1" meaning very dissimilar and "7" meaning very similar'). In a triadic similarity scaling task, panellists are provided with three products and are asked to select the most similar and the most dissimilar pair of products.

By contrast with *direct* methods of similarity scaling, such as pairwise or triadic approaches, projective mapping and sorting tasks are *indirect* similarity measurements. In projective mapping, panellists are asked to position the products on a large sheet of paper according to their similarities and dissimilarities. In the sorting task, panellists are asked to sort/group products according to their similarities. Similarity scaling techniques have the advantage of providing a direct similarity measure but these techniques can be very time consuming (and are therefore likely to fatigue the assessors who could, consequently, produce poor-quality data) because the number of pairs or triads of products to be evaluated increases as an exponential function of the number of products in the set. For example, a set of 10 products would require each panellist to perform 45 pair comparisons and 120 triads. Projective mapping and the sorting task are much less time consuming but still provide reliable inter-product similarity.

In this chapter we will focus on these last two methods. After presenting them in detail, we discuss their advantages and disadvantages. In general, both the sorting task and projective mapping can be used with a number of products, similar to standard evaluation practice, with the caveat that all products need to be simultaneously presented to the assessors and therefore assessor fatigue or carry-over effects need to be taken into account when using these techniques (Chollet et al. 2014). We first describe projective mapping, then the sorting task, and finally we briefly compare these two techniques. We also illustrate these techniques with small examples.

15.2 Projective Mapping

The idea of projective mapping was first mentioned by Dunn-Rankin (1983) under the name of 'placing' to describe a technique in which assessors are asked to express the similarity structure of a set of stimuli by the stimuli relative positions on a plane; an equivalent idea can be found in the classic work of Coombs (1964). It was then reintroduced independently and simultaneously in the mid-1990s under the names 'projective mapping' by Risvik et al. (1994, 1997) in sensory evaluation and 'spatial arrangement procedure' (SAP) in psychology by Goldstone (1994). With projective mapping – also called more recently Napping® by Pagès (2003, 2005) (see also Lê et al. 2015), a term that reflects an amusing mixture of French and English (i.e. '*nappe*' means tablecloth in French and so 'napping' means 'table-clothing' rather than taking a nap…) – the assessors are asked to place the stimuli on a piece of paper in order to express the similarity structure of the stimuli. The procedure in SAP is similar, but the assessors position the stimuli on a computer screen. Because the name 'projective mapping' (or Napping) has been used mostly in sensory evaluation, we will use this denomination. In a recent variation of projective mapping, called 'Partial Napping' (Dehlholm 2015; Dehlholm et al. 2012), the assessors are asked to evaluate the products several times, according to one perceptual domain at a time (e.g. assessors can be asked to evaluate a set of wines on colour, aroma and mouthfeel) (Louw et al. 2013). In Partial Napping, the final analysis integrates all results into a global framework.

15.2.1 Methodology

Projective mapping is performed in a single session. All products (with a minimum number of eight and a maximum number depending upon the nature of the products) are presented simultaneously and are displayed on a table with a different random order for each assessor. Each assessor is presented with a large sheet of paper, often 60 × 40 cm, but of course, any size can be used as long as it provides ample space for panellists to comfortably separate products. A computer can also be used to record the data; see Savidan and Morris (2015) for an evaluation of the modalities used for projective mapping, and Lê et al. (2016) for a new computer-assisted way of collecting data. Assessors are asked first to evaluate by looking at, smelling, feeling and/or tasting (depending on the objectives of the study and on the products) all the products. They are then asked to position the products on the sheet of paper according to the products' patterns of similarities or dissimilarities. Assessors are told that two products should be placed very close to each other if they are perceived as identical and far from each other if they are perceived as very different. There is no further instruction as to how the samples should be separated in this space, and so each assessor chooses his/her own criteria.

Additionally, panellists can be instructed to write a few attributes/descriptors that they consider characteristic of each product or group of products. Integration of descriptive data into the analysis can enrich the interpretation of the product map generated. If a description step is implemented, it is crucial that (for both projective mapping and sorting) the description be performed *after* the participants have completed the first (non-verbal) part. The importance of separating these two steps is illustrated by a, now classic, experiment of Wilson and Schooler (1991). In this experiment, untrained participants had to rate, on a hedonic scale, five jams that had been previously evaluated by a panel of expert assessors (the experts performed this task for the magazine *Consumer Reports,* as part of a standard article). One group of naive participants simply evaluated the jams and their overall evaluation was significantly similar (with an average rank correlation of $r = 0.55$) to the experts' evaluation. By contrast, another group of naive participants were asked to provide reasons for their choice, and the evaluation performed by this group showed no correlation with the experts' evaluation.

15.2.2 Data Analysis

The X and Y co-ordinates of each sample are recorded on each assessor map, and compiled in a product-by-assessors table where each assessor contributes two columns representing respectively his or her X and Y co-ordinates (by tradition, the co-ordinates are measured with the upper left corner of the sheet of paper representing the origin of the measurements, but of course, any system will do as long as the units of measurement are the same for both dimensions). The data matrix is then submitted to a multivariate analysis to provide a sensory map of the products.

Originally, projective mapping data were analysed with principal component analysis (PCA) (Abdi 2007e; Abdi & Williams 2010a) or generalized Procrustes analysis (GPA) (Abdi 2003) and SAP with non-metric multidimensional scaling (MDS). More recently, Pagès (2003, 2005) proposed the use of multiple factor analysis (MFA) (Abdi & Valentin 2007a; Abdi et al. 2013; Escofier & Pagès 1990; Lê & Worch 2015; Pagès 2014) because this technique takes into account the differences between assessors. Other equivalent methods could be used such as INDSCAL (Bárcenas et al. 2004; Nestrud & Lawless 2011), STATIS (Abdi et al. 2012; Lavit 1988) or DISTATIS (Abdi & Valentin 2009; Abdi et al. 2005, 2007). The common goal of all these techniques is to provide a map of the products (called the 'compromise' map) such that the positions of the products on the map best reflect the similarity of the products as perceived by the group of assessors. The descriptors can also be projected on this map. Most techniques will also provide an indication of how each assessor 'interprets' the common space and some techniques (e.g. STATIS, DISTATIS and MFA) can also provide, around each product, confidence ellipsoids that can be used to assess the differences between products (Abdi et al. 2012; Cadoret & Husson, 2013; Husson et al. 2005). Some of these techniques (e.g. STATIS and DISTATIS) will also provide MDS-like maps of the assessors that can be used, for example, to identify outliers or groups of assessors.

15.2.3 Applications

Projective mapping has been applied to diverse food products such as, for example, chocolate (Kennedy & Heymann 2009; Risvik et al. 1994), commercial dried soup samples (Risvik et al. 1997), snack bars (Kennedy 2010; King et al. 1998), ewe's milk cheeses (Bárcenas et al. 2004), citrus juices (Nestrud & Lawless 2008), apples and cheese (Nestrud & Lawless 2011), and wines (Morand & Pagès 2006; Pagès 2003).

15.2.4 An Example: Projective Mapping and Curry Powder

In this example, projective mapping was carried out to evaluate the similarity between 13 curry powders in order to select four different curry powders for an upcoming 'spiciness' scoring test. Ten untrained assessors participated in the mapping. The order of presentation of the samples was performed according to a Latin square in order to distribute potential carry-over and adaptation effects (nevertheless, we could also use a random presentation if, for example, the number of samples is too high to allow an efficient and rapid service).

15.2.4.1 Instructions

Participants were provided with the following instructions.

> *You have in front of you 13 curry powders. Please smell all the samples and position them on the paper in such a way that similar samples are located near one another and different samples are placed far apart. You are free to evaluate the samples according to any criteria that you choose, and you will not need to specify your criteria. Feel free to use as much of the paper as is necessary to express the differences you may perceive. When you are finished, please mark the location of each sample with the corresponding number. You are free to take all the time that you need, and, if necessary, you can smell the crook of your elbow between each sample.*

15.2.4.2 Score Sheet

The score sheet was a 70 × 55 cm white sheet of paper (Figure 15.1).

15.2.4.3 Data Matrices

For each judge, the co-ordinates (X, Y) of each curry powder were measured and collected in a grand data table (Figure 15.2).

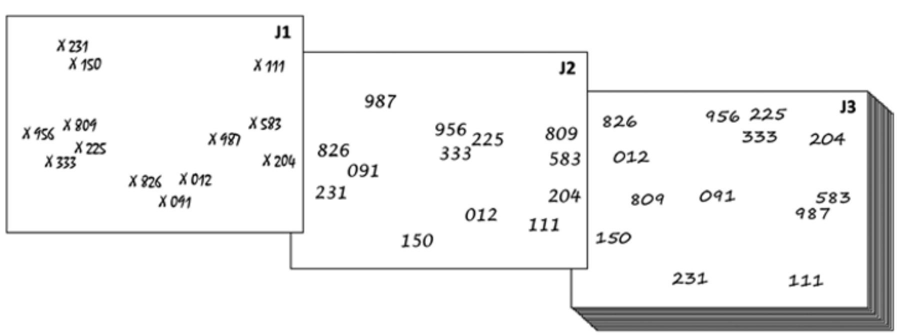
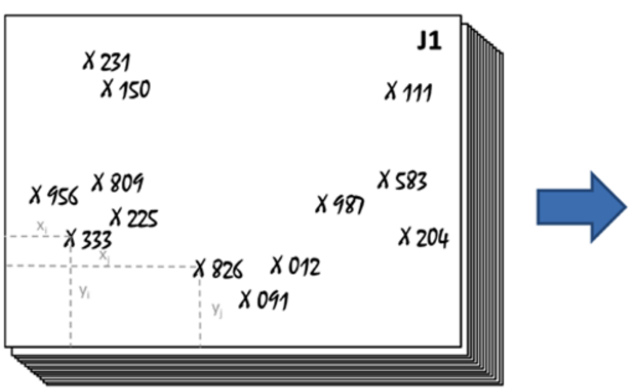


Figure 15.1 Example of score sheets obtained in the projective mapping test.



	J1		J2		J3		J4		J5		J6		J7		J8		J9		J10	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
P1	56	4	65,5	21	25,5	28,5	74,5	11,5	12	24	4,5	36,5	74	29	52	45,5	29,5	34	25,5	32
P2	48,7	30	46,5	14,8	16,2	35,5	78	6,5	15,5	45,5	22,5	26	11,5	36	2,5	2,5	9	7	17	38,5
P3	28	21,8	71,8	22,7	73,2	33	15,5	5	40	43,5	33,5	33	72,8	21,8	62	25	53	35	28	21
P4	34,5	27,5	31,5	14,8	83	33	12,5	11	56	17	47	30,8	79,5	28	79	32,5	61,5	27	43,5	28,5
P5	28	26	7	28,5	82,2	27	9,2	48	40	30	55,8	14,5	43,5	36	84,5	6	66,5	26,5	59	21,5
P6	52,2	7	62	46,2	25	35,5	6,5	30,2	55,2	44	46	54	67,5	18	39	40,5	42	10,8	15,5	42,5
P7	50	26	68,8	46,8	24,8	22	10,5	33	10,5	43,5	42,5	56	73	16,8	23	42,5	48,5	11,5	19,5	43
P8	76	23,2	20,5	36,5	72	25,8	35,5	47	52,5	21,5	57,5	17,5	42	16	82,5	23,5	72,5	42	57	27
P9	81	22,5	16	37	82	19,5	41	47	75,5	12	75	10	55,5	34,5	57,5	17	66	44,5	39	32
P10	53,8	29,2	60	21,5	36,5	35	78,8	10,5	35,2	29	48,5	21	66,5	25,5	40	47	25,5	38	45	33
P11	57,2	7	41,5	14,8	6,5	34,5	73,2	7	17	23,5	34,5	7,5	19	14,8	9,5	39	18	24	22,5	23
P12	33,5	22,5	36,5	14,8	16	28,5	7,5	8,2	17	29	10	33	31,5	35	5,5	21,5	15,5	28,5	22	19
P13	77,5	20,5	25,5	36	83,2	39	62	42,5	53	23	87	2	57,5	18,5	50	39	70,5	47	63	28

Figure 15.2 Example of data obtained in the projective mapping test.

15.2.4.4 Results

The data were analyzed with MFA, with each judge corresponding to one data table comprising two columns. As a preliminary step, an analysis of the judges was performed by first computing a matrix of the R_V coefficients computed between the judges (recall that the R_V coefficient can be interpreted as the squared coefficient of correlation between two configurations of data) (Abdi 2007a, 2010). A PCA (specifically, an eigen-decomposition) (Abdi 2007b) of this R_V matrix provides a map (Figure 15.3) in which the proximity between judges represents their similarity. Figure 15.3 shows that judges 2 and 7 differed from the other judges because these two judges positioned their curry powders differently from the other judges. Consequently, in the following steps of the analysis, the data of these two judges were eliminated. Even if the position of judge 1 is rather eccentric, we decided to keep this judge because in the second analysis this judge's position was close to the other judges (and, as we did not have very many judges, we needed to keep as many as possible; interestingly, however, the analysis performed without this judge was essentially the same as the analysis with this judge).

Inspection of Figure 15.4 suggests that there are four groups of curry powders: the first is composed of P1, P10 and P11, the second of P2, P6, P7 and P12, the third of P3, P4 and P5, and the last one of P8, P9 and P13. From this map, we can select four very different curry powders, for example, P11, P13, P5 and P2, but if we wanted to select very similar products we could choose P2, P6, P7 and P12.

15.3 Sorting Task

The sorting task originated in psychology (Hulin & Katz 1935), and this field has used it routinely ever since; see Coxon (1999) for a thorough review. The sorting task requires little effort and, actually, children as young as four years old can perform it (Best & Ornstein 1986). It was first used in the field of sensory evaluation in the early 1990s to investigate the perceptual structure of odors (Abdi et al. 2007; Chrea et al. 2005; Ishii et al. 1997; Lawless 1989; Lawless & Glatter 1990; MacRae et al. 1992; Stevens & O'Connell 1996). Lawless et al. (1995) were the first to use a sorting task with a food product. In sensory evaluation, the focus of the analysis is to derive similarity maps between the products and, in some cases, to derive, in addition, similarity maps among the panellists. By contrast with sensory evaluation, other disciplines (e.g. psychology, marketing) are sometimes interested only in analysing the similarity patterns between assessors (Hubert & Arabie 1985; Hubert & Levin 1976).

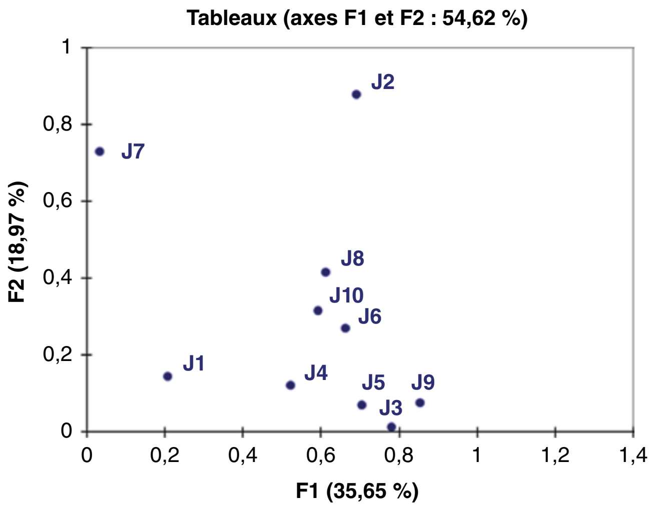


Figure 15.3 Map of the between-judges similarity obtained from analysis of the R_V coefficient table computed between the judges.

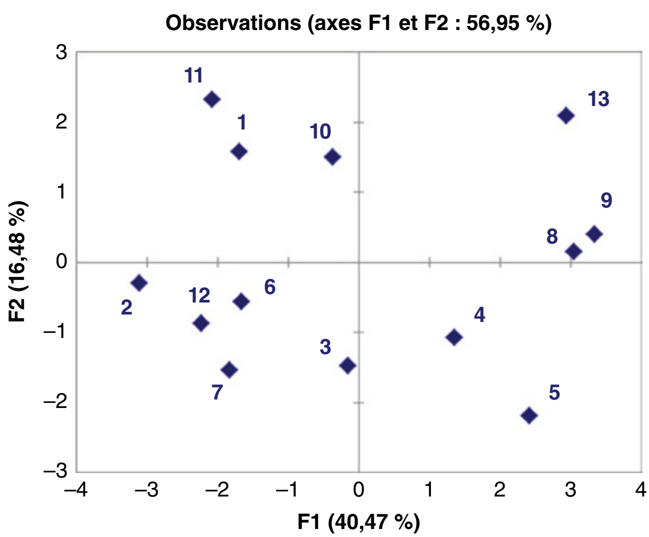


Figure 15.4 Compromise map for MFA. The data tables of judges 2 and 7 were not used to position the products.

15.3.1 Methodology

The sorting task is performed in a single session. All products are presented simultaneously and are displayed on a table with a different random order for each assessor. Assessors are asked first to look at, smell and/or taste (depending on the objectives of the study and the products) all the products and then to sort them into mutually exclusive groups based on perceived product similarities (see Figure 15.4). There is no further instruction as to how the samples should be grouped and so each assessor chooses his/her own criteria. Several variations of the sorting task exist where (1) the assessors are asked to further subdivide or regroup the products to produce a hierarchical distance matrix between objects (Rao & Katz 1971) (see also Santosa et al. 2010 for an application to sensory evaluation), or (2) to provide several sorts (Rosenberg & Kim 1975; Steinberg 1967) (see also Dehlholm et al. 2012 for an application to sensory evaluation under the name of free multiple sorting). Both of these methods have the advantage of providing a finer grained evaluation of the products because the distances between two products can take a larger range of values than with the standard sorting task.

15.3.2 Data Analysis

An overall similarity matrix is generated by counting the number of product co-occurrences (i.e. the number of times each pair of products was sorted in the same group). This similarity matrix is generally analysed by MDS (see Abdi 2007c,d for a review; see also Daws 1996 and Courcoux et al. 2015 for possible alternatives), which is a technique used to visualize proximities or distances (for a definition of distances, see Abdi 2007d) between objects in a low dimensional space. In MDS, each object is represented by a point on a map. The points are arranged on this map so that objects that are perceived to be similar to each other are placed near each other, and objects that are perceived to be different from each other are placed far away from each other. Different algorithms can be used to obtain the visual representation of the objects.

These algorithms can be classified into two main categories: metric MDS (also called classic MDS or principal co-ordinate analysis) and non-metric MDS. In metric MDS, the proximities are treated directly as distances. The input matrix is first transformed into a cross-product matrix and then submitted to a PCA. This method is optimum (i.e. it gives the best possible map to represent the data; see Abdi et al. 2007) when the distance matrix is a squared Euclidean distance matrix (which is the case for the distance matrix derived from a sorting task, as well as for hierarchical versions of the sorting task; see Abdi et al. 2007 for a proof).

The goal of non-metric MDS is to find a map that preserves, as well as possible, the *order* of the distances rather than their magnitude (as metric MDS does). Non-metric MDS is best suited for the analysis of dissimilarities or non-Euclidean distances rather than for the analysis of Euclidean distances. In non-metric MDS, the proximities are treated as ordinal data. An iterative stepwise algorithm is used to create a visual representation of the objects and roughly proceeds as follows:

1. the algorithm creates an arbitrary configuration
2. it computes the distances among all pairs of points

3 it compares the input matrix and the distance matrix using a stress function: the smaller the value of the stress, the greater the correspondence between the two matrices (the stress is roughly equivalent to a non-linear coefficient of alienation or the inverse of a squared coefficient of correlation)

4 it adjusts the object in the configuration in the direction that best decreases the stress.

Steps 2–4 are repeated until the value of the stress is small enough or cannot be decreased any more. Different authors have different standards regarding the amount of stress to tolerate. A common rule of thumb used in sensory evaluation is that any stress value below 0.2 is acceptable.

What method should we choose for analysing sorting data? Because the co-occurrence matrix is equivalent to a squared Euclidean distance (Abdi et al. 2007), metric MDS will provide the best map (i.e. a two-dimensional Euclidean distance) approximation of the data and should, therefore, be preferred. However, the differences between metric or non-metric MDS are rarely of importance in practice because the solutions provided by both methods are very similar when using sorting tasks and so the choice between methods is mostly a matter of convenience, personal preferences or habit (metric MDS, however, makes it easier to evaluate the stability of the results by cross-validation methods). Some authors have also used additive tree representations to describe sorting data (see Abdi 1990 for a review of these techniques, Chrea et al. 2004 for an example, and Chollet et al. 2011 for a review concerning sensory evaluation) or hierarchical analysis performed on the products factor co-ordinates obtained from a metric MDS (Lelièvre et al. 2009).

Multi-block analyses that take into account individual data such as DISTATIS (Abdi et al. 2007; Mielby et al. 2014), multiple correspondence analysis (MCA) (Abdi & Valentin 2007b; Cadoret et al. 2009; Lê & Worch 2015; Takane 1982) or common components and specific weights analysis (SORT CC) (Qannari et al. 2009) have also been used recently (INDSCAL could also be used). These techniques provide a common map (often called a compromise) and also show how each assessor positions the products in the common space. Some of these techniques (i.e. DISTATIS and FAST) (Cadoret et al. 2009) also provide a map of the assessors. The advantages of having a map of the assessors are:

- to be able to identify if some assessors differ from others (e.g. one assessor having a cold and whose tasting ability is, consequently, reduced)
- to be able to evaluate if groups of assessors systematically differ (e.g. do women and men differ in their way of sorting products?)
- to discover if there are some natural groups of assessors (as can be revealed with techniques such as *K*-means or hierarchical analysis) (Lê & Worch 2015; Beaton et al. 2016; Abdi & Beaton 2016).

Recall that in some cases, assessors are also asked to provide verbal descriptors for the groups of products after the sorting task has been performed. To facilitate the task, the assessors can be given a pre-established list of relevant terms (Lelièvre et al.

2008). The analysis methodology of the descriptors associated with the groups of products depends upon the authors and on the statistical techniques used for analysing the data (unfortunately, there is no current relevant review of the literature to compare these methodologies). However, current practices can essentially be regrouped in two main categories: most analyses will perform a separate analysis on the co-occurrence between words, and a few analyses, such as DISTATIS (Abdi & Valentin 2009) and FAST (Cadoret et al. 2009; Lê & Worch 2015), will project the words directly onto the factor map of the products.

The analyses using a co-occurrence table will first build a contingency table with descriptors in rows and products in columns. The values in the contingency table indicate the frequency with which each descriptor was employed for a stimulus. The descriptors given for a group of stimuli are assigned to each stimulus of the group and descriptors given by several assessors are assumed to have the same meaning. The resulting contingency tables are quite large and so the list of descriptors is generally reduced by grouping together terms with similar meanings (a procedure known as lemmatization) and by discarding descriptors used by fewer than a certain number of assessors (e.g. less than 10%). The frequency data can then be projected onto the similarity maps by computing the correlations between the occurrence of descriptors and the stimuli factor scores (Cartier et al. 2006; Faye et al. 2004). Alternatively, the contingency table can be submitted to correspondence analysis (CA) (Abdi & Béra 2014; Abdi & Williams 2010b), which will position both stimuli and descriptors on a descriptor-based space (Picard et al. 2003; Soufflet et al. 2004), to MCA (Cadoret et al. 2009) or to an MDS-like technique such as DISTATIS (Abdi & Valentin 2009; Abdi et al. 2005), which will also produce maps with stimuli and descriptors. Analyses projecting the words directly onto the factorial map will project these words using a barycentric rule: a word will be positioned such that its co-ordinates on the factors are proportional to the weighted average of the factor co-ordinates of the products with which this word was associated (Abdi & Valentin 2007b; Cadoret et al. 2009).

As an illustration, Figure 15.5 displays the results of the analysis of seven beers that were sorted by 11 assessors (Abdi & Valentin 2007a). After completion of the sorting task, the assessors were asked to describe with words the groups of beers that they made. Here, when more than one assessor used the same word, this word was counted as several occurrences of the same word; this procedure is equivalent to lemmatization. Note that other authors (e.g. Cadoret et al. 2009) prefer to project all the terms of a given assessor without equating equivalent terms. The co-ordinate of a word on a factor was computed as the weighted mean of the co-ordinates of the beers that were described by this word. This figure indicates that Dimension 1 separates Beers A and B (described as light, lemony, citrusy, etc.) from Beers C, D, E and F (perceived as rich in alcohol, bitter, coffee-like, etc.). Dimension 2 opposes Beers F and G (described as fruity, alcoholic and heavy) to Beers C and D (described as tasting like honey or coffee).

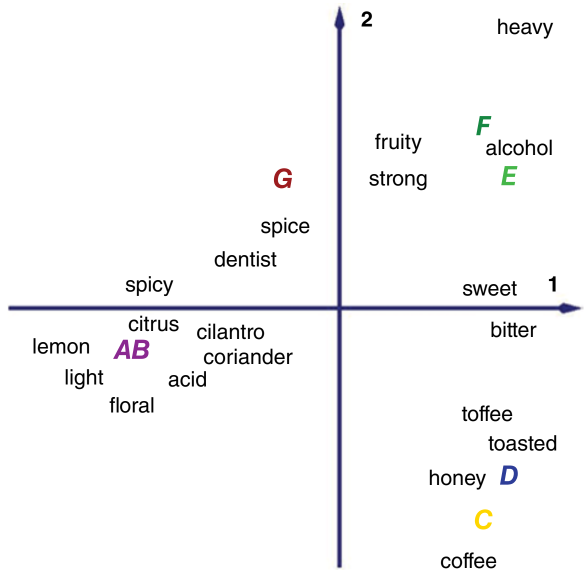


Figure 15.5 An example of the analysis results of a sorting task with products and descriptors. Here, seven beers (letters A–G) were sorted by 11 assessors who also described the groups of beers that they provided. The words have been lemmatized and projected as barycenters of the beers to which they have been associated. The analysis was performed with DISTATIS (for details and original data, see Abdi & Valentin 2007a).

15.3.3 Application

The free sorting task is used when the investigators want to rapidly obtain a sensory map of a set of products (often without searching for any verbal description). The simplicity of the sorting task made it particularly suited to be used with consumers (as opposed to trained assessors) and with participants who do not know the products and are unable or unwilling to verbally describe them in a reliable way, such as novice consumers or even children (Lê & Worch 2015; Valentin & Chanquoy 2012). The sorting task has been used on a large variety of food products (Chollet et al. 2014) including vanilla beans (Heymann 1994), cheese (Lawless et al. 1995; Nestrud & Lawless 2010), drinking waters (Falahee & MacRae 1995, 1997; Teillet et al. 2010), fruit jellies (Blancher et al. 2007; Tang & Heyman 1999), beers (Abdi et al. 2007; Chollet & Valentin 2001; Lelièvre et al. 2008, 2009), wines (Ballester et al. 2005; Bécue-Bertaut & Lê 2011; Piombino et al. 2004), yoghurts (Saint-Eve et al. 2004), spice aromas (Derndorfer & Baierl 2006), cucumbers and tomatoes (Deegan et al. 2010), apples (Nestrud & Lawless 2010) and perfumes (Veramendi et al. 2014). The sorting task has also been used to understand how consumers perceive food products such as meat or meat substitute products (Hoek et al. 2011), wine (Ballester et al. 2008; Campo et al. 2008) and beer (Lelièvre et al. 2009). Finally, the sorting task has been used for the sensory evaluation of non-food products such as, for example, automotive fabrics

(Giboreau et al. 2001; Picard et al. 2003), cloth fabrics (Soufflet et al. 2004), plastic pieces (Faye et al. 2004) and pictures of products such as olive oil bottles and containers (Mielby et al. 2014; Santosa et al. 2010).

15.3.4 An Example: Sorting Task and Orange Juice

In this example, in order to gain further insight into the development of new products, a manufacturer of orange juices wanted to know how consumers perceive the current orange juices available on the market. In particular, the manufacturer wanted to know if consumers perceived the sensory difference between, on the one hand, pure orange juices and orange juices from concentrate, and, on the other hand, flash pasteurized and pasteurized juices. In order to answer this question, a sorting task was performed with 12 orange juices varying in terms of formulation, manufacturing processes and presence or absence of pulp (see Table 15.1 for a description of the juices used in this experiment). Thirty-one orange juice consumers participated in the sorting task. They received 150 mL of each sample in a plastic tumbler and were provided with mineral water for rinsing their mouth between samples if necessary. A tasting session lasted approximately 15 minutes.

15.3.4.1 Questionnaire/Instructions

The participants in the experiment were given the following instructions.

> *You have 12 samples of orange juices in front of you. Please, look at, smell and taste these samples. Then make groups according to their similarity. You are free to make the groups according to any criteria that you may choose, and you do not need to specify your criteria. You can make as many groups as you want and group together as many orange juices as you want. You can take as much time as you want.*

15.3.4.2 Score Sheet

Examples of score sheets are shown in Figure 15.6.

Table 15.1 Information about the 12 orange juices used in the sorting task (the names of orange juices with pulp are printed in italics).

Code	Formulation	Process	Presence of pulp
P1	Pure juice	Flash pasteurized	Yes
P2	Pure juice	Flash pasteurized	Yes
P3	Pure juice	Flash pasteurized	Yes
P4	From concentrate	Flash pasteurized	No
P5	Pure juice	Pasteurized	Yes
P6	Pure juice	Pasteurized	Yes
P7	Pure juice	Pasteurized	Yes
P8	From concentrate	Pasteurized	No
P9	From concentrate	Pasteurized	No
P10	From concentrate	Pasteurized	Yes
P11	From concentrate	Pasteurized	No
P12	From concentrate	Pasteurized	No

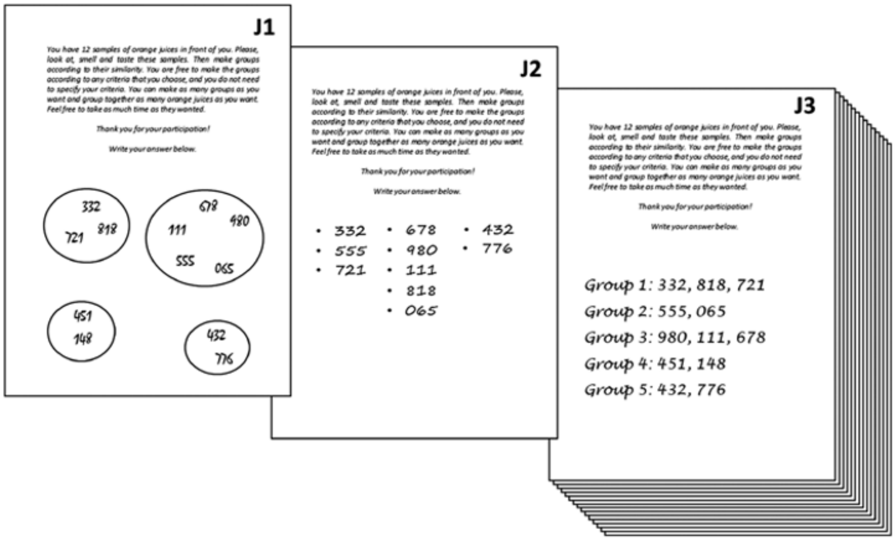


Figure 15.6 Example of score sheets obtained in a sorting task.

15.3.4.3 Data Matrices

For each judge, the results of the sorting task were encoded in an individual co-occurrence matrix where both the rows and the columns represent the orange juices. In this matrix, a value of 1 at the intersection of a row and a column indicates that the judge put these two orange juices together in the same group whereas a value of 0 indicates that the orange juices were not put together. To obtain the global similarity matrix, all the individual matrices were summed (Figure 15.7).

15.3.4.4 Results

Results were analysed with two statistical methods: metric MDS and DISTATIS.

15.3.4.4.1 Metric MDS

The data were analysed with a metric MDS performed on the global similarity matrix; we used metric MDS because the sorting task data generates a squared Euclidean distance (Abdi et al. 2007). In this representation, two orange juices that have often been sorted together by the judges will be positioned near each other and two oranges juices that have rarely been sorted together will be positioned far apart.

Figure 15.8 represents the MDS co-ordinates for the 12 orange juices used in this experiment. The figure suggests that there are three groups of orange juices: the first one composed of P1, P2, P6, P7 and P10, the second one of P3, P5, P8, P9, P11 and P12 and the last one of P4 only. The first group consists of orange

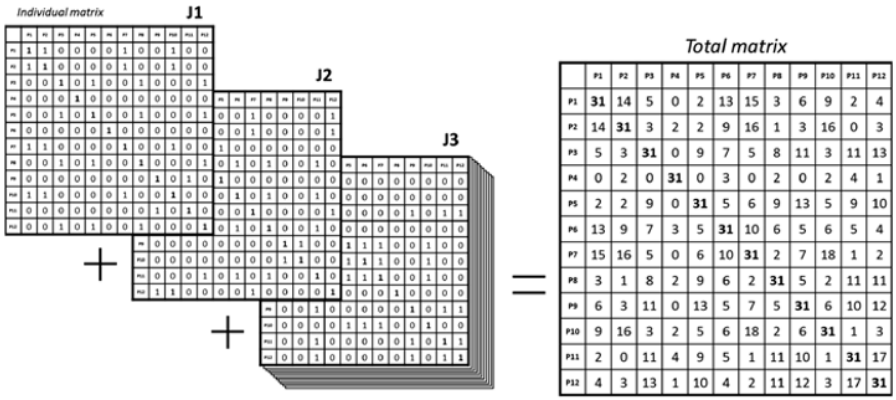


Figure 15.7 Example of data obtained in sorting task.

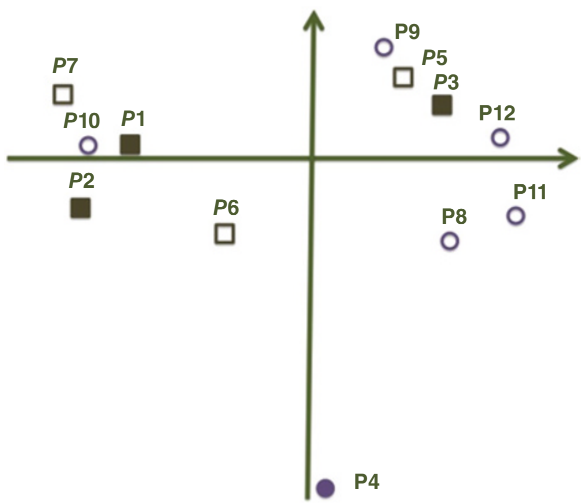


Figure 15.8 Two-dimensional metric MDS map. Squares represent pure juices, circles represent juices from concentrate, full symbols represent pasteurized juices, empty symbols represent sterilized juices and the labels printed in italics indicate the juices.

juices with pulp and rather pure juices (except P10). The second group essentially consists of sterilized orange juices (except P3). Product P4, positioned alone on the bottom of Dimension 2, is characterized by a very typically chemical orange aroma.

15.3.4.4.2 DISTATIS

DISTATIS (Abdi et al. 2012, 2007, 2005) is a method specifically developed for the analysis of sorting task (and projective mapping) data. It provides a compromise MDS-like map that optimally integrates all the judges' co-occurrence data

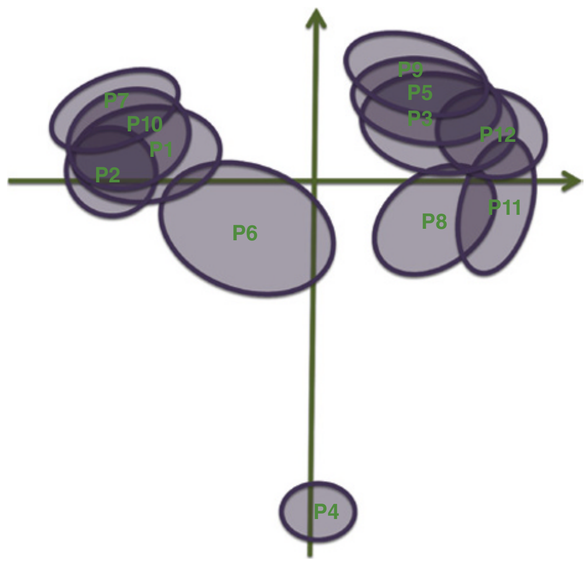


Figure 15.9 Two-dimensional DISTATIS map showing the products with their 95% confidence ellipsoids. When the confidence ellipsoids of two products do not intersect, these two products are perceived as significantly different by the judges. The configuration of the ellipsoids suggests that there are three groups of products.

tables in the sense that the matrix equivalent of an average squared coefficient of correlation (i.e. the R_V coefficient; see Abdi 2007a, 2010), computed between each judge's map and this map, is the largest possible one. On this map, we can also display 95% confidence ellipsoids for the products (obtained by using the cross-validation bootstrap technique of repeatedly sampling with replacement from the set of judges) (Abdi & Valentin 2009). In such a map, when the confidence ellipsoids of two products do not intersect, these two products can be considered as significantly different (Abdi et al. 2009). In addition, DISTATIS provides, for the judges, an MDS-like map where the proximity between the judges on the map reflects the similarity of their respective sorting data.

Figure 15.9 shows the products along with their confidence ellipsoids. As is often the case, the general solution of DISTATIS is very close to the metric MDS map; however, see Mielby et al. (2014) for a case where DISTATIS provided a slightly different and easier to interpret result than plain metric MDS. The configuration of the ellipsoids clearly confirms that the set of products comprises three clusters (as suggested by the MDS analysis). Figure 15.10 displays the map of the judges (obtained from the eigen-decomposition of the between-judge R_V matrix). In this map, the judges far to the right of Dimension 1 have a large commonality with the other judges whereas the judges close to the origin would be atypical. As most of the judges are positioned to the right, this indicates that they

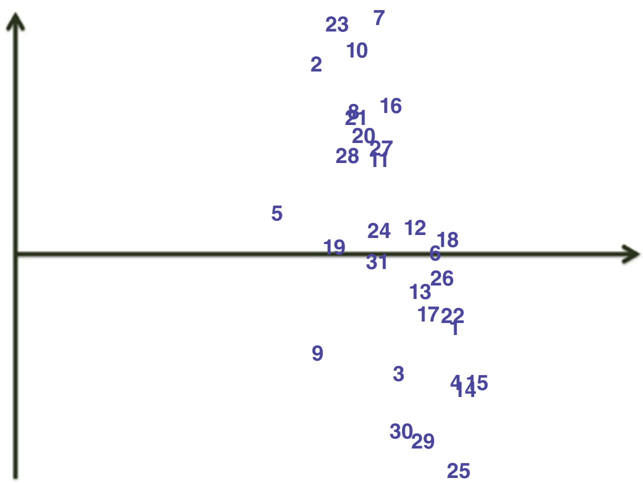


Figure 15.10 Two-dimensional DISTATIS map of the judges. The map suggests that the judges are rather homogeneous because no judge is far from the other judges.

mostly agree on their sorting. If a few judges were outliers, their data may be eliminated and the analysis re-run (as was done for the projective mapping example presented above).

15.4 Pros and Cons of Projective Mapping and the Sorting Task

Intuitively, these two methods (projective napping and sorting task) are somewhat similar. The main difference is that the sorting task forces the assessors to categorize the products in an all/none fashion and provides qualitative data whereas projective mapping does not force such a categorization and provides more graded data. Accordingly, we could expect projective mapping to provide more precise data. Yet, in a recent paper comparing the two approaches, Nestrud and Lawless (2010) reported that sorting and projective mapping of apple and cheese gave similar sensory maps but that a cluster analysis performed on the sensory maps was more easily interpretable for projective mapping than for sorting task, possibly due to a dimension reduction. However, this result still needs to be replicated and generalized to other products. As indicated by Valentin et al. (2012), a difficulty of projective mapping is to constrain the assessors to use two dimensions to discriminate between the products but this constraint may not be too much of a problem. For example, Goldstone (1994) compared SAP with two different types of pairwise comparisons of stimuli and found that the spaces, recovered by non-metric MDS, for these different techniques were highly correlated and that the solution obtained from SAP had the same dimensionality as the other methods. More recently, Nestrud and Lawless (2011) reached the same conclusion using 3D shapes.

From a practical point of view, Veinand et al. (2011) reported that projective mapping was difficult to perform, especially for assessors having problems with handling spatial information. Specifically, Veinand et al. observed that, *de facto*, many assessors performed a sorting task (grouping together samples on the paper map) rather than a real projective mapping. The sorting task, however, might also not be as easy as previously suggested. For example, Patris et al. (2007) filmed trained and novice assessors performing a beer sorting task and then invited them to comment on their behaviour at particular instances during the task. Both groups of assessors found the sorting task difficult to perform, especially novice assessors who reported memory difficulties as well as taste fatigue and saturation problems. The same problem might occur with projective mapping in addition to the spatial difficulties. Up to around 10% of subjects may have problems with both tasks, although these subjects are not always easy to identify in a particular study.

To sum up, both the sorting task and projective mapping completed or not with a verbalization step are time-effective ways of describing products as long as only a coarse description of the products is required. These methods can be used with both consumers and trained panellists on a relatively large set of products but might lead to memory difficulties, fatigue and saturation problems, especially with consumers (as opposed to trained panellists) if the number of products to be evaluated is too large (Chollet et al. 2014; Valentin et al. 2012). Also, some products may generate carry-over effects when tested simultaneously, and some types of products are likely to cause fatigue (e.g. beers or wines, because of alcohol and tannin; chili, mustard and other products with a trigeminal effect). The number of assessors required to perform the task differs depending upon the study, the products and the level of expertise of the participants. However, current literature (Chollet et al. 2014) suggests that 20 untrained assessors might be sufficient to reach stable configurations even though larger numbers (from 60 to 100) of participants are often reported; see Chollet et al. (2011, 2014) and Simiqueli et al. (2015) for a more detailed discussion of these aspects.

A common and important problem for both methods is that the whole set of products needs to be presented at the same time. Therefore, these methods are not suitable for hot products or for quality control, for instance. One solution is to use an incomplete block design but in this case, to obtain relevant results, a very large number of assessors is necessary. Another solution when evaluating products that can create fatigue, such as beers or wine, is to split the set of products in several smaller sets and to add in each smaller set the same product (called a prototype) and to compare the products to this prototype. In this case, the choice of prototype is obviously a crucial step. For both methods, we do not need to perform replication (as in conventional profile where we, in general, study the judges' repeatability) because several studies indicate that similar results are obtained in different replications as long as the products are quite different from each other (Chollet et al. 2011; Lelièvre et al. 2009). However, in order to have some estimation of the validity of the method, it is possible to duplicate one of the products and to check that these two duplicate products are close together on the product map.

All in all, projective mapping, free sorting and their diverse avatars (Ares 2014; Varela & Ares 2014) are now becoming part of the standard toolbox of sensory evaluation and are likely to be even more relevant as the field moves to rely more on (untrained) consumers to evaluate current products and help in developing new products.

15.5 Summary

Projective mapping and sorting tasks are used to obtain information about the similarity structure of a set of products. These tasks can be performed by experts, untrained consumers and adults as well as children. These tasks are, in general, analysed by statistical techniques related to PCA and provide maps that describe the similarity structure of the products under study (and, in some cases, the similarity structure of the assessors).

15.6 Future Developments

Projective mapping, the sorting task and the numerous variations on these techniques have been used for some time in psychology and related domains, and so these methodologies are unlikely to change much in the near future. Possible new developments could include new ways of acquiring the data (i.e. it is likely that computers will be used more often for collecting data,) and maybe a dynamical component in the data collection (i.e. looking at how assessors organize their answers across time) (Lê et al. 2016). Interesting developments are also likely to originate from new statistical techniques to analyse similarity data in order, for example, to reveal if different groups of assessors organize the perceptual space in different ways, or if some products can be eliminated with statistical techniques related to regularization or sparsification.

References

Abdi, H. (1986) Faces, prototypes and additive-tree representations. In: Ellis, H.D., Jeeves, M.A., Newcombs, F. & Young, A.W. (eds), *Aspects of Face Processing*. Dordrecht: Nijhoff.

Abdi, H. (1987) Do we really need a contingency model for concept formation? *British Journal of Psychology*, **78**, 113–125.

Abdi, H. (1990) Additive-tree representations (with an application to face processing). *Lecture Notes in Biomathematics*, **84**, 43–59.

Abdi, H. (2003) Multivariate analysis. In: Lewis-Beck, M., Bryman, A. & Futing, T. (eds) *Encyclopedia for Research Methods for the Social Sciences*. Thousand Oaks: Sage, pp. 699–702.

Abdi, H. (2007a) R_V coefficient and congruence coefficient. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 849–853.

Abdi, H. (2007b) Eigen-decomposition: eigenvalues and eigenvectors. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 304–308.

Abdi, H. (2007c) Metric multidimensional scaling. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 598–605.

Abdi, H. (2007d) Distance. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 280–284.

Abdi, H. (2007e) Singular value decomposition (SVD) and generalized singular value decomposition (GSVD). In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 907–912.

Abdi, H. (2010) Congruence: congruence coefficient, R_V coefficient, and Mantel coefficient. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 222–229.

Abdi, H. & Beaton, D. (2018) *Principal Component and Correspondence Analyses Using R*. New York: Springer Verlag.

Abdi, H. & Béra, M. (2014) Correspondence analysis. In: Alhajj, R. & Rokne, J. (eds) *Encyclopedia of Social Networks and Mining*. New York: Springer Verlag, pp. 275–284.

Abdi, H. & Valentin, D. (2007a) Multiple factor analysis. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 657–663.

Abdi, H. & Valentin, D. (2007b) Multiple correspondence analysis. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 651–657.

Abdi, H. & Valentin, D. (2009) Some new and easy ways to describe, compare, and evaluate products and assessors. In: Valentin, D., Nguyen, D.Z. & Pelletier, L. (eds) *New Trends in Sensory Evaluation of Food and Non-Food Products*. Ho Chi Minh City: Vietnam National University-Ho Chi Minh City Publishing House, pp. 5–18.

Abdi, H. & Williams, L.J. (2010a) Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, **2**, 433–459.

Abdi, H. & Williams, L.J. (2010b) Correspondence analysis. In: Salkind, N.J. (ed.) *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage, pp. 267–278

Abdi, H., Valentin, D., O'Toole, A.J. & Edelman, B. (2005) DISTATIS: the analysis of multiple distance matrices. Proceedings of the IEEE Computer Society: International Conference on Computer Vision and Pattern Recognition, San Diego, USA, pp. 42–47.

Abdi, H., Valentin, D., Chrea, C. & Chollet, S. (2007) Analyzing assessors and products in sorting tasks: DISTATIS, theory and application. *Food Quality & Preference*, **18**, 62–640.

Abdi, H., Dunlop, J.P. & Williams, L.J. (2009) How to compute reliability estimates and display confidence and tolerance intervals for pattern classiffers using the Bootstrap and 3-way multidimensional scaling (DISTATIS) *NeuroImage*, **45**, 89–95.

Abdi, H., Williams, L.J., Valentin, D. & Bennani-Dosse, M. (2012) STATIS and DISTATIS: optimum multi-table principal component analysis and three way metric multidimensional scaling. *Wiley Interdisciplinary Reviews: Computational Statistics*, **4**, 124–167.

Abdi, H., Williams, L.J. & Valentin, D. (2013) Multiple factor analysis: principal component analysis for multi-table and multi-block data sets. *Wiley Interdisciplinary Reviews: Computational Statistics*, **5**, 149–179.

Albert, A., Varela, P., Salvador, A., Hough, G. & Fiszman, S. (2011) Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality & Preference*, **22**, 463–473.

Ares, G. (2014) Methodological challenges in sensory characterization. *Current Opinion in Food Science*, **3**, 1–5.

Ballester, J., Dacremont, C., Le Fur, Y. & Etievant, P. (2005) The role of olfaction in the elaboration and use of the Chardonnay wine concept. *Food Quality & Preference*, **16**, 351–359.

Ballester, J., Patris, B., Symoneaux, R. & Valentin, D. (2008) Conceptual vs. perceptual wine spaces: does expertise matter? *Food Quality & Preference*, **19**, 267–276.

Ballester, J., Abdi, H., Langlois, J., Peyron, D. & Valentin, D. (2009) The odors of colors: can wine experts or novices distinguish the odors of white, red, and rosé wines? *Chemosensory Perception*, **2**, 203–213.

Bárcenas, P., Pérez Elortondo, F.J. & Albisu, M. (2004) Projective mapping in sensory analysis of ewes milk cheeses: a study on consumers and trained panel performance. *Food Research International*, **37**, 723–729.

Beaton, D., Dunlop, J., Abdi, H., Alzheimer's Disease Neuroimaging Initiative (2016) Partial least squares-correspondence analysis: a framework to simultaneously analyze behavioral and genetic data. *Psychological Methods*, **21**, 621–651.

Bécue-Bertaut, M. & Lê, S. (2011) Analysis of multilingual labeled sorting tasks: application to a cross-cultural study in wine industry. *Journal of Sensory Studies*, **26**, 299–310.

Best, D.L. & Ornstein, P.A. (1986) Children generation and communication of mnemonic organizational strategies. *Developmental Psychology*, **22**, 845–853.

Blancher, G., Chollet, S., Kesteloot, R., Nguyen, D., Cuvelier, G. & Sieffermann, J.M. (2007) French and Vietnamese: how do they describe texture characteristics of the same food? A case study with jellies. *Food Quality & Preference*, **18**, 560–575.

Cadoret, M. & Husson, F. (2013) Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality & Preference*, **28**, 106–115.

Cadoret, M., Lê, S. & Pagès, J. (2009) A factorial approach for sorting task data (FAST). *Food Quality & Preference*, **20**, 410–417.

Campo, E., Do, V.B., Ferreira, V. & Valentin, D. (2008) Sensory properties of a set of commercial Spanish monovarietal young white wines. A study using sorting tasks, lists of terms and frequency of citation. *Australian Journal of Grape and Wine Research*, **14**, 104–115.

Cartier, R., Rytz A., Lecomte A. et al. (2006) Sorting procedure as an alternative to quantitative descriptive analysis to obtain a product sensory map. *Food Quality & Preference*, **17**, 562–571.

Chollet, S. & Valentin, D. (2000) Le degré d'expertise a-t-il une influence sur la perception olfactive? Quelques éléments de réponse dans le domaine du vin. *L'Année Psychologique*, **100**, 11–36.

Chollet, S. & Valentin, D. (2001) Impact of training on beer flavor perception and description: are trained and untrained assessors really different? *Journal of Sensory Studies*, **16**, 601–618.

Chollet, S., Valentin, D. & Abdi. H. (2005) Do trained assessors generalize their knowledge to new stimuli? *Food Quality & Preference*, **16**, 13–23.

Chollet, S., Lelièvre, M., Abdi, H. & Valentin, D. (2011) Sort and beer: everything you wanted to know about the sorting task but did not dare to ask. *Food Quality & Preference*, **22**, 507–520.

Chollet, S., Valentin, D. & Abdi, H. (2014) Free sorting task. In: Tomasco, P.V. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: Taylor & Francis, pp 207–227.

Chrea, C., Valentin, D., Sulmont-Rossé, C., Ly Mai, H., Hoang Nguyen, D. & Abdi, H. (2004) Culture and odor categorization: agreement between cultures depends upon the odors. *Food Quality & Preference*, **15**, 669–679.

Chrea, C., Valentin, D., Sulmont-Rossé, C., Ly, M.H., Nguyen, D. & Abdi, H. (2005) Semantic, typicality and odor representation: a cross-cultural study. *Chemical Senses*, **30**, 37–49.

Cohen, H. & Lefebvre, C. (2005) *Handbook of Categorization in Cognitive Science*. New York: Elsevier.

Coombs, C.H. (1964) *A Theory of Data*. New York: Wiley.

Courcoux, P., Qannari, E.M. & Faye, P. (2015) Free sorting as s sensory profiling technique for product development. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods*. New York: Elsevier, pp. 153–185.

Coxon, A.P.M. (1999) *Sorting Data: Collection and Analysis*. Thousand Oaks: Sage.

Daws, J.T. (1996) The analysis of free-sorting data: beyond pairwise co-occurrences. *Journal of Classification*, **13**, 57–80.

Deegan, K.C., Koivisto, L., Näkkilä, L.J., Hyvönen, L. & Tuorila, H. (2010) Application of a sorting procedure to greenhouse-grown cucumbers and tomatoes. *Food Science & Technology*, **43**, 393–400.

Dehlholm, C. (2015) Free multiple sorting as a sensory profiling technique. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods*. New York: Elsevier, pp. 187–196.

Dehlholm, C., Brockhoff, P.B., Meinert, L., Aaslyng, M.D. & Bredie, W.L.P. (2012) Rapid descriptive sensory methods: comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. *Food Quality & Preference*, **26**, 267–277.

Derndorfer, E. & Baierl, A. (2006) Development of an aroma map of spices by multidimensional scaling. *Journal of Herbs, Spices & Medicinal Plants*, **12**, 39–50

Dunn-Rankin, P. (1983) *Scaling Methods*. Hillsdale: Lawrence Erlbaum.

Escofier, B. & Pagès, J. (1990) Multiple factor analysis. *Computational Statistics & Data Analysis*, **18**, 121–140.

Falahee, M. & MacRae, A.W. (1995) Consumer appraisal of drinking water: multidimensional scaling analysis. *Food Quality & Preference*, **6**, 327–332.

Falahee, M. & MacRae, A.W. (1997) Perceptual variation among drinking waters: the reliability of sorting and ranking data for multidimensional scaling. *Food Quality & Preference*, **8**, 389–394.

Faye, P., Brémaud, D., Durand-Daubin, D., Courcoux, P., Giboreau, A. & Nicod, A. (2004) Perceptive free sorting and verbalization tasks with naive subjects: an alternative to descriptive mappings. *Food Quality & Preference*, **15**, 781–791.

Giboreau, A., Navarro, S., Faye, P. & Dumortier, J. (2001) Sensory evaluation of automotive fabrics: the contribution of categorization tasks and non-verbal information to set-up a descriptive method of tactile properties. *Food Quality & Preference*, **12**, 311–322.

Goldstone, R. (1994) An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments & Computers*, **26**, 381–386.

Gombert, J.E., Fayol, M. & Abdi, H. (1990) Emploi des temps verbaux: approche typologique. Les catégories verbales et leur développement: étude descriptive. *CUMFID*, **16**, 87–98.

Heymann, H. (1994) A comparison of free choice profiling and multidimentional scaling of vanilla samples. *Journal of Sensory Studies*, **9**, 445–453.

Hoek, A.C., van Boekel, M., Voordouw, J. & Luning, P.A. (2011) Identification of new food alternatives: how do consumers categorize meat and meat substitutes? *Food Quality & Preference*, **22**, 371–383.

Hubert, L.J. & Arabie, P. (1985) Comparing partitions. *Journal of Classification*, **2**, 309–320.

Hubert, L.J. & Levin, H. (1976) Evaluating object set partitions: free-sort analysis and some generalizations. *Journal of Verbal Learning & Verbal Behavior*, **15**, 459–470.

Hulin, W.S. & Katz, D. (1935) The Frois-Wittmann pictures of facial expression. *Journal of Experimental Psychology*, **18**, 482–498.

Husson, F., Le Dien, S. & Pagès, J. (2005) Confidence ellipse for the sensory profiles obtained by Principal Component Analysis. *Food Quality & Preference*, **16**, 245–250.

Ishii, R., Kemp, S.E., Gilbert, A.N. &. O'Mahony, M. (1997) Variation in sensory conceptual structure: an investigation involving the sorting of odor stimuli, *Journal of Sensory Studies*, **12**,195–214.

Kennedy, J. (2010) Evaluation of replicated projective mapping of granola bars. *Journal of Sensory Studies*, **25**, 672–684.

Kennedy, J. & Heymann, H. (2009) Projective mapping and descriptive analysis of milk and dark chocolates. *Journal of Sensory Studies*, **24**, 220–233.

King, M.C., Cliff, M.A. & Hall, J.W. (1998) Comparison of projective mapping and sorting data collection and multivariate methodologies for identification of similarity-of-use snack bars. *Journal of Sensory Studies*, **13**, 347–358.

Lavit, C. (1988) *Analyse Conjointe de Tableaux Quantitatifs*. Paris: Masson.

Lawless, H.T. (1989) Exploration of fragrances categories and ambiguous odors using multidimensional scaling and cluster analysis. *Chemical Senses*, **14**, 349–360.

Lawless, H.T. & Glatter, S. (1990) Consistency of multidimensional scaling models derived from odor sorting. *Journal of Sensory Studies*, **5**, 217–230.

Lawless, H.T., Sheng, N. & Knoops, S.S.C.P. (1995) Multidimensional scaling of sorting data applied to cheese perception. *Food Quality & Preference*, **6**, 91–98.

Lê, M.T., Husson, F. & Lê, S. (2016) Digit-tracking: interpreting the evolution over time of sensory dimensions of an individual product space issued from napping and sorted napping. *Food Quality & Preference*, **47**, 73–78.

Lê, S. & Worch, T. (2015) *Analyzing Sensory Data with R*. Boca Raton: CRC Press.

Lê, S., Lê, M. T. & Cadoret, M. (2015) Napping and sorting napping as a sensory profiling technique. In: Delarue, J., Lawlor, J.B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods*. New York: Elsevier, pp. 197–213.

Lelièvre, M., Chollet, S., Abdi, H. & Valentin, D. (2008) What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality & Preference*, **19**, 697–703.

Lelièvre, M., Chollet, S., Abdi, H. & Valentin, D. (2009) Beer trained and untrained assessors rely more on vision than on taste when they categorize beers. *Chemosensory Perception*, **2**, 143–153.

Louw, L., Malherbe, S., Naes, T., Lambrechts, M., van Rensburg, P. & Nieuwoudt, H. (2013) Validation of two Napping techniques as rapid sensory screening tools for high alcohol products. *Food Quality & Preference*, **30**, 192–201.

MacRae, A.W., Howgate, P. & Geelhoed, E.N. (1990) Assessing the similarity of odours by sorting and by tridiadic comparison. *Chemical Senses*, **15**, 661–699.

MacRae, A.W., Rawcliffe, T., Howgate, P. & Geelhoed, E.N. (1992) Patterns of odour similarity among carbonyls and their mixtures. *Chemical Senses*, **17**, 119–125.

Mervis, C. & Rosch, E. (1981) Categorization of natural objects. *Annual Review of Psychology*, **32**, 89–115.

Mielby, L.H., Hopfer, H., Jensen, S., Thybo, A.K. & Heymann, H. (2014) Comparison of descriptive analysis, projective mapping and sorting performed on pictures of fruit and vegetable mixes. *Food Quality & Preference*, **35**, 86–94.

Morand, E. & Pagès, J.(2006) Procrustes multiple factor analysis to analyze the overall perception of food products. *Food Quality & Preference*, **17**, 36–42.

Moussaoui, K.A. & Varela, P. (2010) Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality & Preference*, **21**, 1088–1099.

Murphy, G.L. (2002), *The Big Book of Concepts*. Cambridge: MIT.

Nestrud, M.A. & Lawless, H.T. (2008) Perceptual mapping of citrus juices using projective mapping and profiling data from culinary professionals and consumers. *Food Quality & Preference*, **19**, 431–438.

Nestrud, M.A. & Lawless, H.T. (2010) Perceptual mapping of apples and cheeses using projective mapping and sorting. *Journal of Sensory Studies*, **25**, 390–405.

Nestrud, M.A. & Lawless, H.T. (2011) Recovery of subsampled dimensions and configurations derived from napping data by MFA and MDS. *Attention, Perception & Psychophysics*, **73**, 1266–1278.

Pagès, J. (2003) Recueil direct de distances sensorielles: application à l'évaluation de dix vins blancs de Val de Loire. *Science des Aliments*, **23**, 679–888.

Pagès, J. (2005) Collection and analysis of perceived product inter-distances using multiple factor analysis: application to the study of 10 white wines from the Loire Valley. *Food Quality & Preference*, **16**, 642–664.

Pagès, J. (2014) *Multiple Factor Analysis by Example Using R*. Boca Raton: CRC Press.

Patris, B., Gufoni, V., Chollet, S. & Valentin, D. (2007) Impact of training on strategies to realize a beer sorting task: behavioral and verbal assessments. In: Valentin, D., Nguyen, D.Z. & Pelletier, L. (eds) *New Trends in Sensory Evaluation of Food and Non-Food Products*. Ho Chi Minh City: Vietnam National University-Ho Chi Minh City Publishing House.

Perrin, L. & Pagès, J. (2009) Construction of a product space from the Ultra-flash profiling method: application to ten red wines from the Loire Valley. *Journal of Sensory Studies*, **24**, 372–395.

Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F. & Pagès, J. (2008) Comparison of three sensory methods for use with the Napping® procedure: case of ten wines from the Loire valley. *Food Quality & Preference*, **19**, 1–11.

Picard, D., Dacremont, C., Valentin, D. & Giboreau, A. (2003) Perceptual dimensions of tactile textures. *Acta Psychologica*, **114**, 165–184.

Piombino, P., Nicklaus, S., LeFur, Y., Moio, L. & Le Quéré, J. (2004) Selection of products presenting given flavor characteristics: an application to wine. *American Journal of Enology & Viticulture*, **55**, 27–34.

Qannari, E.M., Cariou, V., Teillet, E. & Schlich, P. (2009) SORT-CC: a procedure for the statistical treatment of free sorting data. *Food Quality & Preference*, **21**, 302–308.

Rao, V.R. & Katz, R. (1971) Alternative multidimensional scaling methods for large stimulus sets. *Journal of Marketing Research*, **8**, 488–494.

Risvik, E., McEwan, J.A., Colwill, J.S., Rogers, R. & Lyon, D.H. (1994) Projective mapping: a tool for sensory analysis and consumer research. *Food Quality & Preference*, **5**, 263–269.

Risvik, E., McEwan, J.A. & Redbottena, M. (1997) Evaluation of sensory profiling and projective mapping. *Food Quality & Preference*, **8**, 63–71.

Rosch, E. (1973) On the internal structure of perceptual and semantic categories. In: Moore, T.E. (ed.) *Cognitive Development and the Acquisition of Language*: New York: Academic Press, pp. 111–144.

Rosch, E. (1978) Principles of categorization. In: Rosch, E. & Lloyd, B.B. (eds) *Semantic Factors in Cognition*. Hillsdale: Erlbaum, pp 137–168.

Rosch, E.H. & Mervis, C.B. (1981) Categorization of natural objects. *Annual Review of Psychology*, **32**, 89–113.

Rosenberg, S. & Kim, P. (1975) The method of sorting as a data gathering procedure in multivariate research. *Multivariate Behavioral Research*, **10**, 489–502.

Saint-Eve, A., Paçi Kora, E. & Martin, N. (2004) Impact of the olfactory quality and chemical complexity of the flavouring agent on the texture of low fat stirred yogurts assessed by three different sensory methodologies. *Food Quality & Preference*, **15**, 655–668.

Santosa, M., Abdi, H. & Guinard, J.X. (2010) A modified sorting task to investigate consumer perceptions of extra virgin olive oils. *Food Quality & Preference*, **21**, 881–892.

Savidan, C.H. & Morris, C. (2015) Panellists' performances and strategies in paper-based and computer-based projective mapping. *Journal of Sensory Studies*, **30**, 145–155.

Sha, L., Haxby, J.V., Abdi, H. et al. (2015) The animacy continuum in the human ventral vision pathway. *Journal of Cognitive Neuroscience*, **27**, 665–678.

Simiqueli, A.A, Minim, V.P.R., de Cássia dos Santos Navarro da Silva, R., da Silva, A.N. & Minim, L.A. (2015) How many assessors are necessary for the Optimized Descriptive Profile when associated with training? *Food Quality & Preference*, **44**, 62–69.

Soufflet, I., Calonnier, M. & Dacremont, C. (2004) A comparison between industrial experts' and novices' haptic perceptual organization: a tool to identify descriptors of handle of fabrics. *Food Quality & Preference*, **15**, 689–699.

Steinberg, D.D. (1967) The Word Sort: an instrument for semantic analysis. *Psychonomic Science*, **8**, 541–542.

Stevens, D.A. & O'Connell, R.J. (1996) Semantic-free scaling of odor quality. *Physiological Behavior*, **60**, 211–215.

Takane, Y. (1982) IDSORT: an individual differences multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation*, **14**, 546.

Tang, C. & Heymann, H. (1999) Multidimensional sorting, similarity scaling and free choice profiling of grape jellies. *Journal of Sensory Studies*, **17**, 493–509.

Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010) Sensory methodologies and the taste of water. *Food Quality & Preference*, **21**, 967–976.

Tversky, A. (1977) Features of similarity. *Psychological Reviews*, **84**, 327–352.

Valentin, D. & Abdi, H. (2001) Face recognition by myopic baby neural networks. *Child & Infant Development*, **10**, 19–20.

Valentin, D. & Chanquoy, L. (2012) Olfactory categorization: a developmental study. *Journal of Experimental Child Psychology*, **113**, 337–352.

Valentin, D., Chollet, S., Lelièvre, M. & Abdi, H. (2012) Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, **47**, 1563–1578.

Varela, P. & Ares, G. (2014) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press.

Veinand, B., Godefroy, C., Adam, C. & Delarue, J. (2011) Highlight of important product characteristics for consumers. Comparison of three sensory descriptive methods performed by consumers. *Food Quality & Preference*, **22**, 474–485.

Veramendi, M., Herencia, P. & Ares, G. (2014) Perfume odor categorization: to what extent trained assessors and consumers agree? *Journal of Sensory Studies*, **28**, 76–89.

Wilson, T.D. & Schooler, J.W. (1991) Thinking too much: introspection can reduce the quality of preferences and decisions. *Journal of Personality & Social Psychology*, **60**, 181–192.

Wittgenstein, L. (1953) *Philosophical Investigations*. Oxford: Blackwell.

CHAPTER 16

Polarized Sensory Positioning

Gastón Ares, Lucía Antúnez, Luis de Saldamando and Ana Giménez

16.1 Introduction

Sensory characterization has been traditionally performed using descriptive analysis (DA) (Lawless & Heymann 2010). In this methodology, assessors are trained in the identification and quantification of specific sensory attributes and references are used for attribute recognition and scaling (Heymann et al. 2014). This approach enables detailed, reliable and repeatable information to be obtained (Murray et al. 2001). Data from DA are stable in time and therefore results obtained in different sessions can be compared (Lawless & Heymann 2010).

Interest in rapid methodologies for sensory characterization has largely increased in the last decade (Valentin et al. 2012; Varela & Ares 2012). These methodologies can be implemented in short time frames and with naive assessors. Among rapid approaches for sensory characterization, holistic methodologies have become salient. They are based on the evaluation of global similarities and differences among samples (Ares & Varela 2014). In these methodologies, assessors should form a synthetic representation of the similarities and differences among samples by relying on a process of synthesis for analysing and processing sensory information (Jaeger et al. 2000). This process determines the relative importance of the perceived sensory characteristics for estimating the similarities and differences among samples.

One of the main disadvantages of holistic methodologies is that all samples should be simultaneously evaluated in the same session. This reduces the number of samples that can be considered in the sensory characterization study, particularly when dealing with complex products (e.g. wine), samples that have intense or persistent sensory characteristics (e.g. chili or distillates) or that require careful temperature control. Besides, conclusions about samples evaluated in different sessions cannot be compared due to the comparative nature of

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

the methodology. This poses limitations to the application of holistic methodologies in new product development, which has been characterized as an iterative process (Costa & Jongen 2006).

In this context, polarized sensory positioning was developed by Teillet et al. (2010).

16.2 Description of Polarized Sensory Positioning

Polarized sensory positioning (PSP) is a reference-based method for sensory characterization (Varela & Ares 2012). It is based on the comparison of samples with a set of fixed references, called poles (Teillet et al. 2010). The comparison of samples and poles is based on a holistic evaluation since assessors evaluate global differences, without any indication about the sensory attributes that should be considered in the evaluation or their relative importance.

There are basically two types of PSP, which rely on different types of evaluation, and consequently, data analysis techniques.

16.2.1 Polarized Sensory Positioning Based on Degree of Difference Scales

In the original type of PSP, assessors have to quantify the overall degree of difference between each sample and the poles using unstructured scales, ranging from 'exactly the same' to 'totally different'. An example of the scales is shown in Figure 16.1.

Although most applications of PSP have used unstructured 10 cm scales, any type of intensity scale can be used, including structured scales (Lawless & Heymann 2010).

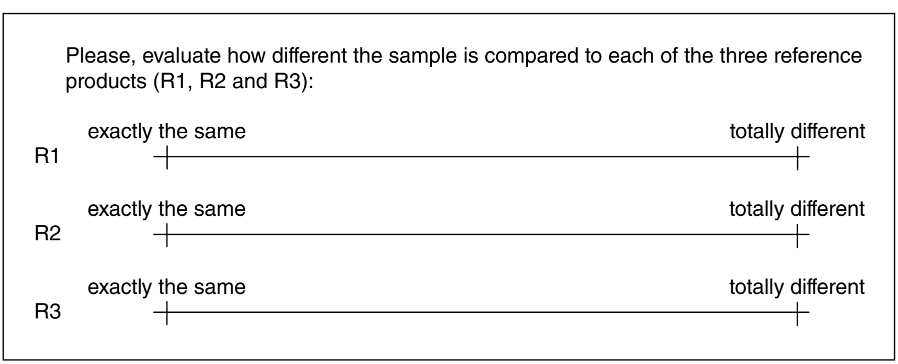


Figure 16.1 Example of the evaluation sheet used in polarized sensory positioning with continuous scales to compare samples with three reference products or poles (R1, R2 and R3).

Please, indicate to which of the reference products (R1, R2 and R3) each sample is most similar to and most different from.

Sample 325

To which of the reference products is this sample most similar to?

☐ ☐ ☐
R1 R2 R3

To which of the reference products is this sample most different from?

☐ ☐ ☐
R1 R2 R3

Figure 16.2 Example of the evaluation sheet used in triadic polarized sensory positioning to evaluate samples.

16.2.2 Triadic Polarized Sensory Positioning

Triadic polarized sensory positioning is an alternative approach to PSP, which consists of asking assessors to indicate to which of the poles a sample is most similar and from which of the poles it is most different. An example of an evaluation sheet is shown in Figure 16.2.

The advantage of this approach is that it does not require untrained assessors to use scales and, therefore, it is easier and simpler for assessors (de Saldamando et al. 2013).

16.3 Practical Considerations for the Implementation of Polarized Sensory Positioning

Several practical considerations should be taken into account when using PSP for sensory characterization, which are discussed below.

16.3.1 Selection of the Poles

Polarized sensory positioning is based on comparison of samples with a fixed set of reference products or poles. Thus, selection of the poles is probably the most important step for the implementation of the methodology. Although research is still lacking on the influence of the poles on results from PSP, some basic recommendations can be provided.

The usual number of poles in PSP studies is three. This number is used for stabilizing a two-dimensional sensory space based on the idea that three points define a plane. If three or four dimensions are expected to be necessary to represent the sensory space of interest, the number of poles should be augmented. However, Ares et al. (2015) showed that it is possible to obtain a stable bidimensional sensory based on two poles and that the number of poles should be related to the number of sensory dimensions that discriminate among samples. As a general rule, the number of poles should be at least equal to the number of sensory dimensions that are necessary for representing the perceptual space. For example, if the sample set is expected to be represented in a two-dimensional space, a minimum of two poles would be necessary. Considering additional dimensions usually increases ability to discriminate among samples, particularly in terms of sensory characteristics that are not well represented by the poles (Ares et al. 2015). However, it should be considered that increasing the number of poles makes the task more difficult and tedious, as well as increasing sensory fatigue.

Regarding the selection of the poles, they should represent the main sensory characteristics responsible for the expected similarities and differences among samples of interest (Ares et al. 2015). As an example, when studying the sensory characteristics of commercial mineral waters, Teillet et al. (2010) selected three poles with different degrees of mineralization. According to these authors, the sensory characteristics of mineral water are mainly determined by the concentration of dissolved minerals. Waters with low mineralization are characterized by their metallic and bitter taste, waters with medium mineral content show neutral and fresh taste, while high mineralization provides a salty taste. For this reason, the authors selected three samples with different degrees of mineralization as poles: one with low mineral content, one with intermediate mineralization and a third one with high mineral content. Similarly, de Saldamando et al. (2013) selected three distinct textures for sensory characterization of make-up foundations: liquid, mousse and cream.

The selection of the poles can be carried out using results from previous sensory characterization studies with the same product category or preliminary studies using rapid methodologies. For example, Ares et al. (2013) used a projective mapping task to identify the sensory characteristics responsible for the main differences among commercial samples of orange-flavoured powdered drinks. Three main groups of samples were identified: one characterized by its low total flavour intensity, a second one by its sourness and a third group was described as sweet and with intense orange flavour. This grouping was in agreement with market positioning and product composition. Samples from the economy segment were characterized by their low total flavour intensity, samples from the premium segment formulated with sugar were characterized by their sweetness and intense orange flavour, whereas samples from the premium segment formulated with low-calories sweeteners were described as sour. Based on these results, Ares et al. (2013) selected one pole from the economy sector and two samples of the premium segment, one formulated with sugar and the other with low-calorie sweeteners.

Recent research has shown that the selection of the poles does not largely affect results from sensory characterizations as long as they represent the main characteristics responsible for similarities and differences among samples (Ares et al. 2015). In this sense, de Saldamando et al. (2013) studied the influence of the poles on results from PSP with consumers. They reported that sensory spaces obtained with different set of poles were highly similar.

Another relevant aspect that should be considered when selecting samples to serve as poles in PSP tasks is availability. Ideally, poles should be stable during the evaluation period and available any time.

16.3.2 Number of Samples

There are basically no limitations on the number of samples that can be evaluated using PSP. The main advantage of PSP over other comparative methodologies is that samples can be evaluated in different sessions. Therefore, if the number of samples is large, the evaluation can be split into different subsets, to be evaluated in different sessions. Teillet (2014) has shown that data from different sessions can be aggregated and that results are similar to those obtained in a single session.

One consideration that should be taken into account when selecting the number of samples for a sensory characterization study with PSP is that at least five samples are necessary if the aim is to obtain a representation of the samples in a bidimensional sensory space.

16.3.3 Type and Number of Assessors

Polarized sensory positioning can be implemented with both trained assessors and naive consumers. The first application of PSP involved assessors trained in the evaluation of the sensory characteristics of mineral water (Teillet et al. 2010). However, de Saldamando et al. (2013), Ares et al. (2013, 2015) and Cadena et al. (2014) have used naive consumers.

Research on the minimum number of assessors needed for performing PSP tasks is still lacking. Traditionally, the number of trained assessors considered for the implementation of rapid methodologies is similar to that considered in descriptive analysis, ranging from 10 to 15 (Lawless & Heymann 2010; Varela & Ares 2012).

Regarding the number of naive assessors that should be considered when using PSP, research on novel methodologies for sensory characterization has shown that a minimum of 30–60 consumers should be considered (Ares et al. 2014; Blancher et al. 2012; Vidal et al. 2014). Cadena et al. (2014) used 81 consumers for the evaluation of yoghurts and concluded that stable sample configurations could be obtained with only 25. However, it should be taken into account that sensory complexity and degree of difference among samples are expected to have a large influence on the number of consumers needed to obtain reliable results. The number of consumers is expected to increase as complexity increases and the degree of difference among samples is reduced.

16.3.4 Replication

Replication is not usually performed in PSP tasks due to constraints in the time and/or resources available for conducting the study and analysing results. However, it is easier to replicate the study when working with trained assessors than when carrying out a consumer study. When naive consumers are used, tools for studying the internal reliability of the data are necessary. A recent approach for evaluating data reliability based on a bootstrapping resampling approach is presented in the next section.

16.4 Data Analysis

Data analysis techniques vary according to the specific type of PSP being considered. Strategies for analysing data from PSP based on degree of difference scales and triadic PSP are discussed below.

16.4.1 Polarized Sensory Positioning Based on Degree of Difference Scales

When scales are used to estimate the degree of difference between each sample and the poles, data basically consist of three degrees of difference scores, one for each of the poles.

Differences among samples can be analysed using ANOVA based on mixed models (Brockhoff 2003) to identify significant differences among samples in their degree of difference with each of the poles. However, this approach has not been commonly used for analysing PSP data.

The most common analysis of PSP data aims at obtaining a low-dimensional representation of samples, which enables similarities and differences among samples to be determined.

Sample configurations can be obtained from averaged data using principal component analysis (PCA) or multidimensional scaling (MDS) unfolding, which is used for rectangular matrices (Green et al. 1989). An example of the data matrix needed to implement these techniques is shown in Figure 16.3. Both analyses can be performed in R language (R Core Team 2013). The `FactoMineR` package (Lê et al. 2008) can be used to perform PCA, whereas the `Smacof` package can be used for performing MDS unfolding (Mair et al. 2013).

Alternatively, data can be analysed using multiblock statistical techniques. These techniques are highly recommended for PSP data since each assessor evaluates the degree of difference between samples and the poles considering their own personal criteria. For example, one assessor can attribute a high relative importance to sweetness for their difference ratings, while another assessor can mainly consider thickness for scoring the degree of difference. Therefore, a consensus representation of the similarities and differences among samples can be obtained using STATIS, multiple factor analysis (MFA) or generalized Procrustes

Sample	R1	R2	R3
1	0.4	1.3	9.9
2	2.2	5.8	7.8
…	…	…	…
J	4.2	9.8	0.4

Figure 16.3 Example of the data matrix used for analysing average PSP data using principal component analysis or multidimensional scaling unfolding. Each column (R1, R2, R3) represents the average degree of difference between a sample and one of the reference products or poles (R1, R2 and R3 respectively), across all assessors.

	Assessor 1			Assessor 2				Assessor N		
Sample	R1	R2	R3	R1	R2	R3	…	R1	R2	R3
1	0.4	1.3	9.9	4.0	8.6	7.5	…	1.3	4.8	7.4
2	1.2	5.8	7.8	0.9	2.6	5.4	…	2.4	3.5	8.7
…	…	…	…	…	…	…	…	…	…	…
J	3.2	4.8	6.4	2.8	1.9	9.4	…	1.9	2.5	7.4

Figure 16.4 Example of the data matrix used for analysing PSP data using multiple factor analysis, STATIS or generalized Procrustes analysis. Each group of columns (R1, R2, R3) contains the degree of difference between a sample and one of the reference products or poles (R1, R2 and R3 respectively), for one of the assessors.

analysis (GPA). In this analysis, the data matrix should be constructed by stacking the scores of each assessor next to each other, as indicated in Figure 16.4.

When MFA is used for obtaining sample configurations, confidence ellipses around samples can be constructed to represent the area of the space with a certain probability of containing the real position of samples, which enables significant differences among samples to be identified. Confidence ellipses can be constructed in R language using partial or total bootstrapping. Details about how to construct confidence ellipses can be found in Dehlholm et al. (2012) and Cadoret and Husson (2013).

Furthermore, the stability of sample configurations can be evaluated using a bootstrappping resampling approach, as suggested by Faye et al. (2006) and Blancher et al. (2012) for sorting tasks. Sample configurations can be considered stable if similar results are obtained in repeated experiments under the same conditions. Therefore, repeated experiments can be simulated by repeatedly sampling from the population of interest using bootstrapping. Random subsets of different size (m) are generated from the original data set of N assessors. The bootstrapping process consists of random draws with replacement of the corresponding number of assessors (m) from the original data set. For each m, a large number of random subsets (usually 1000) are obtained and for each subset, a sample configuration is obtained. The agreement between each of these configurations and the reference configuration (obtained with all the assessors) is evaluated by computing the RV coefficient (Abdi 2010). The RV coefficient is a measure of the similarity between two factorial configurations, which takes the value of 0 if the configurations are uncorrelated and the value of 1 if the configurations are homothetic. This coefficient depends on the relative position of the points in the configuration and therefore is independent of rotation and translation (Robert & Escoufier 1976). The average RV coefficient across simulations is calculated for each number of assessors in the subset. Blancher et al. (2012) proposed an RV coefficient of 0.95 for a sample configuration to be regarded as stable.

16.4.2 Triadic Polarized Sensory Positioning

Data from triadic PSP are qualitative and correspond to the pole that each assessor considered the most and least similar to each of the samples. There are basically two ways of analysing these data.

One of the options is to construct a frequency table that summarizes assessors' responses and then apply correspondence analysis (CA). In this approach, the number of assessors who selected each of the poles as the most similar to each of the samples is determined, as well as the number of assessors who selected each of the poles as the least similar to each of the samples. An example of this frequency table is shown in Figure 16.5. As can be seen in the data matrix, 23 assessors regarded Pole R1 as the most similar to Sample 1, while 45 assessors regarded Pole R2 the most different from Sample 1. This frequency table is analysed using CA to obtain a representation of the similarities and differences among samples in a low dimensional space. CA can be performed in R software using the function CA of `FactoMineR` package.

Sample	R1+	R1–	R2+	R2–	R3+	R3–
1	23	0	13	45	29	20
2	0	55	25	10	40	0
…	…	…	…	…	…	…
J	41	0	10	45	14	20

Figure 16.5 Example of the data matrix used for analysing data from triadic polarized sensory positioning data using correspondence analysis. Each column contains the number of assessors that regarded a pole as most similar to (+) and the most different from (–) each of the samples.

	Assessor 1		Assessor 2		...	Assessor 3	
Sample	+	−	+	−	...	+	−
1	R1	R3	R1	R2	...	R1	R3
2	R3	R2	R3	R2	...	R1	R2
...	...	...	...	...	...	...	...
J	R1	R2	R3	R2	...	R2	R1

Figure 16.6 Example of the data matrix used for analysing data from triadic polarized sensory positioning data using multiple factor analysis. Each pair of columns (+ and -) contains the pole (R1, R2 or R3) that one of the assessors regarded as most similar to (+) and most different from (-) each of the samples.

Alternatively, data from triadic PSP can be analysed using MFA. The data matrix needed to apply this approach is shown in Figure 16.6.

The stability of sample configurations can be evaluated using a bootstrapping resampling approach, as discussed in section 16.4.1 for PSP based on degree of difference scales.

16.5 Applications of Polarized Sensory Positioning

Polarized sensory positioning is mostly useful when information about the sensory characteristics of products is needed within short time frames or when a rough description is sought. The distinctive feature of PSP over holistic methodologies such as projective mapping or sorting is that it allows comparison of data collected in different sessions. For this reason, it is a good methodological choice for sensory characterization of sample sets that require multiple sessions to be evaluated due to their complexity or intense/persistent sensory characteristics.

In the case of new product development, due to its iterative nature, prototypes are obtained in different moments in time and comparison of their sensory characteristics is usually necessary. In these situations, holistic methodologies cannot be used because they rely on the comparative evaluation of the whole sample set. However, data collected in different sessions using PSP can be aggregated and compared, which makes this methodology appropriate for sensory characterization with untrained or semi-trained assessors during new product development.

Published applications of PSP are limited to a few products, which include mineral water (Teillet et al. 2010), cosmetic creams (Chrea et al. 2011), chocolate-flavoured milk (Antúnez et al. 2015), make-up foundations, vanilla milk desserts (Ares et al. 2015) and orange-flavoured powdered drinks (Ares et al. 2013; de Saldamando et al. 2013).

16.6 Advantages, Disadvantages and Limitations

The main advantage of PSP is that it allows data aggregation from different sessions due to the comparative nature of the task with fixed references and the sequential monadic presentation of the tested samples. This technique could potentially allow aggregation of data obtained from the same assessors in different sessions but also data obtained in different sessions with different assessors, which would be relevant when working with consumer-based sensory characterization.

Antúnez et al. (2015) evaluated data aggregation from different consumer groups. These authors reported that RV coefficients between sample configurations from the evaluation of the whole set and aggregated data from the evaluation of the split set by different consumer groups were significant and higher than 0.79. In addition, samples were sorted in similar groups considering their configuration in the first two dimensions of the PCA or CA when they were evaluated by a single consumer group and when two groups of consumers evaluated the split set. These results suggest that aggregation of data from the evaluation of split sample sets by different consumer groups could provide similar results to the evaluation of the whole sample set. However, the degree of difference among samples could affect the stability of data aggregations, suggesting that care must be taken when aggregating data from the evaluation of similar samples by different consumer groups. Further research in this respect is necessary.

In the usual applications of PSP, assessors are not asked to describe samples and therefore information about the sensory characteristics responsible for similarities and differences between samples and each of the poles is not gathered. However, assessors can be asked to describe the poles or each of the samples after completing the PSP task, as it is usually performed in projective mapping tasks (Perrin & Pagès 2009; Varela & Ares 2012).

The main limitation of PSP is the fact that stable reference products are needed. Samples selected as poles should not change their sensory characteristics and should be available whenever the methodology is applied. For this reason, it would be difficult to apply PSP for sensory characterization of fresh natural products due to the high variability in their sensory characteristics. For example, when working with fruits and vegetables, reference products are difficult to obtain and maintain due to product variability and limited shelf-life (Hampson et al. 2000). Also, variability from one product to another makes it difficult to select homogeneous poles for all assessors (Dever et al. 1995; Le Moigne et al. 2008).

16.7 Case Study

Eight yoghurts enriched with a prebiotic ingredient were formulated following a 2^3 full factorial design with the following factors: sugar concentration (4.0% versus 8.0%), type of prebiotic ingredient (native inulin, Frutafit IQ, Sensus,

Table 16.1 Formulation of the yoghurt samples considered in the study.

Sample	Prebiotic component	Commercial sugar (%)	Stabilizer (%)
1	Inulin	4.0	0
2	Inulin	4.0	0.075
3	Inulin	8.0	0
4	Inulin	8.0	0.075
5	Fructo-oligosaccharide	4.0	0
6	Fructo-oligosaccharide	4.0	0.075
7	Fructo-oligosaccharide	8.0	0
8	Fructo-oligosaccharide	8.0	0.075

The Netherlands, and fructo-oligosaccharide, Orafti® P95, Beneo GmbH, Mannheim, Germany) and stabilizer concentration (Dairy Blend YG LP, TIC Gums, White Marsh, Maryland, USA). The experimental design is shown in Table 16.1. The concentration of prebiotic ingredient in all formulations was 6.0%. All yoghurts contained 1% modified starch and 2% skim milk powder. The rest of the formulation consisted of skimmed pasteurized milk (0.1% fat).

Three of the eight samples were selected as poles (samples 1, 4 and 7) based on their formulation (see Table 16.1). These samples represented the main differences in the sensory characteristics of the samples. The poles differed in their sugar and stabilizer concentration and type of prebiotic component. Sample 1 was characterized by its low sweetness and high acidity and was formulated without stabilizer, while samples 4 and 7 had high sugar content and differed in their texture due to the different prebiotic component included in their formulation.

A consumer study with 50 participants was carried out. Each consumer received 30 g of each one of the three poles and approximately 20 g of the eight samples, coded with three-digit random numbers. Samples were presented in monadic sequence, following an experimental design balanced for order and carry-over effects (Williams' Latin square). Consumers were asked to try the three poles first to familiarize themselves with their sensory characteristics. Then, they had to try each of the samples and rate the overall difference between the samples and the three different poles using a 10 cm line scale ranging from 'exactly the same' to 'completely different'. Consumers were allowed to retaste the poles during the test, as many times as they deemed necessary. Still mineral water was used for rinsing between samples.

For each consumer, the distance from 'exactly the same' to the mark on the scale was measured for each sample and each of the poles. Data were analyzed using MFA, considering data from each consumer as a separate group of variables. Confidence ellipses were calculated using parametric bootstrapping (Dehlholm et al. 2012).

The first and second dimensions of the MFA explained 60.5% of the variance of the experimental data. Samples were sorted into two main groups according

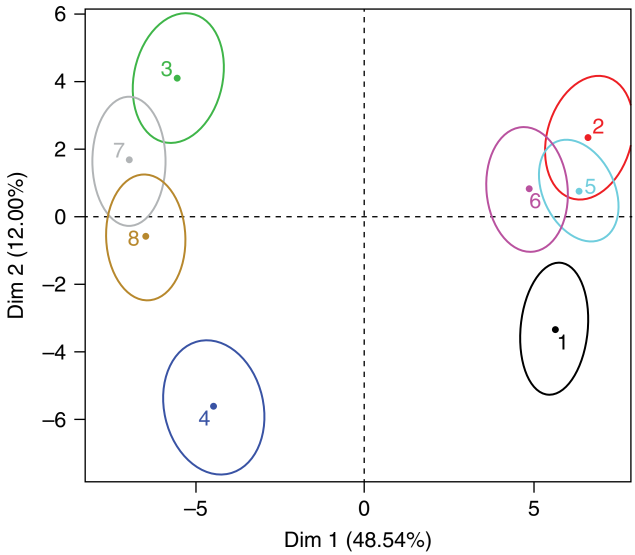


Figure 16.7 Representation of eight yoghurt samples in the first two dimensions of the multiple factor analysis performed on polarized sensory positioning data from a consumer study with 50 participants.

to their sucrose concentration. Samples 1, 2, 5 and 6, formulated with 4.0% sucrose, were located at positive values of the first dimension, as shown in Figure 16.7. Meanwhile, samples formulated with 8.0% sucrose (samples 3, 4, 7 and 8) were located at negative values of the first dimension. Sample 1 was sorted apart from the rest of the samples formulated with 4% sucrose. Similarly, sample 4, formulated with 0.075% stabilizer, was located at negative values of the second dimension, showing different sensory characteristics from the rest of the samples formulated with 8% sucrose.

The stability of the sample configuration was evaluated using a bootstrapping resampling approach, as suggested by Blancher et al. (2012). The bootstrapping process consisted of obtaining 1000 subsets of size equal to the total number of consumers using random sampling with replacement. As shown in Figure 16.8, average RV coefficient of sample configurations rapidly increased as the number of consumers in the virtual panel increased, while standard deviations decreased. If an RV coefficient of 0.95 is considered as stability threshold (Blancher et al. 2012), the configuration could be regarded as stable and only 15 consumers would be necessary to obtain stable results.

16.8 Extensions of Polarized Sensory Positioning

The philosophy behind PSP is being combined with other comparative rapid methodologies to allow comparison of results obtained in different sessions. Teillet et al. (2013) proposed the use of fixed references in the flash profile to

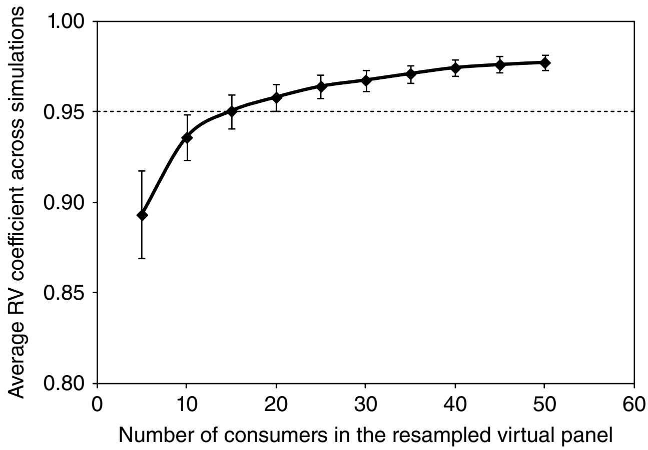


Figure 16.8 Average RV coefficient of sample configurations with respect to the reference configuration of samples as a function of the number of consumers considered in the resampled virtual panels for polarized sensory positioning data obtained in a consumer study in which 50 participants evaluated eight yoghurt samples. Vertical bars correspond to standard deviations.

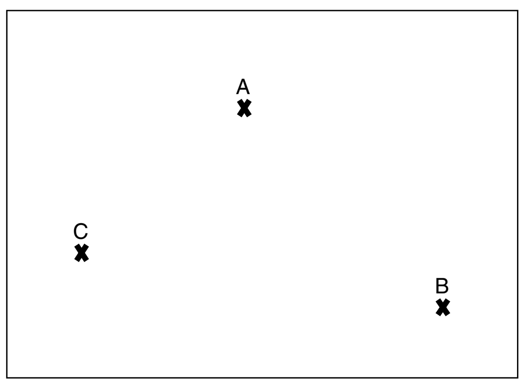


Figure 16.9 Example of the evaluation sheet used in a polarized projective mapping task.

allow comparison of results of different stages of a new product development process. Similarly, Chollet et al. (2014) suggested the use of a set of reference products in sorting tasks.

Ares et al. (2013) proposed the combination of PSP with projective mapping to enable the evaluation of samples in different sessions. This combination has been named polarized projective mapping (PPM). In this methodology, assessors are given a sheet of paper (usually 60 × 40 cm), on which three reference products (A, B and C) have been previously located. An example of the evaluation sheet is shown in Figure 16.9. They are asked to try the poles and to remember their main sensory characteristics. Then, they are asked to try samples and locate them on the sheet of paper considering the position of the poles and taking into account that samples which are located close to each other are perceived as

Sample 483	Less	More
Appearance		
Odour		
Texture		
Flavour		

Figure 16.10 Example of the evaluation sheet used in the pivot profile.

similar and those placed far from each other are perceived as different. After positioning the samples, assessors can also be asked to provide a description of each sample. The X and Y co-ordinates of the samples on the evaluation sheet of each of the assessors are analysed using MFA, considering descriptive data as supplementary variables, as suggested by Pagès (2005) for projective mapping.

Another extension of PSP is the pivot profile, which was proposed by Thuillier (2007) to characterize the sensory properties of champagne with wine experts. In this method, assessors are asked to provide a free description of the differences between samples and a reference product, which is called the 'pivot'. They should describe the appearance, odour, texture and flavour characteristics they perceive as less intense in the product compared to the pivot, as well as those that they perceive as more intense. An example of the evaluation sheet is shown in Figure 16.10. A nice application of pivot profile has been recently published by Lelièvre-Desmas et al. (2017).

16.9 Summary

Polarized sensory positioning is based on the evaluation of global differences between samples and a fixed set of reference samples. Its main advantage is the possibility of aggregating data obtained in different sessions, which is particularly relevant when using rapid methodologies in new product development, quality control or when studying complex products or samples with intense or persistent sensory characteristics. However, the number of published applications of the methodology is still limited and further research on its validity and reliability in different situations is necessary.

16.10 Future Developments

Polarized sensory positioning is a recent methodology for sensory characterization, which has been used in a limited number of applications. For this reason, further methodological research on PSP is necessary to develop recommendations for best practice. Although Ares et al. (2015) have provided insights on the cognitive strategies used by assessors to evaluate the degree of difference between samples, research is still necessary to establish clear recommendations on the number and characteristics of the poles.

Different statistical approaches have been proposed for the analysis of PSP data (e.g. PCA, MFA, GPA, STATIS). However, it is unclear if they differ in performance as no research comparing statistical approaches for the analysis of PSP data has been published. Similarly, the influence of the approach used for evaluating similarities and differences between samples and the poles (unstructured scales or triadic PSP) in the validity and reliability of the methodology requires further study.

Finally, the influence of sample complexity and degree of difference among samples on the applicability of PSP needs to be explored as most published applications have considered relatively simple products.

Acknowledgements

The authors are indebted to Comisión Sectorial de Investigación Científica (Universidad de la República, Uruguay) and Agencia Nacional de Investigación e Innovación (Uruguay) for financial support.

References

Abdi, H. (2010) Congruence: congruence coefficient, RV coefficient, and Mantel coefficient. In: Salkind, N.J., Dougherty, D.M. & Frey, B. (eds) *Encyclopedia of Research Design*. Thousand Oaks: Sage, pp. 222–229.

Antúnez, A., Salvador, A., de Saldamando, L., Varela, P., Giménez, A. & Ares, G. (2015) Evaluation of data aggregation in Polarized Sensory Positioning. *Journal of Sensory Studies*, **30**, 46–55.

Ares, G. & Varela, P. (2014) Comparison of novel methodologies for sensory characterization. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 365–389.

Ares, G., de Saldamando, L., Vidal, L., Antúnez, L., Giménez, A. & Varela, P. (2013) Polarized Projective Mapping: comparison with polarized sensory positioning approaches. *Food Quality & Preference*, **28**, 510–518.

Ares, G., Tárrega, A., Izquierdo, L. & Jaeger, S. (2014) Investigation of the number of consumers necessary to obtain stable sample and descriptor configurations from check-all-that-apply (CATA) questions. *Food Quality & Preference*, **31**, 135–141.

Ares, G., Antúnez, A., Oliveira, D. et al. (2015) Pole selection in Polarized Sensory Positioning: insights from the cognitive aspects behind the task. *Food Quality & Preference*, **46**, 48–57.

Blancher, G., Clavier, B., Egoroff, C., Duineveld, K. & Parcon, J. (2012) A method to investigate the stability of a sorting map. *Food Quality & Preference*, **23**, 36–43.

Brockhoff, P.B. (2003) Statistical testing of individual differences in sensory profiling. *Food Quality & Preference*, **14**, 425–434.

Cadena, R., Caimi, D., Jaunarena, I. et al. (2014) Comparison of rapid sensory characterization methodologies for the development of functional yogurts. *Food Research International*, **64**, 446–455.

Cadoret, M. & Husson, F. (2013) Construction and evaluation of confidence ellipses applied at sensory data. *Food Quality & Preference*, **28**, 106–115.

Chollet, S., Valentin, D. & Abdi, H. (2014) Free sorting task. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 207–228.

Chrea, C., Teillet, E. & Navarro, S. (2011) Application of the Polarized Sensory Positioning in the Cosmetic Area. Presented at the 9th Pangborn Sensory Science Symposium, 4–8 September, Elsevier, Toronto, Canada.

Costa, A.I.A. & Jongen, W.M.F. (2006) New insights into consumer-led food product development. *Trends in Food Science & Technology*, **17**, 457–465.

Dehlholm, C., Brockhoff, P.B. & Bredie, W.L.P. (2012) Confidence ellipses: a variation based on parametric bootstrapping applicable on multiple factor analysis results for rapid graphical evaluation. *Food Quality & Preference*, **26**, 278–280.

De Saldamando, L., Delgado, J., Herencia, P., Giménez, A. & Ares, G. (2013) Polarized sensory positioning: do conclusions depend on the poles? *Food Quality & Preference*, **29**, 25–32.

Dever, M.C., Cliff, M.A. & Hall, J.W. (1995) Analysis of variation and multivariate relationships among analytical and sensory characteristics in whole apple evaluation. *Journal of the Science of Food & Agriculture*, **69**, 329–338.

Faye, P., Brémaud, D., Teillet, E., Courcoux, P., Giboreau, A. & Nicod, H. (2006) An alternative to external preference mapping based on consumer perceptive mapping. *Food Quality & Preference*, **17**, 604–614.

Green, P.E., Carmone, F.J. & Smith, S. (1989) *Multidimensional Scaling: Concepts and Applications*. Boston: Allyn & Bacon.

Hampson, C.R., Quamme, H.A., Hall, J.W., MacDonald, R.A., King, M.C. & Cliff, M.A. (2000) Sensory evaluation as a selection tool in apple breeding. *Euphytica*, **111**, 79–90.

Heymann, H., King, E.S. & Hopfer, H. (2014) Classical descriptive analysis. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 9–40.

Jaeger, S.R., Wakeling, I.N. & MacFie, H.J.H. (2000) Behavioural extensions to preference mapping: the role of synthesis. *Food Quality & Preference*, **11**, 349–359.

Lawless, H.T. & Heymann, H. (2010) *Sensory Evaluation of Food. Principles and Practices*. New York: Springer.

Lê, S., Josse, J. & Husson, F. (2008) FactoMineR: an R package for multivariate analysis. *Journal of Statistical Software*, **25**, 1–18.

Lelièvre-Desmas, M., Valentin, D. & Chollet, S. (2017) Pivot profile method: what is the influence of the pivot and product space? *Food Quality & Preference*, **61**, 6–14.

Le Moigne, M., Symoneaux, R. & Jourjon, F. (2008) How to follow grape maturity for wine professionals with a seasonal judge training? *Food Quality & Preference*, **19**, 672–681.

Mair, P., de Leeuw, J., Borg, I. & Groenen, P.J.F. (2013) Package 'smacof'. Available at: http://cran.r-project.org/web/packages/smacof/smacof.pdf (accessed 16 June 2017).

Murray, J.M., Delahunty, C.M. & Baxter, I.A. (2001) Descriptive sensory analysis: past, present and future. *Food Research International*, **34**, 461–471.

Pagès, J. (2005) Collection and analysis of perceived product inter-distances using multiple factor analysis: application to the study of 10 white wines from the Loire Valley. *Food Quality & Preference*, **16**, 642–649.

Perrin, L. & Pagès, J. (2009) Construction of a product space from the ultra-flash profiling method: application to 10 red wines from the Loire Valley. *Journal of Sensory Studies*, **24**, 373–395.

R Core Team (2013) *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Robert, P. & Escoufier, Y. (1976) A unifying tool for linear multivariate statistical methods: the RV coefficient. *Applied Statistics*, **25**, 257–265.

Teillet, E. (2014) Polarized sensory positioning methodologies. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 255–270.

Teillet, E., Schlich, P., Urbano, C., Cordelle, S. & Guichard, E. (2010) Sensory methodologies and the taste of water. *Food Quality & Preference*, **21**, 967–976.

Teillet, E., Petit, C. & Delarue, J. (2013) Combining PSP and flash profile… Why? How does it work? Presented at 10th Pangborn Sensory Science Symposium, 11–15 August, Elsevier, Rio de Janeiro, Brazil.

Thuillier, B. (2007) Rôle du CO2 dans l'appréciation organoleptique des champagnes – expérimentation et apports méthodologiques. Thesis, URCA, Reims, France. Available at: www.abt-sensory-analysis.com/uploads/pdf/AR0._hesis_B._THUILLIER_2007.pdf (accessed 21 June 2017).

Valentin, D., Chollet, S., Lelièvre, M. & Abdi, H. (2012) Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, **47**, 1563–1578.

Varela, P. & Ares, G. (2012) Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization. *Food Research International*, **48**, 893–908.

Vidal, L., Cadena, R.S., Antúnez, L., Giménez, A., Varela, P. & Ares, G. (2014) Stability of sample configurations from projective mapping: How many consumers are necessary? *Food Quality & Preference*, **34**, 79–87.

CHAPTER 17

Check-All-That-Apply and Free Choice Description

Dominic Buck and Sarah E. Kemp

17.1 Introduction

Verbally based qualitative methods have been used for many years in market research, and extensively in sensory and consumer product research. They have recently received attention in sensory science as simple, rapid techniques for delivering a sensory description of a product as stand-alone methods, integral parts of some methods and to supplement or complement other methods (Bécue-Bertaut 2014).

These techniques are particularly well suited for use with consumers, as they do not involve rating scales. Although rating scales purport to provide quantitative measurement of the human perception of product attributes, they suffer from many drawbacks. Using words alone, and asking about the presence or absence of attributes, are much simpler approaches for individuals to comprehend and undertake and ultimately may provide better and more relevant description. Verbally based qualitative methods can also be used with trained panels, to provide sensory characterization of a product more quickly than a full descriptive analysis, although in less detail.

Over the last decade, non-scaled descriptive approaches have enjoyed an upsurge of interest among the scientific community and practitioners alike for use in providing sensory-based product descriptions and to help to interpret why consumers like a product. As a result, sophisticated analysis options have become available. However, even using basic analysis tools, useful and insightful information can still be gleaned.

This chapter sets out the rationale for using verbally based qualitative descriptive methods and gives an overview of two prominent examples: 'check-all-that-apply' (CATA) and open-ended questioning, a type of free choice description (FCD). It provides methodological guidelines, analysis possibilities, practical considerations, advantages and disadvantages, case studies showing the approaches in action and future directions.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

17.2 Rationale for Use

Rating is an obvious way of measuring the extent to which individuals perceive aspects of products. In consumer studies, rating scales are used on hedonic and diagnostic variables. Hedonic rating scales are often presented as five-, seven- or nine-point scales, sometimes completely labelled or sometimes just with the end points labelled. Diagnostic scales are often presented as five-point 'just about right' (JAR) scales, where the midpoint is 'just about right' and either side, the scale runs two points towards 'not at all enough' on one side and two points to 'much too much' on the other.

In sensory 'trained panel' studies using descriptive analysis, panellists often use a scoresheet, with each variable's assessment measured and recorded on a continuous line scale, by drawing a mark on the scale at the appropriate point. The line scale can be labelled with endpoints, sometimes inset from the end of the line scale in an attempt to minimize end-of-scale bias. Additionally, marks can be shown for benchmark or reference products' intensities.

An example of a continuous line scale for 'aroma strength' with the assessor's score marked towards the stronger end of the scale is shown in Figure 17.1.

For analysis, the lines are effectively divided into 100 or more points and a numeric score is derived based on the position of the mark made by the panellist. The results can therefore be treated as ordinal in nature and they give an impression of precision, because the results are numeric. As with other forms of rating scale, the mean and a measure of variation can be estimated. They can be analysed statistically, allowing hypothesis testing and inference of confidence intervals, or the extent to which they correlate with other descriptive variables or overall hedonic measures, such as overall liking.

Beneath this assumption of precision, however, lie a number of issues with rating scale data that need to be considered, as potentially they could reduce the value of the information collected and distilled from the numbers.

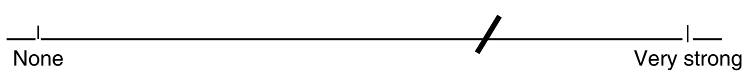


Figure 17.1 Example of a continuous line scale for 'aroma strength' with the assessors score marked towards the stronger end of the scale.

17.2.1 Fast versus Slow Thinking

In *Thinking: Fast & Slow*, Kahneman (2011) suggests two systems of thinking, which have their origins in the evolution of the human brain.

- System I: fast thinking – rapid response, emotional, subconscious
- System II: slow thinking – rational, conscious, considered, logical

By its nature, consumer or sensory research conducted using any form of questionnaire to obtain information from human subjects, and rating scales in particular, require considered, rational thought, that is, System II slow thinking. The problem here is that the more assessors rationalize their subconscious response, the less the information that they provide reliably reflects their true response, due to a number of biases, including:

- halo effects
- social norms
- desire to be 'correct'
- providing answers that are a logical fit with other ratings already provided
- desire to fulfil the task requirements.

Responding to a rating scale requires a high degree of cognitive effort. Easier tasks, such as CATA and other verbal qualitative methods, are said to better reflect consumers' true response.

17.2.2 'Noise' versus 'Signal'

Usual practice for sensory analysis dictates that individuals complete every question in a questionnaire or scoresheet unless specifically directed to avoid them. This means that an individual will provide an evaluation for an attribute regardless of:

- the relevance or importance of the attribute to the assessor
- their understanding of the exact nature of the characteristic
- their sensory acuity to the characteristic
- their fatigue in answering the questions.

In effect, this means a proportion of assessors, sometimes a sizeable proportion, are providing answers that are only contributing noise rather than signal to the results. In turn, this increases variation and reduces discrimination if product-to-product comparisons are being made (see case study 1). Sensory panellists are trained to understand attributes and rate products in a reproducible and reliable fashion, which enhances the signal, whereas consumers receive no such training and may perform better using a simpler task.

In summary, many methods use rating scales to gather information from consumers, but verbal methods have the advantage of being simpler to perform, thereby potentially providing more meaningful data, and are quicker, thus saving time and resources. CATA provides frequency data on the number of assessors selecting sensory terms and therefore gives more information about sensory characteristics than hedonic rating, although CATA does not provide an assessment of perceived intensity (see section 17.6.2), and so provides less information than direct intensity rating (assuming reliable and reproducible scale use).

17.3 CATA Methodology

17.3.1 Overview

The application of CATA, sometimes also called 'tick-all-that-apply' (TATA), for detailed sensory description was sparked by Adams et al. (2007). It involves assessors selecting all terms that apply to a product from a list of descriptors or associations. For purely sensory characterizations of products, the terms are sensory attributes or descriptors. CATA questions provide multivariate binary data that can be analysed using Cochran's Q test (greater than two products) and correspondence analysis (CA). CATA has been shown to be reliable and stable for sensory product characterization by consumers (Jaeger et al. 2013a) and to produce results that are similar to traditional descriptive analysis with a trained panel (Ares et al. 2010a, 2015a; Bruzzone et al. 2012; Dooley et al. 2010). Several methodological variations have been proposed (see section 17.3.8). Reviews are given by Meyners and Castura (2014) and Ares and Jaeger (2014).

In general, the performance of a CATA test is similar to the general methodology employed in other product tests, although questionnaire design is a crucial task. The method consists of several stages which are described below.

17.3.2 Descriptor Elicitation

There are several ways in which a list of sensory descriptors can be generated for use in CATA questions, which may be used alone or in combination:

- *FCD using open-ended questions*: samples are evaluated one at a time and a personal vocabulary of descriptors is built as new characteristics are perceived. Generally, this means the first sample will generate many descriptors. Subsequent samples are likely to generate fewer and fewer additional attributes as more samples are evaluated. This method may be used to generate personal or group descriptor lists (see section 17.4.3 for an overview of analysis). The group list may be generated by different consumers to those undertaking the CATA test.
- *Method of triadic elicitation*: in this methodology, samples are presented in triads. Each sample in the triad is assessed against the other two, and assessors are asked to list all attributes that are the same and those that are different between the sample and the remaining pair. This method is typically used to generate personalized lists as an integral part of other methods, such as FCD and flash profiling, although group lists can also be generated. Again, the group list may be generated by different consumers to those undertaking the CATA test.
- *Focus groups*: these can be used to generate terms during a structured discussion lead by a moderator and may be run with different consumers to those undertaking the CATA test.
- *Trained panel*: typically, a trained panel will generate a list of well-defined sensory attributes with physical references, and will have a shared understanding of these attributes. Although the list of terms may be detailed and well defined, the terms may be less easy for consumers to understand and the list may need to be simplified.
- *R&D experts*: R&D experts may have valuable experience from which to suggest terms, although care needs to be taken to ensure the terms are not too technical and are representative of consumer product perceptions.
- *Desk research*: terms may be identified from published literature, previous studies and marketing communication.

17.3.3 Questionnaire Design

17.3.3.1 Type of Terms

The type of terms selected will depend on the objectives of the study. If a sensory description of a product is required, then only sensory terms should be included. Terms used with consumers must be easily understood and similar to those they might use spontaneously. Terms with similar meanings should be avoided.

Jaeger et al. (2013a) found some differences in the reliability of attributes and there was some evidence that attribute reliability may be moderated by the form of the CATA task. Complex attributes, such as 'creamy', and intensity-graded attributes, such as 'sweet: moderate', 'chocolate: strong', showed the largest differences between sessions and the lowest attribute stability indices. Intensity-graded attributes were less reliably used by consumers. The research developed indices that could be used to refine lists of terms for repeated use based on frequency and reproducibility of attribute use (see section 17.3.7).

If the relationship between sensory characteristics and other product characteristics/associations, such as hedonic, emotional, fit to concept, functional, etc., is to be investigated with consumers, then non-sensory attributes might be included. It is interesting to note that Popper et al. (2011) found that the sensory terms in CATA were more important than other types of terms, such as emotional terms, when describing unbranded products, whereas the reverse was true when describing branded products.

17.3.3.2 Number of Attributes

Short lists might encourage consumers to use all terms and thereby decrease discriminatory ability. Long lists might encourage consumers to focus on 'satisfying' the task by choosing from the beginning of the list without carefully considering the products' sensory characteristics (Ares & Jaeger 2014).

Jaeger et al. (2015a) found that sensory product characterization with 'short' (10–17 terms) versus 'long' (20–28 terms) lists of CATA terms gave largely similar results, including sample configuration and task perception (ease and tediousness). There were some differences in term configuration and the conclusions about sample differences found, depending on sample length. Lower frequency of term use was found for long lists, and there was some evidence that CATA questions with 'long' lists of synonyms (e.g. hard and firm) or antonyms (e.g. hard and not hard) may cause results to be split, which may be associated with reduced discriminatory ability of CATA questions.

Ares et al. (2015b) studied attention in a CATA task using eye tracking and found less attention was given to individual terms in a long list (20 terms) but there was greater sustained attention to the task than with a short list (12 terms). They concluded that list length has little impact on consumers' visual attention.

17.3.3.3 Order of Attributes

The final list of sensory attributes is typically presented on the questionnaire as a 'grid', usually consisting of several columns of terms, with an equal number of terms in each column as far as possible. Figure 17.2 gives an example of a sensory-based CATA grid and instructions.

Traditionally, CATA grids were presented to all assessors with the CATA attributes in the same configuration. Attributes were randomly ordered once, and then that configuration was presented to everyone. Research has consistently shown order biases in citing attributes due to their position in the grid (Ares & Jaeger 2013), with terms at the beginning more frequently selected than those at the end (Ares & Jaeger 2014). Visual eye tracking has shown that consumers (at least from Western cultures) process terms from left to right and top to bottom (Ares et al. 2014a). In general, they do consider the whole list and select the terms that apply, but as the task progresses they make fewer and shorter eye fixations to complete the task. The influence of term order has been found to be dependent on the characteristics of the study (number of samples, number and type of CATA terms) and although conclusions regarding sample differences have been found to vary with term order, the influence of term order was not deemed large enough to invalidate results (Ares & Jaeger 2013; Lee et al. 2013).

Computer-aided personal interviewing (CAPI) technology has made it much easier for individual assessors to receive different configurations of terms in the

Here is a list of words which other people have used to describe **red** wine.
Please indicate which you think apply to the wine you have just tried by placing a tick (✓) in the box next to it. **There are no right or wrong answers. Please tick as many or as few as you wish.**

Light	☐	Easy to drink	☐
Complex	☐	Full bodied	☐
Fruity	☐	Sharp	☐
Smooth	☐	Woody	☐
Fresh	☐	Full of flavour	☐
Rich	☐	Aromatic	☐
Juicy	☐	Robust	☐
Strong	☐	Warming	☐
Oaky	☐	Bitter	☐
Spicy	☐	Heavy	☐

Figure 17.2 Example of a CATA grid and instructions.

grid, which reduces positional bias. Studies using a randomized presentation order for terms within and across consumers have been shown to be reproducible (Ares & Jaeger 2013; Ares et al. 2014b; Lee et al. 2013). Presenting terms in different orders for each sample significantly increases total number and duration of eye fixations (Ares et al. 2014a) and length of time taken to complete the task (Lee et al. 2013). No major differences were found in the sensory characterization comparing balanced presentation order between assessors and both between and within assessors (Ares et al. 2015c), although within-assessor balanced presentation was not recommended for use with more than 50 attributes. However, it may be desirable for each assessor to receive the same configuration across sample evaluations, so that sample-to-sample differences are not biased by positional effects.

Traditionally, all modalities were randomly ordered, so that a texture attribute might be next to a flavour attribute, next to an appearance attribute. Again, research has shown that blocking the attributes by modality allows assessors to find relevant attributes more easily and increases selection frequency (Ares et al. 2013; Buck et al. 2013).

17.3.3.4 Method of Response

Different ways of making a response to a CATA question include simply checking (ticking or marking) attributes that apply; checking explicitly whether an attribute applies or does not apply; and providing an importance or relevance rating for each and every attribute. Research has been conducted into the 'depth' of response to CATA questions (Buck et al. 2013). Clearly, the deeper the response then the greater involvement of System II thinking involved in making that response, so many practitioners believe the simpler response, that is, checking only those attributes that do apply, is the most effective approach.

17.3.3.5 Inclusion of Other Questions

When CATA studies are used with consumers for sensory product description, other types of questions may be included, depending on the objective of the study.

- Hedonic questions in order to investigate consumer liking and sensory drivers of liking. However, using sensory questions introduces the potential for bias of hedonic response. Jaeger et al. (2013b) and Jaeger and Ares (2014) found weak evidence of such bias across a range of products and therefore suggested the suitability of CATA questions for use with hedonic questions. Typically, the hedonic question is asked first to minimize bias, ideally across all products, with the CATA question asked subsequently across all products. In practice, often the hedonic question is asked first and the CATA question asked second for each product in turn, so that only one presentation of each sample is required.
- Other types of terms and associations, such as those related to concept, benefit, emotion, brand, functional, usage, etc.

- JAR scales to increase diagnostics for product guidance. Jaeger et al. (2015b) have also investigated the addition of JAR scales to studies using CATA with hedonic questions and found no bias of hedonic response and no negative effect on consumers' perception of the task.

17.3.4 Samples

Samples for a CATA study to produce a sensory description of a product are presented as for a typical sensory or consumer study. A standard amount of sample is presented. All samples are presented blind in identical, non-biasing containers labelled with random three-digit codes. Samples are presented sequentially and monadically using an appropriate order, such as random, Latin square, complete block, etc. and an experimental design may be used. Lee et al. (2013) found that sample presentation position had an impact on the number of descriptors selected, which was attributed to familiarity with attributes and serving order effects, such as 'first sample' effect. The latter can be reduced by using an initial dummy or warm-up sample. A dummy sample is covertly given to assessors as part of the sample set, and the resulting data are excluded from the data analysis. A warm-up sample(s) is given to assessors to help familiarize them with the test protocol and/or product.

17.3.5 Assessors

Trained assessors from a descriptive panel may be used, who will have been pre-selected for their good sensory abilities and trained to produce reliable and reproducible results. They should have a good, shared understanding of the descriptors. Studies using CATA with a trained descriptive panel include McCloskey et al. (1996), Le Fur et al. (2003) and Campo et al. (2008, 2010).

If consumers are used, a larger number are needed. Ares et al. (2014c) suggest that 60–80 consumers is reasonable to obtain stable sample and descriptor configurations, although research is needed on how the degree of sample differences affects the minimum number of consumers. If CATA is to be included with other questions, then the number of consumers necessary to obtain reliable results with those questions should be taken into consideration. Consumers can be users or non-users depending on study objectives.

The assessors in a CATA study may be the same or different from those who generated the list of CATA descriptors.

17.3.6 Data Collection

Assessors are asked to assess each sample, which may be according to precise evaluation instructions, and choose the descriptive terms that apply using the appropriate questionnaire designed as described above. They are then instructed to wait an appropriate amount of time and/or use a palate cleanser prior to assessing the next sample.

Dynamics of sensory perception has been confirmed as an explanation for bias (Ares et al. 2013). Asking respondents to make the assessment whilst eating versus after eating gave better discrimination.

17.3.7 Statistical Analysis

CATA generates binary data and an individual response matrix of descriptors chosen for products will take the form of a table of zeros and ones. Frequency matrices of the number of assessors selecting each attribute to describe each product are generated. Correspondence Analysis (CA) can be applied to the frequency matrix to obtain a sensory map of the products. Popper et al. (2011) have suggested using a variant of CA called multi-block Hellinger analysis. Multi-block analyses, such as multiple factor analysis (MFA) or multiple correspondence analysis (MCA), have also been used (Ares et al. 2011a,b).

Significance can be tested for each attribute for two products with a simple t test of proportions or McNemar's test to assess product differences. For more than two products, Cochran's Q test provides an overall statistical significance measure of product differences. Meyners and Castura (2014) indicate that at least 15 assessors should be used when comparing two or three samples, which may have implications for the analysis of CATA results generated using trained panels, where fewer assessors may be used.

Methods for evaluating reproducibility and repeatability of CATA assessors have been proposed (Jaeger et al. 2013a; Meyners et al. 2016a; Worch & Piqueras-Fiszman 2015). Jaeger et al. (2013a) proposed indices for assessing reproducibility and repeatability: a global reproducibility index to assess within-assessor reproducibility (based on the work of Campo et al. 2008), an attribute stability index (SI) to determine the reliability of each attribute of the CATA questionnaire across samples and sessions, and an attribute selection stability index (SSI) calculated as the percentage of consumers who selected a term in both sessions for describing samples with respect to the average number of consumers who used the term in at least one of the sessions. It was found that CATA was stable over sessions and consumers performed reliably when producing sensory product characterizations. These types of measures could contribute to the selection of terms to be included in CATA questions by excluding terms which are not used in a stable way by consumers across sessions.

Bar charts are commonly used for visual presentation of results for a few products. With many products, bar charts become difficult to read and it makes sense to use a perceptual mapping approach such as CA. CA maps have added advantages in allowing the relationships between attributes, between products, and between attributes and products to be visualized.

CATA data have been used in combination with hedonic data in more powerful analyses, such as preference mapping (Ares et al. 2011a; Dooley et al. 2010) and penalty/reward analysis (Ares et al. 2014d; Meyners 2016a; Plaehn 2012).

For a detailed description of analysis approaches for CATA data, readers are referred to an excellent paper by Meyners et al. (2013): 'Existing and new approaches for the analysis of CATA data', which describes the various techniques available. However, some of the basic analyses are demonstrated in case study 3.

17.3.8 Variants of CATA Methodology

17.3.8.1 Pick K or Pick K from N

Assessors choose a set number of attributes (K) from the overall list that best describe the product. When the number of attributes to be chosen is much smaller than the list of attributes, this method can quickly identify the main/dominant characteristics of the product, whereas CATA provides a more complete description (see Valentin et al. 2012 for an overview).

17.3.8.2 Forced Choice CATA/Applicability Testing

Assessors are required to answer a yes/no question for every attribute in the list. This method has been proposed to overcome the potential issue of assessors choosing attributes merely to fulfil the task requirements (known as satisficing). Ennis and Ennis (2013) presented a Thurstonian approach for forced choice CATA questions asking whether an attribute 'applies' or 'does not apply'. Jaeger et al. (2014) compared forced choice yes/no questioning, in which assessors were asked to identify if each term on the list was appropriate for describing each sample by answering 'yes' or 'no', and standard CATA questioning. For both methods, frequency of attribute use was determined by counting the number of consumers selecting each attribute. Data analysis was carried out to compare frequency of attribute use (using Fisher's exact test and linear regression), differences among samples (using Cochran's Q test) and similarity and stability of sample and attribute configuration (using CA and RV coefficient). The study found parity overall. Forced choice yes/no questioning was associated with higher term citation frequency, although this did not consistently translate into greater product discrimination.

17.3.8.3 Rate-All-That-Apply (RATA)

In RATA methodology (Ares et al. 2014e), assessors rate the terms they ticked as 'apply'. Reinbach et al. (2014) found that CATA with rating yielded fewer significant descriptors, possibly due to consumers' inconsistent scale use, which might outweigh any advantages of using scales. Guiacalone and Hedelund (2016) found indications that the reproducibility of RATA with semi-trained assessors might be similar to that of a simple checklist, supporting the validity of RATA as a sensory profiling tool. Meyners et al. (2016b) give details of analysis for RATA studies.

17.3.8.4 Temporal Check-All-That-Apply (TCATA)

TCATA (Castura et al. 2016) allows the continuous selection and deselection of multiple applicable attributes simultaneously over time. It builds upon temporal dominance of sensation (TDS) and uses an approach similar to time-quality tracking (Zwillinger & Halpern 1991), an earlier method that also captured a sequence of attribute qualities without intensity scaling. Assessors work with a reduced grid of characteristics and over a controlled, specified time period. They indicate and continually update attributes that apply, thereby tracking sensations in the product as they change over time. Assessors are permitted to check and uncheck attributes whenever necessary. Multiple attributes can be selected simultaneously, which permits descriptions of sensations occurring sequentially or concurrently.

TCATA fading is a further development of this approach, in which selected terms gradually and automatically become unselected over a predefined period of time (Ares et al. 2016). The rationale for this is to avoid the situation where subjects forget to deselect an attribute that no longer applies. Results suggested that automatic deselection can improve discrimination over a standard TCATA task and provide a more accurate dynamic sensory product description. Both TCATA and TCATA fading have potential, but further research is needed to refine the methodology.

17.3.9 Comparison with Other Descriptive Methods

The use of CATA for sensory descriptive analysis has been extensively compared and contrasted to other methods. For reviews, see Valentin et al. (2012), Varela and Ares (2012) and Ares and Varela (2014). Common findings appear to be that CATA produces similar results and conclusions to other methods, but is easier and less time-consuming to perform. It provides more detailed and easily interpretable descriptions than similarity-based methods, such as sorting and mapping techniques, but less detail than methods involving rating on scales, such as traditional descriptive analysis techniques and flash profiling.

CATA has been found to give similar results to descriptive analysis (Ares et al. 2010a, 2015a; Bruzzone et al. 2012; Dooley et al. 2010). However, Ares and Jaeger (2014) state that sensory product description using CATA cannot be regarded as a replacement for traditional sensory descriptive analysis, as the latter will always be more accurate due to extensive training in identification and rating of precisely defined attributes.

Fleming et al. (2015) compared CATA with sorting and polarized sensory positioning (PSP) in a study on astringent stimuli. They concluded that the two-dimensional maps were significantly similar. CATA was able to provide additional evidence on differences between samples being based on side tastes rather than astringent subqualities. CATA was found to be more efficient than sorting, but had similar efficiency to PSP.

17.4 Open-Ended Questioning

17.4.1 Overview

Open-ended questioning, also known as 'free comment', is a type of FCD that gives assessors total freedom to list any descriptors (and in some cases associations) that they think apply to or describe the product they are evaluating. Assessors are asked for an opinion or comment and allowed to answer spontaneously and freely. This is a simple and easy task to perform, which is particularly useful to elicit data from consumers. After initial textual data analysis, which can be complex and lengthy, analysis of frequency data may be carried out using chi-square, chi-square per cell, CA and multifactor analysis. Free comments are often collected as supplementary information to other methods, such as hedonics, sorting and Napping® techniques. Open-ended questioning with subsequent comment analysis has been used to obtain product descriptions in consumer vocabulary (Ares et al. 2010b). See Symoneaux and Galmarini (2014) and Piqueras-Fiszman (2014) for reviews of methodology and analysis.

17.4.2 Methodology

There are several methods that can be used in open-ended questioning to elicit sensory descriptions of products.

i) Voluntarily write comments to describe a product or a set of products, often after another type of assessment, such as hedonic evaluation, preference evaluation or sorting or grouping samples.
   - Use a set number of words (e.g. three or four) to describe a product after evaluation, often after another type of assessment, as above. This enables a more focused response and less complicated analysis.
   - Freely comment on how a product differs from another product.
   - Freely state which sensory attributes of a product are stronger and weaker than another product, such as in pivot profile.
   - Freely state what is liked and not liked about a product, which is used with consumers.

If a personal lexicon of descriptors has been elicited, it is usually sensible to ask assessors to review their descriptors and remove any redundant terms or add terms if they feel they have missed something important. It can then be useful to ask them to conduct additional evaluations, such as projective mapping techniques e.g. Napping, or hierarchical or other sorting tasks, and align the descriptions with the samples thus evaluated. Alternatively, assessors could be asked to reassess the samples using a 'not at all' to 'extreme' rating for each of the terms in their personal lexicon. Personal lexicons are more difficult and cumbersome to interpret and combine across assessors and the verification of terms can be lengthy (Seo et al. 2009).

Trained panellists may be asked open-ended questions to describe products using sensory attributes on which they have already been trained.

Open-ended questions may be asked with other questions, such as liking and preference, and are often positioned last, as there is the potential to bias earlier hedonic questions, which require a more holistic response.

Methodological issues related to samples, assessors and data collection are similar to those described above in section 17.3. Typically, 50–100 consumers are used.

17.4.3 Statistical Analysis

Although open-ended questioning provides a rich and detailed product description, it is difficult and time-consuming to analyse. Varela and Ares (2012) give an overview of the complex process of textual analysis on group data.

1. Transforming open comments into a list of accurate descriptors.
   - Removing typing, orthographic and grammatical mistakes.
   - Elimination of connectors, auxiliary terms and irrelevant words.
   - Identification of phrases and terms.
   - Grouping synonyms, as determined by a dictionary and/or personal interpretation of up to three researchers.
   - Managing ambiguous words.
   - Marking terms of interest for the researcher.
2. Construction of a frequency table.
   - Selecting important categories, for example those mentioned by more than 5% or 10% of the consumers.
   - Determining frequency of mention of each category by counting the number of assessors that used each category to describe each product.
3. Analysis.
   - Global chi-square test to determine significant differences in the product descriptions.
   - Chi-square per cell test to identify significant differences between samples for each of the sensory characteristics used by consumers in product descriptions.
   - CA to produce a two-dimensional map of samples and attributes.

If the objectives of the study require individual lexicons to be developed from open-ended questioning, then a simplified version of stage 1 is used.

In general, statistical analysis is becoming more automated and specialized software exists for the analysis of textual data. Deneulin and Bavaud (2015) proposed novel, automated statistical analyses of textual data from an open-ended survey on a 'minerality' perception by wine tasting experts without tasting, which the authors propose could be applied to questionnaires involving free comments on products.

Data from open-ended questioning can be difficult to interpret. Although consensus may appear to have been reached, consumers may be using the same descriptor for different sensations. The meaning of terms used may be difficult to understand, at least in a way that is actionable, and descriptions may also be idiosyncratic, vague and imprecise. Further research may be needed to interpret and validate them.

17.4.4 Comparison with Other Descriptive Methods

Sample maps obtained from open-ended questioning used as part of a larger study have been found to be similar to those obtained from traditional descriptive analysis with a trained panel (Ares et al. 2010b; Symoneaux et al. 2012; ten Kleij & Musters 2003).

Another direct comparison of free comment description and a classic sensory descriptive profiling technique with wine professionals without prior common training was carried out by Lawrence et al. (2013). The two methods gave similar main odour characteristics, but the free comments method generated more descriptors, was less time-consuming and easier to perform.

Fonseca et al. (2016) compared free comment analysis with pivot profile and found that although free comment analysis showed good results and correlation with pivot profile, its results were more variable and performance was lower as it identified a lower number of main attributes.

17.5 Practical Considerations

The usefulness and efficiency of CATA grids are strongly affected by the choice of attributes to be included. Obviously, an exhaustive list of characteristics would be desirable but as the list grows, assessors are less likely to cite characteristics, particularly those at the bottom of columns and those not in the first column. The authors recommend no more than 25 characteristics to form the grid and preferably fewer than that number. If necessary, create different grids for modalities, such as flavour, texture, etc., to ensure that subjects give all the characteristics sufficient attention. Of course, this raises the question of how to choose the attributes to be placed on a restricted list. Often, this will depend on the importance of the attribute to the product or project.

If characteristics are complex and of particular interest, it may be necessary to include both 'poles' of intensity in the CATA grid. Consider a texture term such as 'hard' for example. Because CATA generates counts of citations of characteristics, the *lack* of citation for 'hard' does not necessarily mean the product is therefore 'soft' – it could be spongy or rubbery, for example. Including 'not hard' may aid understanding. FCD can be a useful precursor to creating the CATA grid in order to understand the range of characteristics in play.

The results of CATA, namely product characterization and differences between samples, are subject to minor biases and it is important to report the exact nature of the questionnaire and experimental conditions under which CATA data are collected, so that future experiments can be compared (Ares et al. 2013).

17.6 Advantages and Disadvantages

17.6.1 Advantages

The major and compelling advantage of CATA and FCD is the intuitive simplicity for assessors. This applies particularly to naive assessors, that is, untrained consumers, but also to some extent to trained sensory panellists.

Data from both methods can be analysed simply, unless textual data analysis is required, although more sophisticated methods of analysis are also available if justified and needed. In particular, CATA, with its structured grid constructed prior to data collection, provides a basis for more powerful analysis techniques of the CATA results alone, or their relationship with hedonic measurement, such as input to penalty analysis (Ares et al. 2014d; Plaehn 2012).

For CATA, the provision of a list of words is helpful to those who find it difficult to verbalize. Assessors are also able to ignore non-relevant descriptors, unlike with scaling methods. CATA is powerful enough to discriminate between samples, whilst remaining quick and simple.

The main advantage for FCD is its total lack of constraint, allowing for individual assessors' perceptions and language to be fully utilized. It has been able to identify important attributes not previously thought to be of interest to consumers and omitted by researchers in other tests (Varela et al. 2014).

Both methods offer an advantage over holistic methods, such as sorting and mapping, in that a larger number of samples can be assessed. For holistic methods, samples need to be assessed together and there is a maximum that can be used without causing fatigue.

17.6.2 Disadvantages

CATA can be biased or produce less valid data due to the choice of terms selected for the CATA grid. Most seriously, this might mean important attributes are omitted. However, there could also be problems if more than one term is included covering the same underlying attribute, which can lead to a 'split vote' in terms of citations. In the consumer context, it is preferable that the CATA grid is informed by sensory evaluation preceding the consumer study, although this is not always possible. In a sensory context, CATA can be used at an early stage to help develop a scoresheet or select products or terms for deeper evaluation. In that sense, the grid can evolve to refine which attributes are covered by the grid. In any event, great care must be taken in the CATA grid's construction to ensure the major attributes are all included.

FCD obviously avoids this potential constraint of CATA. However, the same lack of constraint means that very disparate, idiosyncratic language can be used to express the same sensory perception, sometimes including non-sensory terms. The difficulty with this is in the complex and lengthy analysis and interpretation to determine whether consumers are perceiving the same or different sensory attributes. The meaning of terms may be obscure and descriptions may be vague and imprecise, so that additional research may be needed to understand consumers' responses and enable technical teams to take appropriate action based on them.

There is a significant problem with the use of CATA in strategic sensory studies due to the underlying metric. Unlike traditional descriptive analysis techniques, such as quantitative descriptive analysis (QDA), Spectrum™ methodology or other evaluation methods, that directly measure the intensity of an attribute, CATA measures the *number of assessors who identify the presence* of an attribute. Some correlation with intensity of an attribute is no doubt likely and the number of times assessors checked the presence of an attribute has been found to provide a good estimate of intensity (Bruzzone et al. 2012; Reinbach et al. 2014). However, cautious interpretation is necessary. It may be that an attribute with weak intensity is selected by all assessors because it is important, for example in products with 'a hint of' such as tomato sauce with a hint of garlic, but this does not mean it has a strong intensity. Essentially, CATA is an indirect and blunt measurement, albeit one that, thanks to its simplicity, is robust.

For strategic purposes, perhaps the improvement or optimization of a product, it is clearly much more helpful to a development scientist to be given an intensity target for an attribute. Even if both poles of an attribute are included in a CATA study, for example 'too salty' and 'not salty enough', to try to redevelop a product to reduce and balance CATA criticisms of saltiness would be a difficult task to undertake. This disadvantage is even more extreme in FCD as, for the reason given above, it is difficult to obtain true counts of subjects reporting the same underlying characteristic if the subjects are using many and varied words to describe their perceptions. Dooley et al. (2010) describe an approach to use CATA data to create preference maps and apply them to a study of vanilla ice creams. They also compare the results with preference maps using trained panel descriptive analysis. They conclude that the characterization of the ice creams agrees well with descriptive sensory profiles and cite the advantage of simplicity in completing the CATA task. However, they also conclude that 'the limitation of this approach is that the optimal profile derived from the CATA maps is in terms of response counts and not intensities as given by a trained panel or consumers using attribute scaling'.

In addition, if consumers do not select a term, it cannot necessarily be concluded that they consider it does not apply. Assessors may not have selected it because they are neutral or undecided about the term, or because they did not pay attention to it (Varela & Ares 2012).

Sensory product characterization and differences among samples from CATA studies are subject to multiple minor biases related to how the question is formulated (Ares et al. 2013) as described earlier, including number, choice and order of attributes. Positional bias is present in completing CATA grids. A common experience is that items near the tops of columns and in the leftmost column tend to be checked more than other grid positions. Strategies to overcome these effects are within-assessor and across-assessor randomization of CATA terms, although within-assessor randomization leads to the task taking longer to perform (Lee et al. 2013). Another bias occurs due to sample order presentation, which can be due to lack of familiarity with attributes when assessing initial samples, as well as first sample bias that occurs generally in product testing, and can be overcome using a dummy first sample (Lee et al. 2013).

CATA and FCD are more often used with consumers. However, many more consumers than trained panellists are required to produce valid results. In most instances, new consumers need to be recruited for each study, which can be time and resource intensive, although this needs to be balanced against the time and resources need to set up and maintain a trained panel.

17.7 Applications

Sensory descriptive analysis using CATA and FCD has been applied to many different products, as the references listed at the end of the chapter illustrate. They are quicker than other sensory descriptive methods, particularly traditional descriptive analysis methods, and can be used when time and resources are scarce, or when detailed information is not necessary, such as for a quick 'look-see' study. They are particularly suited for use with consumers as they are simple to perform.

CATA and FCD are often used as an add-on to other methods to provide more detail; for example, they are used with similarity-based methods, such as sorting and mapping techniques, to provide descriptions of product groups. FCD is often used in the early stages of traditional descriptive techniques with trained panels to help them elicit sensory attributes to form a lexicon. Fleming et al. (2015) recommended CATA in combination with PSP to help choose poles and characterize sensory space. When assessors use their personal lexicon to score products, or when FCD is used alongside other methods, more complex statistical analysis techniques are used; for example, it may be possible to overlay the FCD onto the consensus structure or map. FCD is an integral part of free choice profiling (see Chapter 13) and flash profiling (see Chapter 14). An example of the use of FCD as part of another method is shown in case study 2. In this instance, the products' scores on assessors' individual, FCD-derived lexicons are analysed using generalized Procrustes analysis (GPA) to understand how the free choice descriptors are associated with and distinguish between products.

CATA and FCD have received a lot of attention for use in understanding why consumers like products for product development and optimization. They can be used in the early stages to investigate consumer perceptions, and to elicit information for further studies. Studies can often be performed exclusively with consumers, rather than combining trained panel sensory data with consumer hedonic data. CATA has been combined with hedonic data to provide an understanding of the sensory attributes driving liking and consumer sensory segments, as a more rapid alternative to traditional preference mapping (Ares et al. 2011a; Dooley et al. 2010). Open-ended questioning has also been used to provide sensory descriptions for preference mapping (Ares et al. 2010b; ten Kleij & Musters 2003) and to provide a description of 'likes' and 'dislikes' (Symoneaux et al. 2012; Varela et al. 2014), thereby providing information on importance and liking of attributes.

CATA has been used in studies to identify ideal products, in order to list the characteristics of assessors' ideal product, which is a simpler and more intuitive task than rating ideal intensity in classic ideal product profiling (Ares et al. 2011a). Again, it has the advantage that consumers can be used to identify ideal product profiles without the need for trained panel descriptive analysis. Meyners et al. (2013) were able to identify positive and negative hedonic drivers, which can be used for optimization. Ares et al. (2011a) compared ideal profiling using different methods, including CATA, projective mapping and intensity scales. The methods provided similar information about sensory characteristics of the products and ideal products. Assessors were asked to either position their ideal product on their projective map or check all terms they considered appropriate to describe their ideal product. The results were different due to the projective map method placing the ideal product within the map, whereas CATA allowed the ideal product to be placed outside the space. However, the ideal CATA task was found to be less intuitive and more difficult.

Penalty analysis based on the comparison of consumer sensory perception and ideal product can be used to gather information on the impact of deviation from the ideal on liking scores (Ares et al. 2014d; Plaehn 2012).

FCD is particularly useful to generate consumer sensory terms, which can be used for marketing and communication strategies. It has also been suggested for use with experts, in this case wine tasters, as an alternative to providing detailed, common training for use in classic descriptive analysis (Lawrence et al. 2013).

17.8 Case Studies

17.8.1 Case Study 1: Rancidity/Staleness – The Value of CATA Compared with Rating Scales

This is a consumer research example in which salted snack products at different ages (and known to have rancidity problems with ageing) were evaluated by consumers using both rating scales (including a hedonic rating) and a CATA grid.

- Data from the rating scale 'Overall Acceptability' (on a nine-point hedonic scale) are shown in Figure 17.3. A 95% least significant difference was calculated in order to identify which products' overall acceptability mean scores

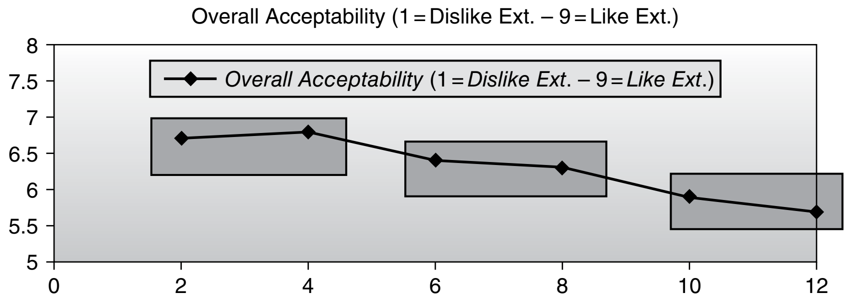


Figure 17.3 Overall acceptability of aged salted snack samples (samples that do not differ significantly (P > 0.05) are boxed together).

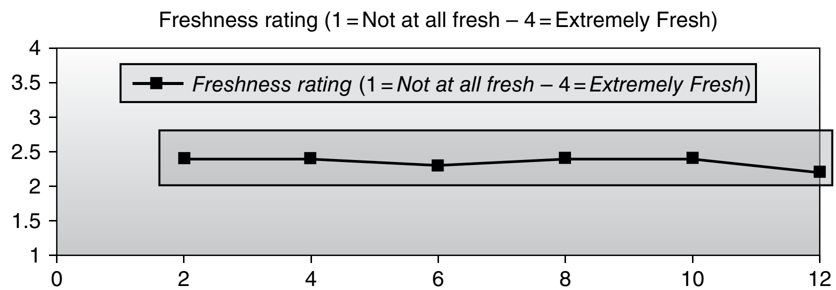


Figure 17.4 Freshness rating of aged salted snack samples (samples that do not differ significantly (P > 0.05) are boxed together).

were significantly different from one another at the P < 0.05 level. They show that consumers are discriminating between the liking scores across the different ages of samples, with acceptability for weeks 2 and 4 significantly higher than those for weeks 6 and 8, and these significantly higher than for weeks 10 and 12.

- Data from the rating scale 'Freshness' (on a four-point evaluative scale) are shown in Figure 17.4. They show that on this measure, consumers fail to discriminate between products across the age range (pairwise differences between means all P > 0.05).

However, if we look at the % CATA endorsements for 'Fresh' (on a binary yes/no basis) across the samples in Figure 17.5, again a clear pattern emerges showing discrimination between the different ages of samples. Here proportions were compared using a z-test to identify significant differences at the P < 0.05 level. Weeks 2, 4 and 6 are not significantly different, but the endorsements for 'Fresh' are significantly higher than for week 8 which, in turn, is significantly higher than for weeks 10 and 12.

We hypothesize that because all subjects rated freshness on the rating scale whether or not they were 'qualified' to rate it (on the basis of sensory sensitivity and cognitive understanding of the term), the results include 'noise' from non-discriminators and hence no significant differences emerge across samples of all ages.

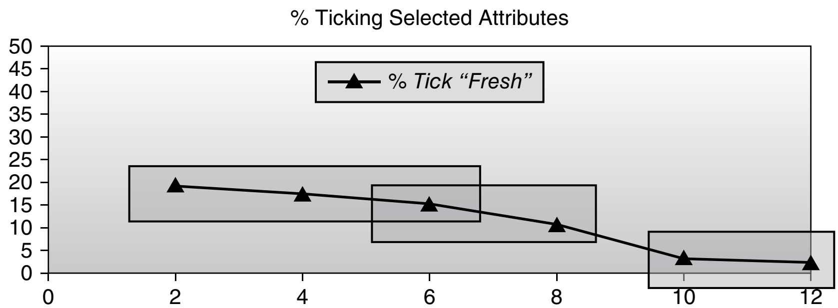


Figure 17.5 CATA endorsements of 'Freshness' of aged salted snack samples (samples that do not differ significantly (P > 0.05) are boxed together).

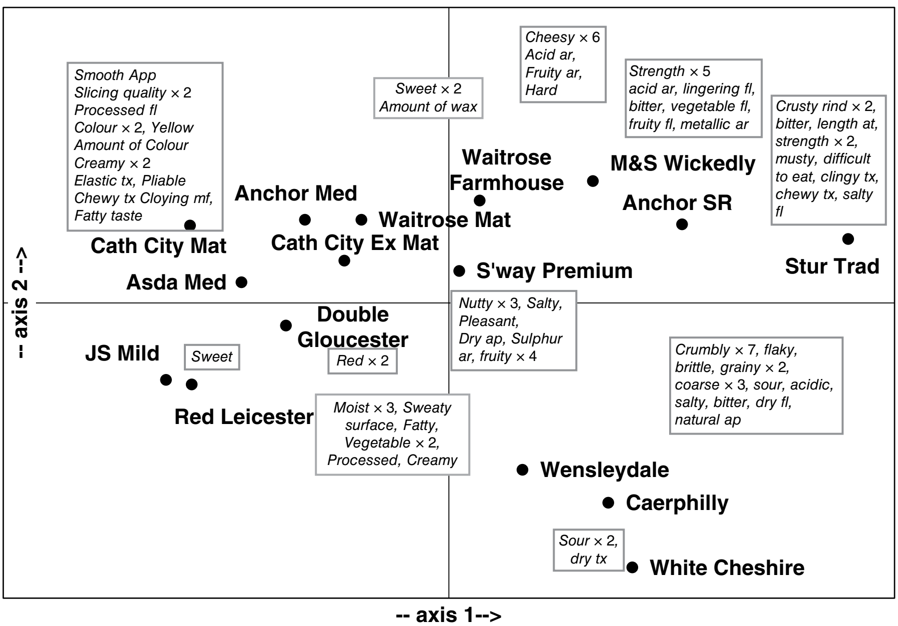


Figure 17.6 GPA consensus plot of British cheddar and territorial cheeses showing associations with free choice descriptions.

By contrast, because the CATA freshness attribute was largely checked only by those subjects who recognized freshness in the younger samples compared with the older samples, it discriminates between different aged samples much more effectively.

17.8.2 Case Study 2: Free Choice Description – Cheese Study with GPA Analysis

In this exercise the first stage was to determine a cheese vocabulary for communication and marketing purposes. Products were presented as triads and subjects were invited to identify attributes using FCD, to create their own personal vocabularies. Each subject then rated the set of cheeses against the attributes in their personal vocabulary. Subsequently, the scores were analysed using GPA which uses scaling, rotation and reflection to create a consensus plot. The resulting plot is shown in Figure 17.6 and provides an insight to how consumer language is used to describe the different cheeses. In the plot, we can see that the horizontal axis is largely driven by cheese maturity, manifested in flavour and texture terms, whereas the vertical axis tends to separate out the British territorial cheeses from the Cheddars. Of course, the principal output, so far as the communication objective is concerned, is to identify closely linked consumer descriptors via the different cheeses. So, for example, we can see the texture terms 'pliable', 'chewy' and 'elastic' describing the less matured cheeses and 'crumbly', 'flaky' and 'brittle' describing the more matured ones.

17.8.3 Case Study 3: Standard CATA – Biplot; Cochran

Fifteen subjects evaluated six meat-based snack products using a sensory CATA grid. The six products comprised the current formulation, two competitors and three development reformulations.

The sensory CATA grid included the following 21 characteristics, suggested by a previous QDA profile of products from the same category, with additions from the manufacturer's development team.

Smoky	Meaty
Bland	Pale
Dark	Rubbery
Watery	Salty
Chewy	Spongy
Tough	Smooth
Greasy	Juicy
Gristly	Dry
Peppery	Fatty
Spicy	Tasty
Moist	

Figure 17.7 shows the (partial) raw data for 15 subjects and three of the CATA characteristics, showing a '1' if the subject checked the attribute for the product or a '0' if the subject didn't check the attribute for the product.

We summarize the data for just one of the attributes, 'salty', in Figure 17.8.

With more than two products, Cochran's Q test provides an appropriate method for evaluating whether the products are significantly different. If an overall significant difference is found, a multiple comparison technique can identify what product differences are responsible, In this case, current is associated with higher citations of 'salty' compared with the other products (Figure 17.9).

Subject		Salty						Chewy						Greasy					
	Sample	A	B	C	D	E	F	A	B	C	D	E	F	A	B	C	D	E	F
1		0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
2		0	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	0	0
3		0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
4		0	0	0	1	0	0	0	1	1	1	0	1	0	0	0	0	0	0
5		0	0	1	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0
6		0	0	0	0	0	0	0	0	1	1	0	1	1	0	0	1	0	1
7		0	1	0	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0
8		0	0	0	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0
9		0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0
10		1	1	1	1	1	1	1	0	1	1	0	1	0	0	0	0	0	0
11		0	0	0	1	0	0	1	1	1	1	1	1	0	0	0	0	0	0
12		0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0
13		0	0	0	1	0	0	1	1	1	1	1	0	0	0	0	0	0	0
14		0	0	0	1	0	0	1	0	1	1	0	1	0	0	0	0	0	0
15		0	0	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0

Figure 17.7 Case study 3: small selection of the input data for three attributes of the six meat-based snack products (where: A = Dev't. 3 ; B = Dev't. 1; C = Comp. 1; D = Current; E = Dev't. 2 and F = Comp. 2).

Sample	Observations	Minimum	Maximum	Mean	Std. deviation
Dev't. 3	15	0.0	1.0	0.07	0.258
Dev't. 1	15	0.0	1.0	0.13	0.352
Comp. 1	15	0.0	1.0	0.13	0.352
Current	15	0.0	1.0	0.67	0.488
Dev't. 2	15	0.0	1.0	0.13	0.352
Comp. 2	15	0.0	1.0	0.13	0.352

(Mean = proportion of subjects checking attribute for product)

Figure 17.8 Case Study 3: summary statistics for one selected attribute – 'salty'.

Q (Observed value)	32.17
Q (Critical value)	11.07
DF	5
p-value (Two-tailed)	< 0.0001
alpha	0.05

The Cochran's Q test shows there is a significant difference between products and the multiple comparison test, below, shows the Current Sample (D) to be more associated with high saltiness than the other samples

Sample	Frequency	Sum of ranks	Mean of ranks	Groups	
Dev't. 3	15	46.0	3.07	1	
Dev't. 1	15	49.0	3.27	1	
Comp. 1	15	49.0	3.27	1	
Dev't. 2	15	49.0	3.27	1	
Comp. 2	15	49.0	3.27	1	
Current	15	73.0	4.87		2

Figure 17.9 Case Study 3: Cochran's Q test for significance for one selected attribute – 'salty'.

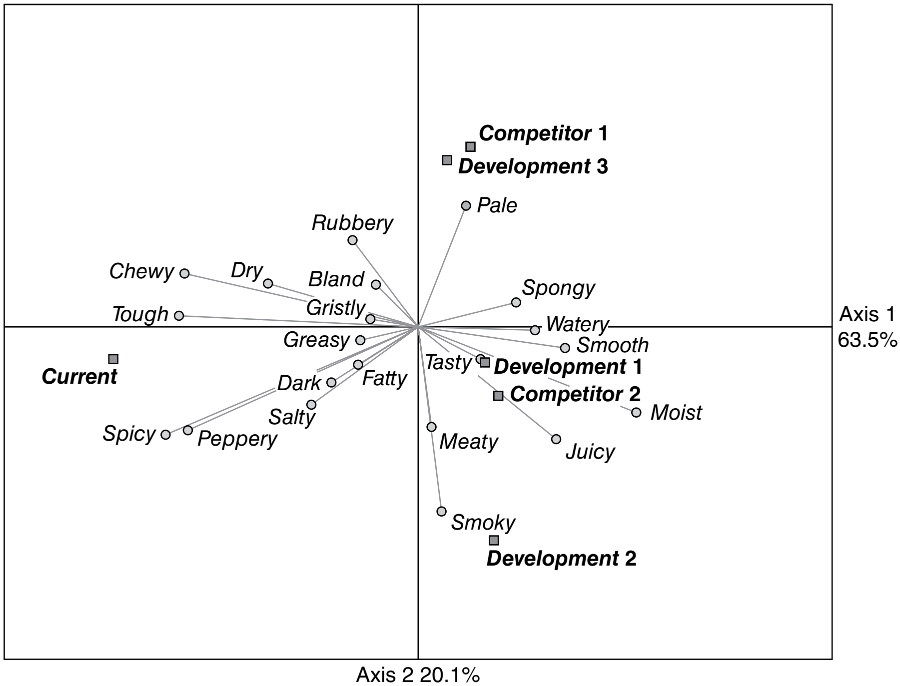


Figure 17.10 Case study 3: correspondence analysis biplot of CATA sensory citations for six meat-based snacks.

With six products, it is possible to create a CA map, showing the associations between sensory characteristics and products. Figure 17.10 shows the current product to be strongly distinguished from the others, mostly due to higher citations of spicy and seasoning characteristics and a more challenging texture. The map also suggests that the product developers have created formulations that closely match the competitors' products. Finally, one of the development products (Dev't 2) offers another, different product experience: more smoky, combined with some juiciness and some spice/seasoning.

In this study, subjects used a second CATA grid to evaluate more emotive characteristics of the products. The emotive CATA grid comprised 13 characteristics chosen from a marketing lexicon in conjunction with the client's marketing team.

Appetizing	Poor quality
Cheap	Processed
Familiar	Tastes artificial
Healthy	Looks natural
Good quality	Unhealthy
Looks artificial	Succulent
Tastes natural	

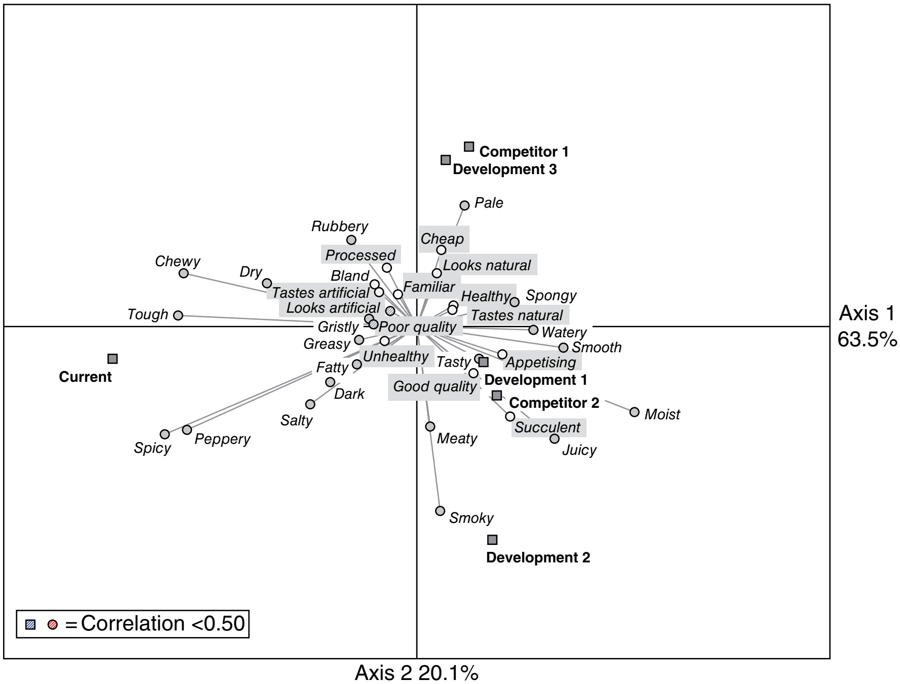


Figure 17.11 Case study 3: correspondence analysis biplot of CATA sensory citations for six meat-based snacks with 'emotional' CATA citations plotted as passive points.

Including data for these terms as passive variables into the correspondence analysis allows us to see how the terms relate to the products and the sensory characteristics without altering the sensory configuration.

Figure 17.11 shows that the terms we could reasonably expect to be positive ('appetizing'; 'good quality'; 'tastes natural', etc.) tend to be associated more with products and sensory attributes towards the right of the map, that is, away from the current product, indicating that the development samples may indeed provide an opportunity for product improvement.

17.9 Summary

CATA has established itself as a powerful and robust technique for rapid description of product characteristics without the need for extensive training. This makes it, along with FCD, particularly suitable for use by 'naive' consumers rather than trained sensory panellists, although there may be circumstances when the techniques may be useful to a trained panel at the earliest stages of vocabulary development or when a quick 'look-see' is needed. For consumer studies, CATA is much simpler than rating scales for consumers to understand

and use and in some circumstances, it could be argued that it provides a better description than the more rationally considered rating scales.

For profiling products, CATA data are simpler to analyse and understand than FCD, due to the difficulties in coping with the diverse and extensive vocabulary generated by the latter. However, a price to pay for this simplicity is that CATA profiles have the potential to be limited by the initial choice of attributes presented to assessors. Note, however, that for communication investigations, this very diversity of language in FCD can be a distinct benefit.

Product profiles generated by CATA and FCD show good correspondence with those from formal, trained panel descriptive analysis. However, there is a major drawback in using either CATA or FCD to characterize products for *strategic* improvement and optimization, because they are based on response counts rather than intensities. Statistically, response counts provide less information than intensities and also counts are less tractable in defining sensory targets.

Research into CATA to improve the methodology and extend its scope, for example into TCATA, continues but the technique has proved itself time and again to be a useful consumer approach to characterizing products in sensory terms.

17.10 Future Developments

Researchers continue to investigate aspects of the CATA and FCD methodology in order to support their scientific basis and application to sensory characterization of products. There has been a recent plethora of studies, cited earlier in this chapter, on the methodology details of CATA to investigate sources of bias and strengthen reliable and reproducibility. To date, less research has been carried out on FCD as a stand-alone or supplemental method to provide sensory description and it would certainly benefit from improved and streamlined analysis techniques for FCD responses. It is expected that both methods will receive continued attention.

New methodological variations are being developed for sensory characterization, such as TCATA and RATA, and different way of asking FCD questions, which are also receiving detailed methodological investigation.

Research is ongoing into the benefits of using CATA and FCD in conjunction with other methods and analyses such as temporal dominance of sensations (Meyners 2016b) and penalty lift analysis (Meyners 2016a), in order to elicit deeper insights into the appreciation and/or characterization of products.

It is anticipated that the application of CATA and FCD in product development and optimization to understand consumer liking and sensory-based segmentation and identify ideal products will increase. CATA and FCD are amenable for use through social media and portable devices due to their simplicity, and there is scope to gather sensory data from consumers in real-life contexts using these techniques.

References

Adams, J., Williams, A., Lancaster, B. & Foley, M. (2007) Advantages and Uses of Check-All-That-Apply Response Compared to Traditional Scaling of Attributes for Salty Snacks. 7th Pangborn Sensory Science Symposium, 12–16 August, Minneapolis, USA.

Ares, G. & Jaeger, S.R. (2013) Check-all-that-apply (CATA) questions: influence of attribute order on sensory product characterization. *Food Quality & Preference*, **28**, 141–153.

Ares, G. & Jaeger, S.R. (2014) Check-all-that-apply (CATA) questions with consumers in practice: experimental considerations and impact on outcome. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Sawston: Woodhead Publishing, pp. 227–246.

Ares, G. & Varela, P. (2014) Comparison of novel methodologies for sensory characterization. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 365–389.

Ares, G., Barreiro, C., Deliza, R., Giménez, A. & Gámbaro, A. (2010a) Application of a check-all-that-apply question to the development of chocolate milk desserts. *Journal of Sensory Studies*, **25**, 67–86.

Ares, G., Gimenez, A., Barreiro, C. & Gambaro, A. (2010b) Use of an open-ended question to identify drivers of liking of milk desserts. Comparison with preference mapping techniques. *Food Quality & Preference*, **21**, 286–294.

Ares, G., Varela, P., Rado, G. & Giménez, A. (2011a) Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. *Food Quality & Preference*, **22**, 581–591.

Ares, G., Varela, P., Rado, G. & Giménez, A. (2011b) Are consumer profiling techniques equivalent for some product categories? The case of orange-flavoured powdered drinks. *International Journal of Food Science & Technology*, **46**, 1600–1608.

Ares G, Jaeger, S.R., Bava, C.M. et al. (2013) CATA questions for sensory product characterization: raising awareness of biases. *Food Quality & Preference*, **308**, 114–127.

Ares, G., Etchemendy, E., Antunez, L., Vidal, L., Giménez, A. & Jaeger, S.R. (2014a) Visual attention by consumers to check-all-that apply questions: insights to support methodological development. *Food Quality & Preference*, **32**, 210–220.

Ares, G., Antunez, L., Giménez, A. et al. (2014b) Further investigations into the reproducibility of check-all-that-apply (CATA) questions for sensory product characterization elicited by consumers. *Food Quality & Preference*, **36**, 111–121.

Ares, G., Tárrega, T., Izquierdo, L. & Jaeger, S.R. (2014c) Investigation of the number of consumers necessary to obtain stable sample and descriptor configurations from check-all-that-apply (CATA) questions. *Food Quality & Preference*, **31**, 135–141.

Ares, G., Dauber, C., Fernández, E., Giménez, A. & Varela, P. (2014d) Penalty analysis based on CATA questions to identify drivers of liking and directions for product reformulation. *Food Quality & Preference*, **32**, 65–76.

Ares, G., Bruzzone, F., Vidal, L. et al. (2014e) Evaluation of a rating-based variant of check-all-that-apply questions: rate-all-that-apply (RATA). *Food Quality & Preference*, **36**, 87–95.

Ares, G., Antúnez, L., Bruzzone, F. et al. (2015a) Comparison of sensory product profiles generated by trained assessors and consumers using CATA questions: four case studies with complex and/or similar samples. *Food Quality & Preference*, **45**, 75–86.

Ares, G., Antunez, L., Giménez, A. & Jaeger, S.R. (2015b) List length has little impact on consumers' visual attention to CATA questions. *Food Quality & Preference*, **42**, 100–109.

Ares, G., Reis, F., Oliveira, D. et al. (2015c) Recommendations for use of balanced presentation order of terms in CATA questions. *Food Quality & Preference*, **46**, 137–141.

Ares, G., Castura, J.C., Antúnez, L. et al. (2016) Comparison of two TCATA variants for dynamic sensory characterization of food products. *Food Quality & Preference*, **54**, 160–172.

Becue-Bertaut, M. (2014) Tracking verbal-based methods beyond conventional descriptive analysis in food science bibliography. A statistical approach. *Food Quality & Preference*, **32**, 2–15.

Bruzzone, F., Ares. G. & Gimenez, A. (2012) Consumers' texture perception of milk desserts. II – Comparison with trained assessors' data. *Journal of Texture Studies*, **43**, 214–226.

Buck, D., Hasted, A. & Taylor, Y. (2013) Poster [P1.3.017]: Improving consumer response to check all that apply (CATA) attribute batteries. 10th Pangborn Sensory Science Symposium, Rio de Janeiro, Brazil, 11–15 August.

Campo, E., Do, B.V., Ferreira, V. & Valentin, D. (2008) Aroma properties of young Spanish monovarietal white wines: a study using sorting task, list of terms and frequency of citation. *Australian Journal of Grape & Wine Research*, **14**, 104–115.

Campo, E., Ballester, J., Langlois, J., Dacremont, C. & Valentin, D. (2010) Comparison of conventional descriptive analysis and a citation frequency-based descriptive method for odor profiling: an application to Burgundy Pinot noir wines. *Food Quality & Preference*, **21**, 44–55.

Castura, J.C., Antúnez, L., Giménez, A. & Ares, G. (2016) Temporal Check-All-That-Apply (TCATA): a novel dynamic method for characterizing products. *Food Quality & Preference*, **47**, 79–90.

Deneulin, P. & Bavaud, F. (2015) Analyses of open-ended questions by renormalized associative and textural networks: a study of perception of minerality in wine. *Food Quality & Preference*, **47**, 34–44.

Dooley, L., Lee, Y.S. & Meullenet, J.F. (2010) The application of check-all-that-apply (CATA) consumer profiling to preference mapping of vanilla ice cream and its comparison to classical external preference mapping. *Food Quality & Preference*, **21**, 394–401.

Ennis, D.M. & Ennis, J.M. (2013) Analysis and thurstonian scaling of applicability scores. *Journal of Sensory Studies*, **28**(3), 188–193.

Fleming, E.E., Ziegler, G.R. & Hayes, J.E. (2015) Check-All-That-Apply (CATA), sorting, and Polarized Sensory Positioning (PSP) with astringent stimuli. *Food Quality & Preference*, **45**, 41–49.

Fonseca, F.G.A., Esmerion, E.A., Tavares Ilho, E.R., Ferraz, J.P., da Cruz, A.G. & Bolini, H.M.A. (2016) Novel and successful free comments method for sensory characterization of chocolate ice cream: a comparative study between pivot profile and comment analysis. *Journal of Dairy Science*, **99**, 3408–3420.

Guiacalone, D. & Hedelund, P.I. (2016) Rate-all-that-apply (RATA) with semi-trained assessors: an investigation of the method reproducibility at assessor-, attribute- and panel-level. *Food Quality & Preference*, **51**, 65–71.

Jaeger, S.R. & Ares, G. (2014) Lack of evidence that concurrent sensory product characterisation using CATA questions bias hedonic scores. *Food Quality & Preference*, **35**, 1–5.

Jaeger, S.R., Chheang, S.L., Yin, J. et al. (2013a) Check-all-that-apply (CATA) responses elicited by consumers: within-assessor reproducibility and stability of sensory product characterizations. *Food Quality & Preference*, **30**, 56–67.

Jaeger, S.R., Giacalone, D., Roigard, C.M. et al. (2013b) Investigation of bias of hedonic scores when co-eliciting product attribute information using CATA questions. *Food Quality & Preference*, **30**, 242–249.

Jaeger, S.R., Cadena, R.S., Torres-Moreno, M. et al. (2014) Comparison of check-all-that-apply and forced-choice Yes/No question formats for sensory characterisation. *Food Quality & Preference*, **35**, 32–40.

Jaeger, S.R., Beresford, M.K., Paisley, A.G. et al. (2015a) Check-all-that-apply (CATA) questions for sensory product characterization by consumers: investigations into the number of terms used in CATA questions. *Food Quality & Preference*, **42**, 154–164.

Jaeger, S.R., Hunter, D.C., Kam, K. et al. (2015b) The concurrent use of JAR and CATA questions in hedonic scaling is unlikely to cause hedonic bias, but may increase product discrimination. *Food Quality & Preference*, **44**, 70–74.

Kahneman, D. (2011) *Thinking Fast and Slow*. New York: Farrar, Straus and Giroux.

Lawrence, G., Symoneaux, R., Maitre, I., Broaasaud, F., Maestrojuan, M. & Mehinagic, E. (2013) Using free comments method for sensory characterisation of Cabernet Franc wines: comparison with classical profiling in a professional context. *Food Quality & Preference*, **30**, 145–155.

Lee, Y., Findlay, C. & Meullenet, J.F. (2013) Experimental considerations for the use of check-all-that-apply questions to describe the sensory properties of orange juices. *International Journal of Food Science & Technology*, **48**, 215–219.

Le Fur, Y., Mercurio, V., Moio, I., Blanquet, J. & Meunier, J.M. (2003) A new approach to examine the relationships between sensory and gas chromatography olfactometry data using generalized procrustes analysis applied to six French Chardonnay wines. *Journal of Agricultural & Food Chemistry*, **51**, 443–452.

McCloskey, I.P., Sylvan, M. & Arrhenius, S.P. (1996) Descriptive analysis for wine quality experts determining appellations by Chardonnay wine aroma. *Journal of Sensory Studies*, **11**, 49–67.

Meyners, M. (2016a) Testing for differences between impact of attributes in penalty-lift analysis. *Food Quality & Preference*, **47**, 29–33.

Meyners, M. (2016b) Temporal liking and CATA analysis of TDS data on flavored fresh cheese. *Food Quality & Preference*, **47**, 101–108.

Meyners, M. & Castura, J.C. (2014) Check-all-that apply questions. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 271–306.

Meyners, M., Castura, J. & Thomas Carr, B. (2013) Existing and new approaches for the analysis of CATA data. *Food Quality & Preference*, **30**, 309–319.

Meyners, M., Castura, J.C. & Worch, T. (2016a) Statistical evaluation of panel repeatability in Check-All-That-Apply questions. *Food Quality & Preference*, **49**, 197–204.

Meyners, M., Jaeger, S.R. & Ares, G. (2016b) On the analysis of Rate-All-That-Apply (RATA) data. *Food Quality & Preference*, **49**, 1–10.

Piqueras-Fiszman, B. (2014) Open-ended questions in sensory testing practices. In: Delarue, J., Lawlor, B. & Rogeaux, M. (eds) *Rapid Sensory Profiling Techniques and Related Methods: Application in New Product Development and Consumer Research*. Sawston: Woodhead Publishing, pp. 247–268.

Plaehn, D. (2012) CATA penalty/reward. *Food Quality & Preference*, **24**, 141–152.

Popper, R., Abdi, H., Williams, A. & Kroll, B.J. (2011) Multi-block Hellinger analysis for creating perceptual maps from check-all-that-apply questions. Presented at 9th Pangborn Sensory Science Symposium, 7 September, Toronto, Canada.

Reinbach, H.C., Giacalone, D., Ribeiro, L.M., Bredie, W.L.P. & Frøst, M.B. (2014) Comparison of three sensory profiling methods based on consumer perception: CATA, CATA with intensity and Napping®. *Food Quality & Preference*, **32**, 160–166.

Seo, H.S., Lee, S.Y. & Hwang, I. (2009) Development of a sensory attribute pool of brewed coffee. *Journal of Sensory Studies*, **24**, 111–132.

Symoneaux, R. & Galmarini, M.V. (2014) Open-ended questions. In: Varela, P. & Ares, G. (eds) *Novel Techniques in Sensory Characterization and Consumer Profiling*. Boca Raton: CRC Press, pp. 307–332.

Symoneaux, R., Galmarini, M.V. & Menhinagic, E. (2012) Comment analysis of consumers' likes and dislikes as an alternative tool to preference mapping. A case study on apples. *Food Quality & Preferences*, **24**, 59–66.

Ten Kleij, F. & Musters, P.A.D. (2003) Text analysis of open-ended survey responses: a complementary method to preference mapping. *Food Quality & Preference*, **24**, 43–52.

Valentin, D., Chollet, S., Lelievre, M. & Abdi, H. (2012) Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, **47**(8), 1563–1578.

Varela, P. & Ares, G. (2012) Sensory Profiling, the blurred line between sensory and consumer science. A review of novel method for product characterization. *Food Research International*, **48**, 893–908.

Varela, P., Beltrán, J. & Fiszman, S. (2014) An alternative way to uncover drivers of coffee liking: preference mapping based on consumers' preference ranking and open comments. *Food Quality & Preference*, **32**, 152–159.

Worch, T. & Piqueras-Fiszman, B. (2015) Contributions to assess the reproducibility and the agreement of respondents in CATA tasks. *Food Quality & Preference*, **40**, 137–146.

Zwillinger, S.A. & Halpern, B.P. (1991) Time-quality tracking of monosodium glutamate, sodium saccharin and a citric acid-saccharin mixture. *Physiology & Behavior*, **49**, 855–862.

SECTION 3
Applications

CHAPTER 18

Application of Descriptive Sensory Analysis to Food and Drink Products

Cindy Beeren

18.1 Introduction

A first-time purchase of a food or drink by a consumer is likely to be driven by information gathered from different sources, such as commercials, recommendations by family or friends or in-store promotions, or the decision may be influenced by the price. At this point of the first purchase, little information would be available about the sensorial characteristics; a consumer would not yet know whether he or she personally would find the product as tasty as a friend made it sound or whether the product really looks as good on the table once unpacked and prepared as the commercial made it look.

A few exceptions exist where sensorial aspects do play an important role on first purchase. In particular, the human senses of sight and/or smell may play a significant part; with our sense of sight, we measure visual aspects of products and therefore many products have appealing packaging and positioning on the shelf. In addition, for a smaller number of products, the product itself is visible to the consumer; for example, a pizza may be seen through a window in the packaging or a cheese could be seen at the counter. These see-through packaging formats appear in general to be favoured by consumers (Beeren 2011). For these occasions, sight would play a major role in the first purchase decision. With our sense of smell, we perceive the aroma of the food or beverage, which plays an important part during a first purchase for a minority of products, but where it does, it may be vital – for example, the aroma coming from a bakery aisle or from a roasted chicken in a deli (Beeren 2010).

During any purchase, first or subsequent, the consumer will have a desire for specific sensorial elements. The consumer may be looking forward to the refreshing flavour of a lemon drink, to the succulent texture of a steak or the creamy, melting mouthfeel of an ice cream. Subsequent food and drink purchases are

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

usually largely based on previous positive sensory experience(s) and thus the right sensory profile is vital for long-term product success. Descriptive sensory analysis is a powerful tool that measures these sensorial desired characteristics and can help to create and improve products.

To develop and position a food or drink accurately, involvement of the target consumer is vital. The consumers' needs or wishes in terms of product ideas, concepts and potential sensory characteristics should be known by the product developer. Descriptive sensory analysis can provide product developers with information that will allow the translation of consumer liking into tangible sensory product characteristics. Any developed prototypes, manufactured food and drinks and competitive products can then be compared against these identified desirable product characteristics.

18.2 General Principles of Descriptive Sensory Analysis Methods as Applied to Food and Drink Products

The general principles of descriptive sensory analysis are very similar to any type of sensory testing. It is important to consider all the different steps to ensure that testing is carried out correctly and thus to obtain meaningful results.

Figure 18.1 shows the suggested process flow. Each of the individual steps is explained hereafter.

18.2.1 Test Objective

Prior to starting any sensory test, it is important to consider the objective of the test. What is the type of data required? Who is likely to use the data and for which purpose? It may, for example, be that one would like to compare a current own ready meal to a competitor, to identify if a new prototype soup adheres to the product characteristics desired by the consumer, to measure the changes in a milk powder during its shelf-life or to determine the impact of a process change on the product. Whatever the purpose for the evaluation, it is vital to take a moment to phrase the objective of the test.

18.2.2 Identification of Test Method

After identification of the objective, an appropriate test method can be chosen. For several of the given example objectives, the only requirement might be to evaluate whether a difference exists between two or more products. For this objective, a descriptive sensory analysis test may not be the most appropriate tool. Discrimination testing is on many occasions more straightforward and is in general more sensitive in measuring the overall difference between two or more foods, drinks or ingredients.

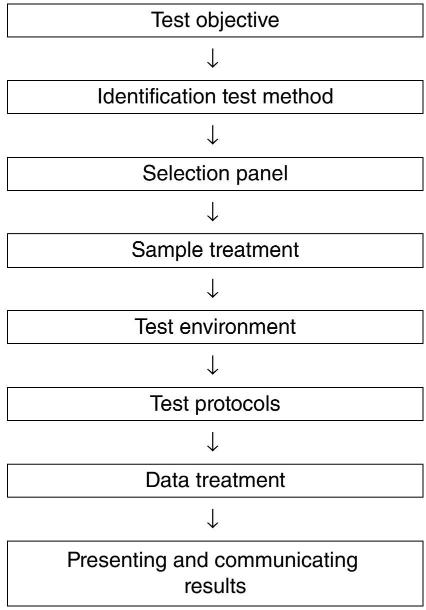


Figure 18.1 Process flow sensory analysis.

When the objective of testing extends to the desire to understand the product characteristics, a descriptive test method is often a more suitable test. Chapter 20 takes a closer look at the various descriptive methods that can be considered. As many types of descriptive sensory analysis have been developed over the years, depending on the objective, a particular test may be more suitable than others. The selection of the test should also consider the practicalities at hand and what might be viable within company limitations, such as accessibility to a (trained) panel (trained in the chosen test method), timing constraints, sample availability and method and data analysis expertise to provide useful and reliable sensory data.

18.2.3 Panel Selection

Most descriptive sensory analysis test methods use trained sensory assessors. Assessors require training in the method to be used, samples, protocols, attributes and scale usage. Assessors can be recruited internally or externally. Advantages of using internal assessors include their presence in case of an urgent need for evaluation and potentially being more cost-effective; disadvantages of using employees are likely time constraints and bias due to additional product and project knowledge. Externally recruited assessors are less likely to have a personal interest in the outcome of the test and they would usually be dedicated to the sensory test at the time of testing, rather than trying to do it quickly between some other tasks. External assessors are thus more likely to concentrate better on the job at hand. Many companies choose to use internal assessors for the purpose of quality control; in this case assessors frequently need to be readily available for short time periods, whilst for larger research and development projects external assessors are regularly chosen as the panel of choice, as large projects often involve longer sessions for assessors, making it less viable for employees to participate.

Other practical considerations when using internal assessors include:

- support from upper management
- communication from upper and middle management to give permission to release employees for the required time
- clarity on time requirements for assessors and their immediate superior
- when departments are small, it may be better to spread participation among a number of departments
- motivation and interest from assessors
- morale boosters such as group lunches together
- participation in only one group to prevent conflict of demands
- technical abilities including taste acuity, ability to deal analytically with complex test situations (test assessors for ability to discriminate and judge samples), good health, stable personality (neither passive nor overly dominant and not too easily distracted) and ability to verbalize perception (Zook & Wessman 1997).

For most descriptive sensory analysis techniques, the use of screened and selected assessors is recommended. Depending on the test objective, it may be appropriate to use assessors who are particularly sensitive to certain stimuli (e.g. certain off-notes when off-notes can be an issue, or sweet and bitter taste when working with particular sweeteners) or to use assessors who can verbalize their perception particularly well.

Assessors may respond differently to test products depending on the individual, likes/dislikes, sensitivity, knowledge of the test (objective and products), mood, familiarity of products, time of day, etc. A panel of assessors should therefore be employed to reduce individual factors.

Initial screening when building an internal or external food or drink sensory panel would normally include questions on food allergies and intolerances, food neophobia (to identify assessors unwilling to taste new/unfamiliar foods), food likes and dislikes, dietary requirements (including vegetarian and religious food practices). All the questions are aimed at building a panel that allows as much general use as possible, in line with foods and drinks usually evaluated by the company's sensory panel(s) and considering health and personal choices of individuals.

Training will facilitate a more objective evaluation of foods and drinks and therefore, in general, the better trained assessors are, the fewer assessors are required. However, panel leaders should always consider that however well trained assessors are, (some) variability will still exist between them, not least in sensitivity levels, daily distractions/concentration and motivation. Factors influencing the panel size include the test method selected, the amount of test samples and the potential consequences of wrong decisions being made.

18.2.3.1 Sensory Screening of Assessors

Assessors used for sensory evaluation of foods and drinks are typically screened and selected on their ability to distinguish small differences in appearance, aroma, taste, flavour and texture and their ability to verbalize these differences. Generally, screening tests include a basic taste recognition test, in which four or five basic tastes would be presented and evaluated, consisting of sweet, salt, sour, bitter and possibly umami (ISO 2012). In some instances, astringent is included as an additional sensory characteristic, for example, for chocolate or wine panels. As assessors are not always familiar with all basic tastes, prior exposure to all components is appropriate.

Other typical screening tests include detection and recognition of aromas, whereby generally between five and ten different compounds are evaluated and assessors try to recognize and describe the perceived aroma (ISO 2012). With this test, assessors are usually graded for their performance, receiving three points for every odour that is correctly identified, two points for an almost correct answer (‘description in general terms’), one point for an odour with appropriate association and no points for an unrelated odour or no odour perception. The ISO standard for the selection, training and monitoring of selected assessors and expert sensory assessors (ISO 2012) suggests using materials such as benzaldehyde for almond/cherry aroma and vanilla for vanilla aroma. Incorporating common food taint aromas, such as trichloroanisoles (TCA) and trichlorophenols (TCP), would be recommended too, as about 70% of taints in food and drinks are caused by these two components. Taint issues should be picked up during quality control (QC) and it is thus important that assessors carrying out sensory evaluation in QC are sensitive to these and other components that may be relevant to the produced foods and/or drinks.

To ensure that assessors are capable of carrying out descriptive analysis, assessment of their ability to describe products is a common part of a screening procedure. Term derivation could be performed on any specific food or beverage or a picture illustration – this could be from a food product or any product that is appropriate to the samples to be evaluated in the panels selected for. Asking assessors for texture and mouthfeel attributes of foods and drinks, ideally those that would be evaluated once participating in panels, may be a good way to measure their descriptive ability. Other modalities (such as appearance, aroma or flavour) could be chosen as well; however, often other modalities are already covered in other screening tests, so focusing a term derivation on the texture modality could balance the testing. In addition, texture is a difficult modality for the general consumer to describe, and thus those potential assessors performing well in this test may stand out.

Other tests that may be used for assessor screening are dependent on the tasks of the assessors, and could include specific discrimination tests, such as triangle or paired comparison, memory tests, tests to identify visual impairments, ranking tests, etc. ISO standards provide further details on screening of assessors (ISO, 2012).

It is important that the sensory analyst keeps personal details of assessors, such as allergy information or strong likes or dislikes. Those with particular allergies may be unsuitable to join a sensory panel for their own health.

18.2.3.2 Sensory Training of Assessors

To fully comprehend the test products and entirely understand the test methods and procedures, sensory training for assessors is required. Any assessor not trained prior to the assessment in the methods or products used may be less confident in using the scales provided and more likely to stay within a small range of the scale-width provided, potentially resulting in indiscrimination of samples evaluated. Some assessors may, in contrast, use large parts of the scale, which, if not trained properly, may lead to providing inconsistent and more variable scores.

Exposure to products will help to identify the varieties of the products concerned. Heckel and Wilson (2002) describe the recognition of deviation from the norm in terms of appearance and flavour by a trained assessor; assessors not exposed to several samples may not catch the potential variation and problems.

Assessor training may include background details on the product itself, for example details on ingredients and the production process to discover the sources of specific product characteristics. If, for example, a panel is created for the evaluation of coffee, it may be helpful for the assessors to learn about the origin, growth and processing of coffee plants and beans in order to understand the source of different sensory characteristics of the coffees. Exposure to several test products and variants is vital during training sessions; this may include spiked test samples. Spiking test samples could be carried out by adding chemicals to induce particular flavour characteristics or basic tastes to check on differentiation of samples with common taste sensations (Word & Gress 1981) and learn new variations.

During training sessions, the test method should be explained and practised and test procedures should be discussed. Any customer feedback, case studies and unacceptable samples could be shared and discussed to gain a better understanding of the product quality variations.

For training purposes, multiple sessions are usually held, with the number of sessions dependent on the type and complexity of the samples, test and protocols, number of attributes, previous knowledge, etc. Panel feedback can be applied to measure assessors' understanding of attributes and within-panel agreement. Using analysis of variance (ANOVA), one could also measure whether selected attributes discriminate different products (Powers 1988). When carried out properly, training also tends to help to motivate assessors. At the time of writing, the American Society of Testing and Materials (ASTM) is working on a standard for panel performance, which may be worth checking.

18.2.4 Sample Treatment

When comparing two or more foods or drinks, it is vital to treat the products equally, which may sound obvious but can often be overlooked. The samples should be obtained in a similar way (e.g. purchase competitor and own products ideally from the same store) and ensure that samples are of a similar age. Products available in different packaging formats, such as soft drinks available in PET bottles, glass bottles and cans, should be purchased in the same packaging format if the objective of the test is to compare product characteristics from own and competitor product. Product variability, including differences due to seasonal variation, should be minimized within a test, so that each assessor evaluates the same stimuli and samples are compared fairly (unless the test is designed to test variability between samples). Fresh products, such as fruit, vegetables, meat and fish, can be variable within and across each individual piece. Care needs to be taken to use samples of the same ripeness, age and size, taken from a similar and representative area, such as minimum distance from stalk/core area, from the same cut of meat.

Storage of food and drink samples should also be identical for all samples within the test, unless the objective is to measure the effects of storage under different conditions. All samples should be stored away from any strong-smelling products or environment.

White or clear materials are usually chosen for product serving and three-digit coding will reduce bias of any association with a brand. If the objective is to focus on the sensory characteristics, samples should be offered in blinded cups (e.g. dark cups with lids).

Sample treatment should be in line with the test objective. Often the food and drink products are prepared in a way that is typical of how a consumer would prepare them at home. However, each assessor participating in the descriptive sensory analysis will evaluate the samples in the same format. For example, products that are likely to be subject to personal customization, such as coffee or tea, would be offered to the assessors in a defined format; the objective of descriptive analysis is not whether the sample is liked, it is to describe the sensory characteristics of a product.

As during the sensory evaluation of food and drink products, samples are consumed, hygienic practices are vital during all stages of sample storage, preparation and serving. Appropriate food hygiene training should be in place for all staff involved during the process of storage, preparation and serving. Periodic auditing of food and drink handling practices is recommended, ideally by someone with a food and safety or quality background.

Food and drink storage and preparation areas should be equipped with the necessary tools to store, prepare and serve the samples in a hygienic and proper manner. All materials should be used for foods only (e.g. not contaminated with chemicals), only food-grade materials should be in the environment and cleaning procedures should be in place.

The presentation of samples should be consistent, using consistent preparation methods delivering the same volume of sample(s), at the same temperature and in a same manner during every evaluation. Within one test, samples are ideally prepared by the same person, avoiding differences due to preparation. When multicomponent products are evaluated, it should be checked that all components are given to each of the assessors.

Amount of samples tested in a session should be kept to a reasonable level, particularly samples with a strong flavour, and acidic products should be restricted to a minimum. If using internal assessors, it may also be difficult for them to spend too much time away from their normal activities. Generally, multiple panels with fewer samples would be favoured above one session with larger sample numbers. To reduce carry-over effects and adaptation, samples of a similar strength are presented in a balanced randomized design.

The procedures to representatively sample food and drinks and to store, prepare and present the products should be documented in company standard operating procedures. Sensory panel leaders working on the identified areas should be aware and trained in the procedures. In addition, records should be kept during sampling to ensure information about samples, storage conditions, preparation details and presentation specifics is kept for future reference.

18.2.5 Test Environment

The environment in which sensory testing is carried out has an influence not only on product perception but also psychologically, on the perceived importance of the testing to the company. A dedicated room for testing shows the company's commitment to sensory testing. Using the same dedicated area for all tests would also lead to more consistent results; the room has an effect on the presentation of products, appearance of the visual characteristics, texture of samples through temperature and humidity, aroma through ambient odour and comfort to assessors. The testing area should be easily accessible to assessors.

Environmental influences should be considered (ISO 1988).

- Consistent lighting: artificial light, same types of lights throughout the room, ideally artificial daylight/north light fluorescent should be considered for optimal colour assessments.
- Noise may distract assessors and thus influence performance.
- Odours coming from production or preparation area should be kept to a minimum.
- Decoration may influence the perception of the visual appearance; neutral colours, such as matt off-white or light neutral grey, are recommended.
- Individual testing booths to limit distraction and avoid communication between assessors.

18.2.5.1 Storage Area

Preparation and storage areas adequate for testing purposes should also be in place to present samples appropriately and as such the following should be considered.

- Storage areas need to be maintained and kept clean and out-of-date samples should be removed. A cleaning schedule should be in place.
- To maintain food safety, good hygiene practices must be in place for all storage areas.
- When samples are stored, appropriate labelling should be used, including use-by dates, product and allergy information.
- Samples are to be stored at the right temperature and be checked on their temperature before placing in the storage area.
- Temperature monitoring and calibration should be in place too for all storage areas, alongside documented corrective actions.
- Certain products should be stored away from each other for hygiene and odour transfer purposes.
  - Ideally, raw meat and fish products should be stored separately in designated fridges/freezers. Alternatively, these products should be stored in designated areas of the fridge/freezer, clearly labelled, on bottom shelves.
  - Products developed for people with allergies should be stored in dedicated areas, and kept fully wrapped.
- All products should be stored away from strong odours.
- Testing sundries (pots, plates, etc.) used for testing purposes should also be stored in appropriate, clean conditions, not near to any strong-smelling materials.

18.2.6 Test Protocols

After identification of the objective, test method, samples and assessors, the test protocols should be set; clearly defined test procedures and protocols will assist with keeping the quality consistent and reducing bias during evaluation. Test protocols will clearly define how samples are handled by the assessors, for example whether beer samples are to be consumed at refrigerated or ambient temperature and whether the samples should be swallowed or spat out. Another example: for a yoghurt sample, it should be clarified whether smoothness should be judged on appearance, on first in-mouth contact or immediately prior to swallowing. Test protocols should also consider the type of utensils to be used; for example, the texture of a chicken nugget may be assessed by using the fingers, by using cutlery or when the product is placed the mouth – each one may lead to different results. Test procedures on how to handle and evaluate the food or drink should be readily available for the assessors. Assessors should be told not to eat, drink or smoke prior to a sensory session, not to use any (strong) fragrances, not to wear lipstick and not test when having a heavy cold.

As discussed earlier, sensory panel leaders should be aware of any allergies that assessors may have, as well as being aware of medication, illness and pregnancy. The panel leader should keep assessors up to date with ingredients within the products being evaluated and assessors must be informed of known food allergens. Under the General Food Law (EC 2002; Food Safety Act 1990), companies need to ensure that the food provided is safe. Determining this includes the information provided to the assessor and thus the panel leader would need some kind of allergen alerting procedure. The Food Information Regulations (FIR) 2014 (TSO 2014) apply to non-prepacked food offered for sale. These regulations form a comprehensive model of good practice for sensory testing. Panel leaders should also ensure that no products are presented to assessors who should not be consuming them (e.g. non-pasteurized cheeses for pregnant women, products that could contain nuts to nut allergy sufferers). In some instances, consent forms will need to be signed by assessors showing their consent to test samples with specific ingredients, such as alcohol, or novel ingredients. In addition, the sensory panel leader is responsible for the samples being provided to assessors, in particular when this may relate to health and well-being. Samples volume should thus be considered when testing products with alcohol, higher calorific and salt contents.

Before providing samples to assessors, it is best practice for the sensory panel leader to assess the product themselves. This will help the panel leader to understand panel language, product strength and sample variability. The panel leader should always ensure they do not influence the opinion of assessors, and understand the importance of individual tastings during sensory sessions.

Records should be kept during testing. Factors such as length of storage, product preparation, serving temperature and evaluation time can be utilized if any abnormalities in data appear or any questions arise at a later stage.

Appropriate palate cleansers should be used to refresh assessors' palates between each test sample. Palate cleansers are used to provide a baseline from which the sensory assessments are made. An effective palate cleanser will help to avoid differences in sensory assessment across different replications. The most common palate cleansers are crackers and water, with table water crackers followed by rinsing with spring water being in general a good palate cleanser (Lucak & Delwiche 2009); however, specific palate cleansers can also be used for specific foods. An overview of useful palate cleansers is shown in Table 18.1 (ASTM 2010a; Brannan et al. 2001; Findlay et al. 2007; Lucak & Delwiche 2009; Narain et al. 2010; Nasrawi & Pangborn 1989). Assessors should always ensure they rinse their palate with water after any other palate cleanser and before evaluating the next product.

To avoid confusing assessors, the same methodology or selection of methods could be utilized and all methods need to be well explained and understood by assessors. Exact methods and protocols used during testing should remain as consistent as possible and assessors should be comfortable using these. Testing at

Table 18.1 Palate cleansers.

Sample evaluated	Potential palate cleanser
General	Filtered water (ideally multiple (5) rinses), water biscuits, bread
Fatty products	Lemon juice
Spicy products	Cucumber, yoghurt, milk
Acidic products	Milk
Soft textured products	Carrots
Sweeteners	Granny Smith apples
Caffeine solutions	0.55% Carboxymethylcellulose (CMC)
Coffee	Warm water
Wine	5 g/L pectin
Astringency	Pectin solution, CMC solution
Burn from capsaicin	Sucrose solutions
Cooling (mint)	Whole milk, pectin solution
Alcoholic samples	Distilled, demineralized, deionized or carbonated water (seltzer /club soda)

regular times, ideally appropriately to the product, but not at a time when assessors are hungry or have just eaten, will help to establish a consistent product assessment.

When foods or drink samples are shared, good hygiene practices should be used by assessors, and sufficient tools should be available, such as a sufficient amount of cutlery to use a new set for each assessment.

Replication of testing can provide more accurate results. Typically, three replicates are used during testing. The exact number of replicates will, however, vary depending on product type and amount of assessors. When products are more variable, assessors are less familiar with the product and/or fewer assessors take part in the evaluation, more replicates are recommended. The replicate data can in addition be used to measure assessor effectiveness and experimental error. For food and drink samples with high natural variability for which preparation has to be well controlled, such as fresh fish, meat or apples, the samples themselves may generate replication differences (Powers 1988). Common sense should be used when interpreting data from samples that may have large natural variation.

18.2.7 Data Treatment

Before and during testing, a lot of data is generated. When evaluating the results, one should look back at all gathered information, as some inconsistencies may influence data treatment when analysing the results; this includes, for example, sample preparation or assessor attendance. Data can be captured using computerized tools or paper. When working on paper, it is essential to double-check on calculations and/or data entry, as human errors are more likely to occur.

Most descriptive sensory analysis methods pool data from all assessors from one or multiple replications together and calculate average scores and identify characteristics that may significantly differentiate products. Individual scores by assessors should also be checked to enhance confidence in the overall results. Non-reproducible results within assessors, such as the same attribute for the same product scored inconsistently, generally show a lack of understanding in the attribute or sample variation within product. Non-discrimination between samples for specific attributes shows that samples are very much alike in a given attribute, that assessors are not using the full range of the scale or that samples themselves are variable (within sample). Consistency between assessors will give higher confidence in results but inconsistency due to differences in sensitivity between assessors may occur for some attributes. Knowing the performance and abilities of assessors well and understanding their sensitivities will help panel leaders in analysing generated data.

18.2.8 Presenting and Communicating Results

Sensory data can be difficult for the non-sensory specialist to interpret and thus reports need to be clear, concise and unambiguous. Standardization of reports will help internal or external clients to better understand reports and will make it easier for the sensory scientist to generate the reports.

The report should give a full overview of testing carried out, and usually includes the following.

- *Objective*: what is the rationale for the study? Was the test a routine or non-routine test? What was the background for this test? Were any specific standard operation procedures (SOPs) used? And if specific SOPs were used, any deviations from these should be included in the report.
- *Test method*: how was the work carried out to obtain the data? Was a specific sensory test method used? A short description of the working of the method should be included.
- *Assessors*: was a specific panel employed? How many assessors participated? Were the assessors trained in the method and the samples?
- *Samples*: how were the samples obtained? Include all details of the samples, including full product description, manufacturing date, best before date, batch code, packaging type and size. For any development samples, ensure that full details are covered, for example any new ingredients are listed. If the samples were prepared in any way, details should be included, as well as sample serving, sample storage and sample presentation details.
- *Test environment*: where was the test carried out? Was it carried out under standardized conditions? Describe whether testing was carried out at a specific temperature, in individual booths, under odour extraction, in a quiet location and whether specific lighting was applied.
- *Data collection and data analysis*: what tools were used to acquire the data? What statistical tools and which methods were used to analyse the data?
- *Results*: the result section of descriptive sensory analysis often shows both numerical data in the form of mean scores with significant differences and

graphical representation of main differentiating attributes, for example by means of spider graphs, barcharts or using graphs obtained from carrying out principal component analysis (PCA).
- *Discussion*: the discussion highlights and interprets any important aspects of the test results; it examines the reliability of the study's findings and builds in any expert knowledge from past experience or previous relevant testing.
- *Conclusion and recommendations*: conclusions give a clear and concise statement of the outcome of the test, tied back to the initially defined objectives. Any uncertainties and interpretation/reasoning for these are included, as is additional relevant expert opinion. When appropriate, clear recommendations are given on any further actions required, such as the need for further developments of the sample(s) or requirements for further testing. It could, for example, be that a product differs on several sensory characteristics from a reference sample; however, consumer testing may identify which sample is most liked by the target consumer audience.

18.3 Descriptive Sensory Analysis Application in Food and Drink Product Development

Descriptive sensory analysis is commonly used in food and drink product development. This type of analysis is very valuable when creating 'me too' products and also to investigate the impact of changes to a product, such as a new ingredient, a process change or a change in packaging. In addition, descriptive sensory analysis may be very valuable to measure changes occurring during shelf-life. If we consider the stages of product development to be design, development, implementation and compliance (Morley & Osborn 2012), it is apparent that there is an important position for descriptive sensory analysis within each of these stages.

18.3.1 Application of Descriptive Sensory Analysis in Product Development Design

The initial stage of product development encompasses brainstorming and idea generation by looking at market trends and/or the functionality of an ingredient. To develop a new concept, information is sought from literature, patents and competitive food or drink products in the market. To give a better understanding of competitive food or drinks and of the market where the potential new food or drink will be placed, descriptive sensory analysis will give an advantageous edge to the product developer. Understanding will be generated of the food or drinks already on the market, their sensory properties and the differences between them. Combining descriptive analysis with consumer acceptability testing will identify which foods or drinks are liked for which reason and may even identify specific consumer group preferences for certain characteristics and provide insight on potential gaps in the market. From the descriptive sensory analysis, a profile of desirable sensory characteristics can be created to aid the development

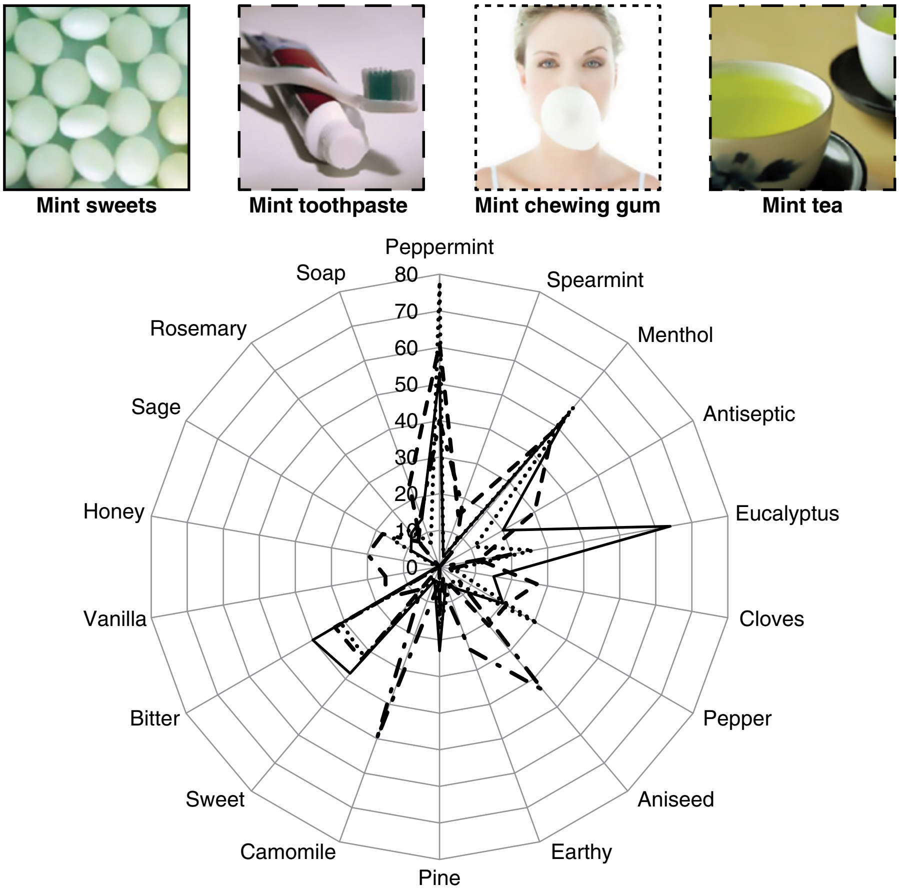


Figure 18.2 Spider graph of different mint products.

of the product and this can act as a reference guide once trial products are being created.

Descriptive analysis can also be carried out on a range of different product types. Figure 18.2 shows an example of a spider graph for different types of mint products which were evaluated for their sensory characteristics. The test included mint sweets, chewing gum, toothpaste and mint tea. Characteristics obtained from a range of products like this can inspire product developers to create a product with a different sensory profile.

18.3.2 Application of Descriptive Sensory Analysis in Product Development

Understanding the science behind food is essential in developing any product. Aspects that need to be considered include ingredient interactions, processing challenges, product microstructure, texture and stability to help ensure the success of the new product. Throughout the product development process, the developed products should be compared with the desirable sensory characteristics identified during the design stage. This comparison can simply be done by carrying out a descriptive analysis test and evaluating the attributes present in the assessed food or drinks. Including a target or main competitive product in the descriptive analysis will make it easier to evaluate the results and assess whether the newly developed samples adhere to the desired product concept. Once the desired product has been developed, a consumer benchmarking study could be carried out to evaluate the liking towards newly developed food or drink compared with current or competitive products. A major advantage of carrying out the descriptive sensory analysis throughout the development stage is to understand the food or drink in detail but there is also a financial advantage, as this test is often more cost-effective than carrying out multiple consumer studies.

Descriptive sensory analysis can compare new sample variations, identify the optimal packaging format and the optimal levels of ingredients; all are likely to have an affect on the appearance, smell, taste, texture and, depending on the product type, even sound of the products.

18.3.3 Application of Descriptive Sensory Analysis in Implementation of New Food or Drink

Taking a food or drink beyond the kitchen is always challenging and is a common point of difficulty within the innovation cycle. During scale-up to pilot plant and when carrying out factory trials, product sensory characteristics may differ from the samples developed earlier in the kitchen. Understanding the sensory differences by carrying out sensory descriptive analysis will assist in bringing the samples back to the originally intended sensory profile.

During this new product implementation stage, the product specifications are also generated.

18.3.3.1 Descriptive Sensory Analysis to Generate Sensory Specifications

It may be difficult for consumers to identify and articulate the specific sensory attributes that influence acceptability of particular food and drink products. The interaction of even simple tastes, such as sweet and acidic taste, or colour and fruit flavours can make the product evaluation and identification of exact characteristics complex (Earle et al. 2001). Developing terms to describe food and drinks' aroma, appearance, flavour and texture is therefore often carried out by trained sensory panels or by panels of product experts using descriptive sensory analysis techniques. The terms derived should be specific enough for efficient use for QC purposes; to merely describe a tomato sauce as 'red' will not give enough information, as there are many varieties of a red colour (Staniforth 2004). Similarly, to describe a flavour as 'typical' does not provide an objective definition of the food's or drink's flavour and thus more detailed descriptions are required

After creation of well-defined product descriptors, they should be correlated with the product liking of consumers. This will create objective measurement criteria for consumers' ideals and will identify the key quality attributes required for product development and quality control testing and to aid with shelf-life determination (Beeren 2010).

18.3.4 Application of Descriptive Sensory Analysis to New Product Compliance

From the very first day that food and drink products reach the supermarket shelves until the end of shelf-life, consumer sensory expectations must be met and the products need to be fundamentally safe. During the new product implementation stage, the shelf-life of the samples needs to be investigated and back-of-pack information to be completed. Although there are different techniques to measure shelf-life at different time points, such as survival analysis (Hough 2010), discrimination testing (Kilcast & Subramaniam 2000; Peryam 1958) and consumer acceptability testing (Kilcast 2000), descriptive sensory analysis is often applied.

18.3.4.1 Descriptive Sensory Analysis to Measure Changes During Product Shelf-Life

The ASTM provides the following definitions to describe shelf-life (ASTM 2005).

- 'Sensory shelf-life is the time period during which the products' sensory characteristics and performance are as intended by the manufacturer.'
- 'The product is consumable or usable during this period, providing the end-user with the intended sensory characteristics, performance, and benefits.'
- 'After this period, however, the product has characteristics or attributes that are not as intended, or it does not perform the same functions as fresh products or those selected before the end of shelf-life.'

Factors that control shelf-life include the product characteristics (structure and ingredients), storage and distribution environment and the properties of the packaging format. To measure how a newly developed product behaves during its ageing process, descriptive sensory analysis is an important tool, tracking the perception of the different sensory characteristics over time. The food or drink can be evaluated in one or more packaging systems and under one or more storage conditions to compare the effects on the product's ageing process. Shelf-life testing is usually carried out at several time points. Six time points will allow for good measuring of trends developing over time – both increases in some attributes, often related to off-flavours, and decreases in favourable ingredients can occur. The end of shelf-life is, as described by the ASTM, when the product characteristics or attributes are not as intended or when it does not perform the same function as the fresh product or those selected prior to the end of shelf-life (ASTM 2005).

To measure products over shelf-life, samples can be tested when they get to the specified age (sequential testing) or in one test (staggered design) where all the samples have been aged to the desired time point. Testing at one time point could be done by storing samples in a condition that does not age the products any further (e.g. frozen) at some point during the ageing process. The advantage of testing during ageing is that the samples are evaluated during their normal ageing process and not during altered conditions. However, from a sensory panel perspective, it means that the panel needs to be very well trained, or retrained prior to each time point, as the scale usage for certain attributes is likely to drift somewhat over time. To aid the panel with the drifting/retraining, a stable reference, such as a frozen standard product, can be useful.

Often manufacturers would like to know the shelf-life of a food or drink prior to allowing the full ageing process to happen. Accelerated tests could be used to estimate the shelf-life. Any sensorial testing carried out during the product development stages is useful to prepare this accelerated testing. The principle of accelerated testing is to store the newly developed food or drink product under a condition that accelerates normal deterioration. It should be noted, however, that when carrying out accelerated testing, food or drinks are often put under conditions which are different from normal, which in itself can create changes to the products which are not (commonly) seen during shelf-life under normal conditions. For example, increased temperature or humidity might start different chemical processes and thus may create different characteristics in the food or drinks.

Changes measured using descriptive sensory analysis should always be compared to ambient storage conditions to create a predictive model first. It is imperative to ensure that microbiological safety is not compromised under the accelerated conditions.

18.4 Application of Descriptive Sensory Analysis in Food and Drink Quality Control and Quality Assurance

The role of sensory analysis in quality assurance (QA) usually encompasses the maintenance of consistency of the food or drink's quality, protection against taints and often evaluation of shelf-life. Before, during and after production, ingredients and manufactured products are evaluated and compared against a sensory specification or control sample.

In most cases, internal employees are utilized to carry out the testing. Testing often occurs immediately when a new ingredient arrives on site and after the food or drink is produced, leading to irregular and often frequent testing; assessors thus need to be readily available.

The type of testing carried out is dependent on the availability of assessors, test frequency, product numbers and type, panel training requirements, test facility availability and data handling requirements. In many cases for a product check, a simple test is carried out (such as in/out of specification or a simple quality rating system); alternatively, a descriptive sensory analysis is used.

Descriptive sensory analysis has several functions in the QC and QA disciplines. The main areas considered here are:

- compliance with descriptive specifications
- troubleshooting and complaint handling
- shelf-life
- competitive benchmarking.

Shelf-life testing has been described earlier and will not be further expanded on in this section.

18.4.1 Application of Descriptive Sensory Analysis for Complying with Descriptive Specifications

The use of descriptive specifications (Muñoz et al. 1992) plays a vital part in ensuring product quality. Specifications provide details on the product's sensory characteristics in terms of appearance, aroma, flavour, texture and after-effects and can be semi-quantitative (see Figure 18.3 for an example) or fully quantitative (see Figure 18.4 for an example). Descriptive specifications are built around attributes known to affect consumer acceptability. Compliance with specification is done by a panel of trained assessors, for example by selecting attributes present in evaluated samples or scoring intensities of selected attributes (Rogers 2010).

18.4.2 Application of Descriptive Sensory Analysis for Troubleshooting and Complaint Handling in Food and Drink Quality Control and Quality Assurance

Upon discovery of an inferior product quality, through internal checks/observation or external complaints, it is imperative to identify the root cause and implement corrective actions. Any sensory evaluation involving potentially tainted food or drinks should first undergo a safety check. Food or drink samples returned by customers should be evaluated carefully and opened packs should never be tasted.

Investigations of the food or drink often begin with an evaluation by key experts, who could include employees from QC, QA, production, technical, research and development departments, product managers, etc. This group could establish if any obvious quality differences are present and possibly already indicate the root cause of the problem. In many cases, the cause is not instantly recognized and further investigation is required. At this point, descriptive sensory analysis can be applied to obtain information about sensory characteristics, which could give an indication of the types of components causing the quality issue, which in turn would assist in finding the root cause. If chemical or instrumental analysis is also conducted, correlation between the descriptive sensory analysis and the chemical or instrumental analysis could provide invaluable information in finding the root cause of the problem.

Product: **Fruit Drink 1**
Pack Type and size: **380 ml PET bottle**
This product must be free from off-odours, taints and foreign particles

Appearance
Evaluated by looking at the product in a clear sampling cup under artificial daylight before and after swirling.
A very strong orange coloured liquid that is bright, clear and still. The liquid appears thin and does not leave a residue upon swirling

Aroma
Evaluated by smelling from the sampling cup before and after swirling.
A moderate lemon and moderate orange aroma. A slight aroma of vitamin C tablets and slightly floral.

Flavour
Evaluated by taking sips from the sampling cup.
A moderate lemon flavour [sour lemon, like fresh lemons], moderate orange flavour [bitter orange, like Seville oranges] that is also moderately acidic [basic taste of citric acid solution] and moderately sweet [basic taste of sucrose solution]. A slight flavour of vitamin C tablets [like Haliborange] and also slightly bitter [basic taste of caffeine solution] in flavour

Mouthfeel/texture
Evaluated at the same time as the flavour and by taking more sips from the sampling cup.
Slightly drying mouthfeel. The liquid feels slightly thicker than water in the mouth.

Aftertaste/Afterfeel
Evaluated after swallowing – no extra sips taken.
A moderate citrus aftertaste [tangy flavour of general citrus fruits such as lemon, lime and orange] that is also moderately acidic and moderately bitter. Moderately drying and a moderate teeth/mouthcoating afterfeel. A slight sweet aftertaste.

Figure 18.3 Example specification for a fruit drink with examples of flavour/aftertaste glossary terms. Source: Adapted from Rogers (2010).

Product 7, Batch 123
Example data and specification (used for decision making and not seen by assessors)

Attribute specification	Results for sample 123	Sensory
Appearance: colour intensity	5.2	4.5–6.0
Appearance: brightness	7.0	6.5–9.5
Aroma: lemon	3.0	2.0–4.0
Aroma: orange	4.3	4.0–6.0
Aroma: vitamin C tablet	1.5	0.0–1.5
Flavour: lemon	3.0	2.0–4.0
Flavour: orange	6.1	5.0–7.0
Flavour: vitamin C tablet	1.0	0.0–1.5
Flavour: acidic	3.0	2.5–4.5
Taste: sweet	8.2	7.5–8.5

Figure 18.4 Example data and specification for fully quantitative descriptive specifications. *Source:* Adapted from Rogers (2010).

18.4.3 Application of Descriptive Sensory Analysis for Competitive Benchmarking in Food and Drink Quality Control and Quality Assurance

Understanding the competitor's product position is crucial in a competitive product environment. As a food and drink manufacturer, it is important to understand consumer preference levels for own and competitor products and the reasoning behind this. In addition, understanding the product characteristics of both own and competitor products will allow a deeper understanding of product behaviour, this being an advantage when changes to process or ingredients are required, when issues arise or when further knowledge of the product is required, for example, how to extend shelf-life or when creating new variants.

18.4.3.1 Applying Descriptive Sensory Analysis to Better Understand Consumers' Preferences

Consumers are very capable of indicating which foods and drinks they like and which they do not like. Using consumer product benchmarking with a group of consumers, liking can be established. Detailed reasoning beyond liking is more difficult to achieve. Some specific techniques have been developed to understand consumer preferences for the detail of product characteristics, such as free choice profiling (Lawless & Heymann 1999; Williams 1975), check-all-that-apply testing (CATA) (Ares et al. 2011) and ideal profile methodology (OP&P 2007). Another powerful approach to gain detailed insight into liking of product attributes is to correlate the consumer research data with data obtained from descriptive sensory analysis. Figure 18.5 shows an example outcome for a preference map of bread products; Figure 18.5a shows the consumer (dots) preferences for each of the products (px) and Figure 18.5b shows the sensory attributes to describe the different products.

18.5 Considerations Relating to Specific Food and Drink Products

Whilst different methods and protocols are described in this book, the practicalities relating to the preparation, serving or characteristics of a product group may favour particular tests or protocols over others in order to obtain good-quality results.

18.5.1 Considerations Relating to Hot Served Products

For all foods and drinks presented hot or in any other form prepared, the same conditions should apply. The same brand, model, wattage/temperature/voltage should be used. All equipment should be calibrated frequently to ensure similar operation. Ideally, all samples should be prepared consistent with the available preparation devices. Temperatures should be monitored and recorded throughout the preparation and also when serving; a serving temperature for entrées would typically be 60–71 °C, meat and poultry at a minimum temperature of 60 °C. Holding times should be consistent for all samples within a test (ASTM 2010b).

18.5.1.1 Considerations Relating to Hot Served Drinks

For foods and drinks that are served hot and for which the aroma is a major factor, such as coffee and tea, the aroma components are usually evaluated first, prior to the appearance, as the aroma will change drastically within the first few minutes. The type of serving vessel used to present the product should also be considered, as this may affect the product's taste profile and product temperature. Normally, samples are to be covered between servings. Thermos flasks can be used to minimize temperature loss. Flasks used within a test should be the same make, volume and weight. Preparation method and serving temperature should also be kept consistent. A serving temperature of 58–60 °C is usually

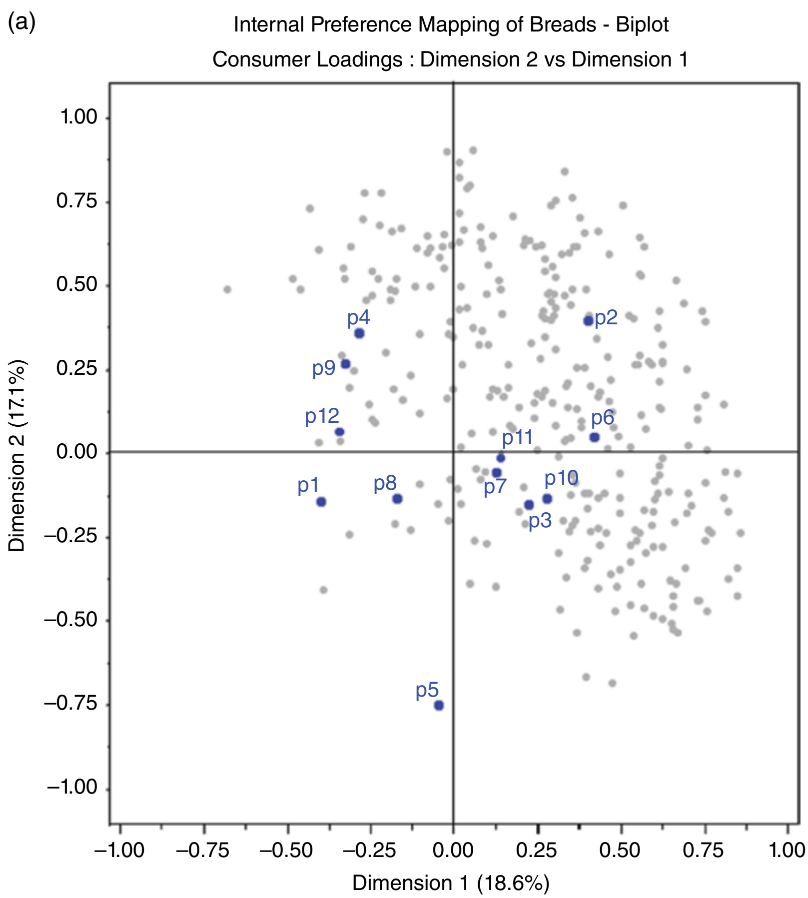


Figure 18.5 Internal preference mapping of breads.

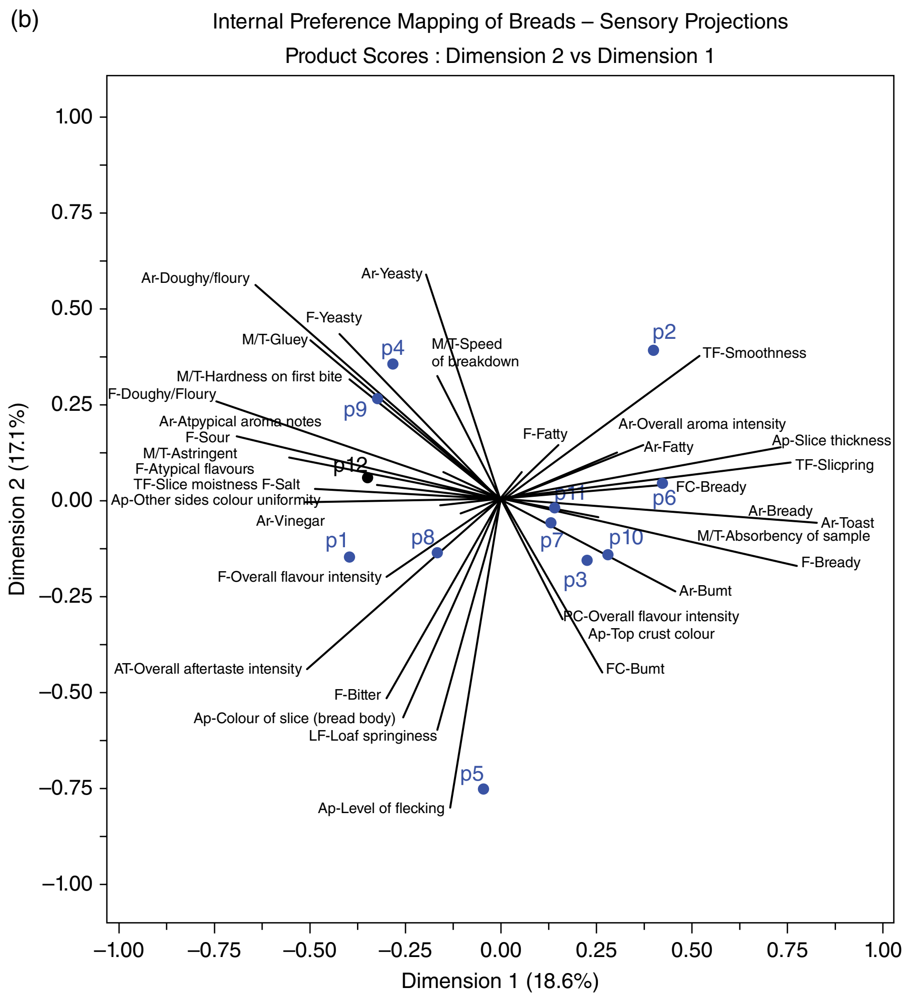


Figure 18.5 (Cont'd)

appropriate for hot beverages. Milk and sugar are often used in hedonic tests for coffee and/or tea. If used in a descriptive test, however, all samples should contain the same amount and types (ASTM 2010b).

18.5.2 Considerations Relating to Food and Drinks Containing Active Ingredients

Foods and drinks that contain active ingredients, such as throat lozenges and medication, should be tested cautiously. In some cases, the product can be tested without the active ingredient; in many others, however, the active ingredient is a major contributor to the sensory perception and hence should be included. Assessors will need to be advised of the active ingredients present in the samples that they are testing and testing may not be suitable for certain assessors (e.g. pregnant or breastfeeding women, assessors on specific medication or with certain allergies). Approval for testing may need to be obtained from an internal and/or external ethics committee and assessors may need to complete and sign a consent form. The amount of consumed material needs to be controlled and it may be favourable to use sip and spit methodology.

18.5.3 Considerations Relating to 'Non-Standard' Foods and Drinks

'Non-standard' foods and drinks are for the context of this section defined as products that are not intended for sale to the public, and that are being used for experimental or developmental purposes – and/or products being developed for a different market and that may have ingredients that are not (yet) approved for sale in the country of testing. Such products must undergo a risk assessment on a case-by-case basis and subjected to specific protocols. 'Non-standard' foods and drinks include novel foods, foods containing non-approved ingredients, foods produced using novel processes, ingredients not normally consumed unless incorporated into foods and foods containing pharmacologically active ingredients. All tests comprising 'non-standard' foods and drinks and the full protocols for testing are usually subject to internal and/or external ethics committee approval.

Potential adverse effects on the health of assessors should be minimized, specifically taking into consideration the following.

- Volunteer recruitment procedures should be designed to identify known health problems and allergies.
- Products should be microbiologically safe and, if necessary, the test products should be approved by a food microbiologist or should be subject to microbiological testing in cases of doubt. This is particularly important for shelf-life and accelerated shelf-life testing.
- Tests should be designed to minimize the amount consumed for health and nutritional reasons. In particular, tests on 'raw ingredients' should consider the risks of consumption above normal levels.
- Chronic effects on health should be considered, for example in long-term testing on products with specific ingredients. If appropriate, medical tests should be included as part of volunteer screening procedures. Records of consumption should be kept and health monitored on an ongoing basis.
- Observation of panellists for a specific time after testing of products can be considered with medical support.
- Testing of appearance and, if safe, odour only could be considered.

When testing non-standard foods, all assessors should be asked to sign a consent form where they declare full understanding of the study and provide consent for use of the results for the purposes stated. The consent form must contain a summary of any known and relevant compositional and toxicity data, an estimate of ingestion levels over the course of the project and any other information relevant to the study (European Commission Guidance 2012; Health Council of the Netherlands 2007; IFST 2010; Leatherhead Food Research 2011).

18.5.4 Considerations Relating to Food and Drinks Containing Alcohol

When testing foods and drinks containing alcohol, assessors should be informed that they contain alcohol. Usually consent forms would include level of alcohol in the products, total amount of alcohol served, legal (government) warnings, indication that testing is not appropriate for certain consumer groups, those individuals who can be excluded for testing (e.g. pregnant women, users of certain medications, current illness and alcohol abuse) and guidance for assessors if they feel intoxicated after consumption of the samples. Local regulations on the lawfulness of handling and serving alcoholic products may need to be checked, as well as the age of assessors to consume alcoholic samples. In some cases, proof of age needs to be obtained (ASTM 2010a).

Panel leaders should calculate and be aware of the amount of alcohol served to assessors. Cold alcohol beverages are usually served at 3–7 °C, whilst hot alcoholic beverages are commonly served at 66–71 °C. For some alcohol drinks, dilution with spring, demineralized or distilled water (for example, to a 50/50 dilution) of the sample is recommended to reduce the alcohol bite and burn, which may affect the perceived sensory characteristics. Glass is usually recommended for the serving of products. Special black glasses can be purchased if appropriate. In particular, carbonated alcoholic samples should be evaluated immediately after serving to reduce the changes in carbonation levels that will alter the sensory characteristics (ASTM 2010a).

The sip and spit method is commonly used for tests comprising food and drink samples containing alcohol. Care must be taken with regard to driving after testing alcoholic samples, for example transport could be arranged for assessors after the test. Some test centres have breathalysers available for assessors to check their alcohol consumption. It is recommended that testing ends 1 hour before leaving the premises and ideally does not take place at the end of the (working) day (ASTM 2010a).

18.5.5 Considerations Relating to Food and Drinks Served Frozen

Foods that are served frozen, such as ice cream, have additional complications regarding serving temperature and length of assessment. Once samples are starting to thaw, the product profile can change drastically. Portioning up of frozen samples may be difficult and special equipment (such as knives or cutting devices) can assist with equal presentation of sample portions. Scraping the sides of a container should be avoided. Coated products should be cut prior to tempering to reduce the possibility of collapse of the product structure.

Foods can be stored at slightly higher temperature just before serving to allow more convenient sample preparation and evaluation; however, for microbiological safety, foods should be held below 4 °C. Some frozen foods are designed to be consumed at a temperature higher than normal freezer temperature and need to be equilibrated at that temperature prior to testing. Standard serving temperatures range between −18 °C and −10 °C. Enough time should be allowed between samples to allow the assessors' mouths to return to normal body temperature.

Temperatures of all storage areas should be monitored closely and storage areas should be loaded properly to ensure adequate air circulation and protect all samples from odours of equipment and other products in the storage area.

If dry ice is used, appropriate safety precautions should be taken to avoid burns and any food that comes in contact with dry ice should be discarded (ASTM 2010b).

18.5.6 Considerations Relating to Fresh Produce

Fresh produce, or products containing fresh produce, such as fruits, vegetables, meat and fish, are often subject to differences due to seasons, harvest, weather, feeding regimes, etc. In addition, differences between samples from the same crop, batch or herd are common, as are differences within products. Assessors therefore often evaluate multiple samples, originating from several products, for example, slices from different cucumbers, or for meat and fish, scientists usually portion from the left and right side of the carcass in a balanced fashion.

Non-metal utensils for fruits and vegetables can be used to reduce browning. Fresh produce samples are usually prepared just before serving (up to 1 hour's prior evaluation) to maintain fresh appearance (ASTM 2010b).

18.5.7 Further Considerations and Practicalities Relating to Food and Drink Sensory Analysis

For testing of any products, microbiological safety should be established prior to the sensory evaluation. This would be particularly relevant during shelf-life testing or when samples are being transported and could have been stored under abnormal conditions, that may have contaminated the samples or accelerated the shelf-life (e.g. higher temperatures). For any product, consistency must be ensured throughout the procurement, storage, preparation, serving and handling processes. Measurements should be taken accurately and data should be recorded on all practices that may influence the end product. Change within samples whilst serving or holding should be considered, including carbonation levels, temperature changes, odour volatile loss, etc.

Sample handling by assessors should be consistent, where appropriate. For example, hard boiled sweets/mints can be held in the mouth for a set time period before evaluating. Typically 45–60 seconds would be appropriate with a time interval of 2–3 minutes between samples (ASTM 2010b). Appropriate length of time between samples should be considered for all products, as well as appropriate palate cleansers, as discussed earlier in this chapter.

Some complex food and drinks and food and drinks that are tested infrequently may be more difficult for assessors to evaluate without a prior training session. Including a labelled control sample in such tests could assist assessors with the evaluation of samples.

Foods and drinks that have a longer lasting taste profile or strong after-effects would ideally be tested with a method that allows for measurement over time, such as time intensity (ASTM 1998), progressive profiling (Jack et al 1994) or temporal dominance of sensations (Pineau et al. 2009). Some product categories have specific protocols that should be adhered to, including speciality products (such as wine and champagne) and samples with strong characteristics (e.g. very spicy products). For the latter product group, the Scoville Scale (British Standard 1989) was specially developed. This applies labelled magnitude scaling for the testing of the 'heat' of products.

18.6 Case Studies

18.6.1 Case Study 1: Improving the Sensory Characteristics of a Chocolate Bar

A chocolate manufacturer, ChocBloc, was looking to supply a certain retailer, Foods-to-go. Even though ChocBloc does have several products on the shelf, before Foods-to-go could be approached, ChocBloc had to ensure that the chocolate bars were in line with the consumer demands of this retailer.

18.6.1.1 Products

To evaluate consumer liking towards the ChocBloc samples, two samples originating from ChocBloc were compared with six competitor chocolate bars. The two samples originating from ChocBloc were both current samples supplied to other retailers.

One of the challenges with the products was the appearance of a logo on some of the samples. Logos/branding on products may bias consumers in their scoring. Therefore, from each of the chocolates, a very thin layer was removed from the top to eliminate this bias.

18.6.1.2 Consumer Acceptability Testing

Consumer acceptability testing was carried out with a group of consumers in hall settings. Testing was carried out at two locations, northern and southern, where respectively 153 and 156 consumers took part. Testing was carried out in groups of 50–55 consumers per session. Testing was carried out on tablets (iPads). A panel host and four servants/assistants were available during all sessions. The panel host explained the test procedures at the start of the sessions, answered queries from consumers and identified to consumers when testing was in progress and when consumers could take a break from testing. Servants/ assistants ensured that the correct samples were given to consumers at the right time and also helped answer simple questions from consumers (e.g. related to tablet use).

The consumers had water biscuits and water available and were encouraged to use those as palate cleansers between the chocolate samples. A 5-minute break was enforced after every sample, and a 15-minute break for the full panel took place midway through the test. During the 15 minutes, consumers were allowed to chat and/or spend time reading books/magazines, etc. During testing, no speaking with other consumers was allowed. All samples were provided to consumers randomized in a sequential monadic design to reduce bias.

The results were broken down by retail loyalty, age, gender and area of living. The results showed that the product could be improved further, as consumers, particular those shopping at Foods-to-go, showed a preference for competitor products.

18.6.1.3 Quantitative Descriptive Analysis Based Testing

Whilst consumers were carrying out the above detailed consumer acceptability testing, trained sensory assessors carried out quantitative descriptive analysis based sensory descriptive analysis on all products to identify the sensorial characteristics of each of the eight different chocolate bars.

Fourteen assessors drawn from the trained sensory panel at Leatherhead took part in this experiment. All assessors have been trained in the sensory evaluation of an extensive range of food and beverage products and were familiarized with the range of chocolate bar samples in this test.

Assessors evaluated each of the samples independently and provided a detailed sensory description of the product. The panel then discussed their views and generated a consensus vocabulary, which consisted of specific attributes and definitions. The final glossary generated by the panel consisted of 33 attributes, comprising four attributes for appearance, four attributes for aroma, six attributes for texture, nine attributes for flavour and 10 attributes for after-effects. In addition, for each modality, the attribute ‘other’ was used, where assessors could evaluate the intensity from any other characteristics they perceived which were not identified in the common list. The use of the attribute ‘other’ helps to prevent dumping of sensory sensations at other attributes. The vocabulary can be seen in Table 18.2.

Training was carried out after glossary formation to ensure that assessors had experience scoring the products using the generated glossary. Reference samples, such as xylitol for cooling effect and sauce flour for powdery feel, were presented to assist with this training.

Table 18.2 Vocabulary.

Appearance		
Colour	Brown chocolate colour intensity	Light-Dark
Gloss	Light scattering on surface	Not-Very
Evenness surface	The presence of pits on the surface	Not-Very
Air	The presence of air after breaking the chocolate	Not-Very
Other		Not-Very
Aroma		
Overall aroma intensity	Total aroma strength	Weak-Strong
Cocoa	Aroma of cocoa powder	Not-Very
Stale	Musty, cardboard	Not-Very
Off	Atypical aroma, e.g. rubber	Not-Very
Other		Not-Very
Texture		
Hardness on first bite	Force required to shear sample	Soft-Hard
Initial smoothness	Initial smoothness assessed when tongue yields chocolate	Not-Very
Brittle	Way sample fractures into small hard-edged pieces (peanut brittle)	Not-Very
Cooling	Effect similar to xylitol powder	Not-Very
Powdery	Chalky feel, e.g. sauce flour	Not-Very
Particle size	Size of particle	Not-Large
Other		Not-Very
Flavour		
Overall flavour intensity	Total flavour strength	Weak-Strong
Cocoa	Flavour of cocoa powder	Not-Very
Nutty	Flavour of almonds/nut skins	Not-Very
Sweet	Sweet basic taste	Not-Very
Bitter	Basic taste (like cocoa powder)	Not-Very
Stale	Musty, cardboard	Not-Very
Animal fat	Flavour of animal fat	Not-Very
Mushroom	Flavour of mushrooms	Not-Very
Off	Atypical flavour, e.g. rubber	Not-Very
Other		Not-Very
After-effects		
Cocoa	Cocoa powder aftertaste	Not-Very
Sweet	Sweet aftertaste	Not-Very
Bitter	Bitter aftertaste	Not-Very
Stale	Musty, cardboard	Not-Very
Mushroom	Aftertaste of mushrooms	Not-Very
Off	Atypical aftertaste, e.g. rubber	Not-Very
Cooling	Sensation similar to the after-effect of xylitol	Not-Very
Powdery	Powdery feel like cocoa powder	Not-Very
Astringent	Drying effect	Not-Very
Irritant	Burning effect	Not-Very
Other		Not-Very

All chocolate bar samples were tested in triplicate, so all assessors evaluated all eight chocolate bar samples three times. The samples were presented in a balanced presentation order.

Each assessor carried out individual evaluations of all the eight samples, which were presented in a sequential monadic randomized fashion.

- Packaging of the samples was removed immediately before the evaluation.
- All samples were presented to the assessors on plates labelled with three-digit coding.
- All samples were tested in triplicate.
- All assessments were carried out in individual sensory booths at a sensory facility.
- Assessors used line scales, varying from 0–100, to indicate intensity of attributes.

Sensory data were collected using a computerized acquisition system, Compusense® five v5.2.

During assessment, assessors evaluated a maximum of three samples in a sample set. Between each sample, assessors had a 5-minute break, where they refreshed their palates with water biscuits and filtered Vivreau water, whilst staying in their individual booths. After each three samples, the assessors gathered in a common panel room where they had a longer break.

The data from each assessor were pooled and analysed to determine the differences between the samples. The data analysis was carried out using ANOVA and Fisher's least significant difference (LSD) for multiple comparisons between samples. A significance level of 5% (95% confidence) was applied for measurement of statistical significant differences between samples. Data analysis was carried out using SenPAQ version 3.9.

Analysis of the relationship between the samples and the attributes was carried out by pooling the sensory data and using the PCA multivariate statistical method.These data were analysed using XLSTAT 2009.2.02.

Further analysis to explore consumer acceptability towards the chocolate bars and correlate this with the intensity of sensory characteristics was carried out using the multivariate preference mapping method and a technique to analyse the drivers for liking.

ChocBloc appeared particularly sweet, nutty and powdery compared to the other chocolate bar samples tested, as shown in Figure 18.6. Correlating the tests using preference mapping and identifying the drivers for liking, it appeared that the less favourable ChocBloc product was also described as having a mushroom note, a characteristic negatively correlated with the liking of the chocolate bars by Foods-to-go shoppers. Tables 18.3 and 18.4 show the drivers for liking.

This process helped ChocBloc to improve the product. After developing new prototypes, the samples were once again tested using QDA-based testing and the results built a great case to present the newly developed chocolate products to Foods-to-go.

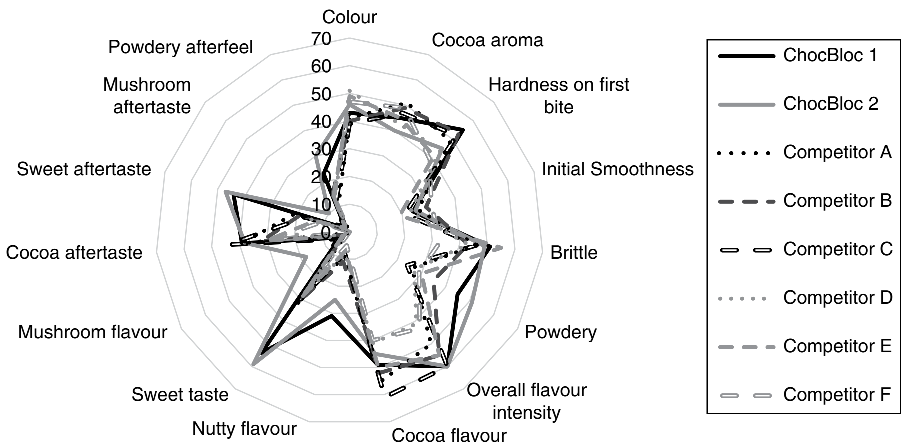


Figure 18.6 Spider chart showing statistically significant attributes for ChocBloc case study.

18.6.2 Case Study 2: Finding the Root Cause of an Undesirable Flavour in Water

This study was undertaken to study an off-flavour which was present in a bottled water product. Although the off-flavour was not unpleasant *per se*, it was undesirable as the product was meant to be clean mineral water. This case study shows the advantage of using descriptive sensory analysis as an aid for troubleshooting and pretesting for chemical flavour analysis.

Several samples of the bottled water were taken from different batches and sent to the sensory research lab for testing. Initial analysis showed that the samples had varying levels of flavour present.

Free choice profiling (Lawless & Heymann 1999; Williams 1975) was then applied to identify differences between the samples with the aim of finding the root cause of the off-flavour. Using generalized Procrustes analysis (Meullenet et al. 2007; Williams & Langron 1984), the profiles of each of the water samples were compared. The samples affected were described as floral, rosy, woody and minty.

The flavour profile identified was used by scientists employing gas chromatography-mass spectrometry to identify the chemical causing the off-odour as beta-damascenone, a component with a very low sensory threshold and thus difficult to measure using instrumental techniques.

The mineral water manufacturer sent in several components and packaging materials from the factory and the culprit appeared to be a rubber material on the production line that had also been in contact with raspberry-flavoured drinks.

Table 18.3 Drivers for liking – ChocBloc case study.

	ChocBloc1	ChocBloc2	Competitor A	Competitor B	Competitor C	Competitor D	Competitor E	Competitor F
Colour	0.51	0.53	0.49	0.46	0.45	**0.88**	**0.72**	0.45
Cocoa aroma	0.45	0.40	**0.71**	0.05	**0.79**	0.58	**0.92**	0.58
Cocoa flavour	0.61	0.15	**0.70**	**0.68**	0.66	0.64	0.61	**0.68**
Nutty flavour	**0.89**	**0.86**	0.29	0.31	0.28	0.28	0.27	0.25
Sweet taste	0.63	0.20	0.64	0.18	**0.79**	0.64	0.54	**0.71**
Mushroom flavour	0.02	**–0.68**	0.00	0.01	–0.02	–0.01	0.01	0.00
Mushroom aftertaste	0.01	–0.60	0.00	0.00	–0.01	–0.01	0.01	0.00

Correlation coefficients $P < 0.1$ shown in bold.

Table 18.4 Drivers for liking, summary all shoppers and Foods-to-go shoppers – ChocBloc case study.

		All shoppers	Foods-to-go shoppers
Appearance	Colour		
Aroma	Cocoa		
Flavour	Cocoa		
	Nutty		
	Sweet		
	Mushroom		
Aftertaste	Mushroom		

key:

Positive correlations

Negative correlations

18.6.3 Case Study 3: Descriptive Sensory Analysis in QC Testing

To measure whether drinks produced by the Refreshed beverage company were in line with consumer expectations, Refreshed applied a simple in/out method to release drinks to their retailer clients. Using this method, three screened assessors with limited training in the product specifications and some knowledge regarding acceptable and unacceptable product standards taste the drinks and identify whether the samples are generally within or outside the product's specification. With increasing demands from Refreshed's clients, but also an increased desire to obtain more knowledge on their products, Refreshed looked at alternative approaches to the in/out method with the aim of finding a simple method that provided more in-depth information.

18.6.3.1 Panel

One of the main considerations that Refreshed had to take on board was their current panel make-up, consisting of internal staff members who were required to carry out short sensory evaluations at short notice. This immediately flagged up another important aspect: staff time. Any testing carried out could not take up too much time from employees. There was the opportunity to grow the panel with a few additional staff members who had joined the company since the last sensory screening sessions.

18.6.3.2 Environment

The sensory testing environment was a room in the corner of the production area. The room was separated by walls but did not have a strong odour extraction system. Ideally, the room would be given an extraction area or located a little further away from the production lines. As a first practical measure, the room was fitted with a self-closing door with a rubber frame, to allow fewer odours to come in directly from the factory environment. The room had a table with four chairs. To remove bias and increase ability to focus on testing, simple collapsible booths were made by Refreshed's own maintenance team.

18.6.3.3 Testing

After considering the requirements of clients and practicalities within the company, it was decided to implement a two-step approach for all beverage releases.

1. Drinks were tested by three assessors, who identified general difference from standard drinks by direct comparison with this product and differences on 3–5 main attributes, which for most drink variants included sweetness, acidity and flavour intensity. For sparkling products, this also included carbonation level, and some specific products had attributes more tailored towards the drink. For each of the attributes, the three assessors indicated on a category scale whether more or less of the attribute was present in the produced sample compared to the standard available.
2. Any samples identified by one or more assessors in step 1 to be more than 1 point different from the standard were subjected to a simple descriptive sensory test. The test created for these occasions by Refreshed used assessors, selected from a pool of 16 assessors, trained in more depth on the sensory characteristics of Refreshed's products. Each product had a standard list of 10–14 sensory characteristics. Assessors scored the intensity of those characteristics in a standard and newly produced sample in duplicate on a nine-point scale. The samples were presented randomized and blind with three-digit coding.

After testing, the QC manager reviewed the test results and calculated mean scores and significant differences between the standard and the new sample. Depending on the results, the sample was released or put on hold.

18.7 Summary

The application of descriptive analysis is a great way to identify and quantify sensory characteristics of food and drink products. The availability of traditional methods, newly introduced rapid methods and methods that can be used with non-trained assessors gives sensory scientists a wide range of options to measure sensory characteristics. To ensure optimal results, protocols suitable to the test and the product should be adhered to, as described in this chapter.

18.8 Future Developments

Sensory science in general has seen an increase in popularity over recent years, where companies often commence with internal panels for QC checks and discrimination testing. The popularity of sensory descriptive analysis is also rising, particularly for companies operating external panels, but also using internally screened assessors. In general, food producers are starting to use descriptive sensory analysis as a tool to gain more in-depth product understanding for marketing purposes – including sensory claim substantiation and technical product knowledge building.

Over recent years, more quick and temporal methods have been introduced, a trend which is likely to continue for some years. The newer techniques are becoming a little more mainstream but more work is needed for food and drink producers to make the right choices about the most beneficial techniques to be used for their requirements.

Consumer research has also seen a steep increase, particularly due to the requirement of retailers to benchmark all new and current products. Many new techniques are also being developed to measure consumer liking, as well as different techniques where consumers are used to obtain data closely related to those more commonly obtained from descriptive sensory analysis using trained assessors.

It is anticipated that in the next few years, more companies will use external panels, which will be trained in one or more sensory descriptive analysis techniques. With more techniques being developed in different areas, we may see the following changes in the (near) future.

- Better understanding in correlating results with consumer insight studies.
- Deeper insight in understanding interaction of the senses better, in particular to develop healthy products.
- Easy to set up, cost-effective techniques for 'beginner panels'.
- In-depth understanding of the senses and genetics involved in sensory perception to correlate with nutritional food and beverage product benefits.
- Understanding and improving texture perception and food structure.
- Using sensory panels to estimate consumer liking and consumer panels to measure sensory perception is an area of active research.
- Testing in context – at the right location and moment.
- Testing using immersive techniques, video links and other interactive devices (VR, augmented reality, etc.).

References

Ares, G., Varela, P., Rado, G. & Gimenez, A. (2011) Are consumer profiling techniques equivalent for some product categories? The case of orange-flavoured powdered drinks. *International Journal of Food Science & Technology*, **46**(8), 1600–1608.

ASTM (1998) *E 1909-97. Standard Guide for Time-Intensity Evaluation of Sensory Attributes*. West Conshohocken: American Society for Testing and Materials.

ASTM (2005) *E 2454-05. Standard Guide for Sensory Evaluation Methods to Determine the Sensory Shelf Life of Consumer Products*. West Conshohocken: American Society for Testing and Materials.

ASTM (2010a) *E1879-00. Standard Guide for Sensory Evaluation of Beverages Containing Alcohol*. West Conshohocken: American Society for Testing and Materials.

ASTM (2010b) *E1871-10. Standard Guide for Serving Protocol for Sensory Evaluation of Foods and Beverages*. West Conshohocken: American Society for Testing and Materials.

Beeren, C.J.M. (2010) Establishing product sensory specifications. In: Kilcast, D. (ed.) *Sensory Analysis for Food and Beverage Quality Control*. Cambridge: Woodhead Publishing, pp. 75–96.

Beeren, C.J.M. (2011) Consumer Attitudes to Food Packaging. PACE Conference, Brussels, 11 February.

Brannan, G.D., Setser, C.S. & Kemp, K.E. (2001) Effectiveness of rinses in alleviating bitterness and astringency residuals in model solutions. *Journal of Sensory Studies*, **16**(3), 261–275.

British Standard (1989) *Methods of Test for Spices and Condiments. Determination of Scoville Index of Chillies*. London: British Standards Institute.

Earle, M., Earle, R. & Anderson, A. (2001) *Food Product Development*. Cambridge: Woodhead Publishing.

EC (2002) *Regulation (EC) No. 178/2002 of the European Parliament and of the Council of 28 January 2002 laying down the general principles and requirements of food law, establishing the European Food Safety Authority and laying down procedures in matters of food safety*. Official Journal of the European Union, L31, 1–24.

European Commission Guidance (2012) *Human Consumption to a Significant Degree, Information and Guidance Document*. Available at: http://ec.europa.eu/food/food/biotechnology/novelfood/documents/substantial_equivalenc_en.pdf (accessed 29 June 2017).

Findlay, C.J., Castura, J.S. & Lesschaeve, I. (2007) Feedback calibration: a training method for descriptive panels. *Food Quality & Preference*, **18**(2), 321–328.

Health Council of the Netherlands (2007) *The Safety Assessment of Novel Foods (2)*. Available at: www.gezondheidsraad.nl/sites/default/files/200723E_SANFoods2.pdf (accessed 20 June 2017).

Heckel, C.B. & Wilson, E. (2002) Characterization of fats and oils, specifications and technical bulletins. *Manufacturing Confectioner*, **82**(6), 43–49.

Hough, G. (2010) *Reduction of food waste through sensory shelf life (PowerPoint slides)*. Epsom: Leatherhead Food Research.

IFST (2010) *IFST Guidelines for Ethical and Professional Practices for the Sensory Analysis of Foods*. Available at: www.ifst.org/knowledge-centre-other-knowledge/ifst-guidelines-ethical-and-professional-practices-sensory (accessed 20 June 2017).

ISO (1988) *8589 Sensory Analysis – General Guidance for the Design of Test Rooms*. Geneva: International Organization for Standardization.

ISO (2012) *8586:2012(E). Sensory Analysis – General guidelines for the Selection, Training and Monitoring of Selected Assessors and Expert Assessors*. Geneva: International Organization for Standardization.

Jack, F.R., Piggott, J.R. & Paterson, A. (1994) Analysis of textural changes in hard cheese during mastication by progressive profiling. *Journal of Food Science*, **59**(3), 539–543.

Kilcast, D. (2000) Sensory evaluation methods for shelf life assessment. In: Subramaniam, P. & Kilcast, D. (eds) *The Stability and Shelf-Life of Food*. Cambridge: Woodhead Publishing, pp. 79–103.

Kilcast, D. & Subramaniam, P. (2000) *The Stability and Shelf-Life of Food*. Cambridge: Woodhead Publishing.

Lawless, H.T. & Heymann, H. (1999) *Sensory Evaluation of Food: Principles and Practices*. Gaithersburg: Aspen Publishers.

Leatherhead Food Research (2011) *Code of Practice for Studies Involving Human Volunteers*. Epsom: Leatherhead Food Research.

Lucak, C.L. & Delwiche, J.F. (2009) Efficacy of various palate cleansers with representative foods. *Chemosensory Perception*, **2**(1), 32–39.

Meullenet, J.F., Xiong, R. & Findlay, C. (2007) *Multivariate and Probabilistic Analyses of Sensory Science Problems*. Ames: Blackwell Publishing Professional.

Morley, W. & Osborn, S. (2012) *Food innovation; design, development, implementation, compliance (PowerPoint slides)*. Epsom: Leatherhead Food Research.

Muñoz, M., Civille, G.V. & Carr, B.T. (1992) *Sensory Evaluation in Quality Control*. New York: Van Nostrand Reinhold

Narain, C., Lawson, S.C. & Beeren, C. (2010) Determining Effective Palate Cleansers for Use with High Intensity Sweeteners. Eurosense Poster presentation, 5-8 September, Vitoria-Gasteiz, Spain.

Nasrawi, C.W. & Pangborn, R.M. (1989) The influence of tastants on oral irritation by capsaicin. *Journal of Sensory Studies*, **3**, 287–294.

OP&P Product Research (2007) Consumer guided product development – the ideal profile method (PowerPoint slides). SPISE 2007 Symposium, Ho Chi Minh City, 26–27 July. Available at: www.opp.nl/artikelen/pdfs/12.pdf (accessed 20 June 2017).

Peryam, D.R. (1958) Sensory difference tests. *Food Technology*, **12**, 231–236.

Pineau, N., Schlich, P., Cordelle, S. et al. (2009) Temporal dominance of sensations: construction of the TDS curves and comparison with time-intensity. *Food Quality & Preference*, **20**(6), 450–455.

Powers, J.J. (1988) Current practices and application of descriptive methods. In: Piggott, J.R. (ed.) *Sensory Analysis of Foods*, 2nd edn. London: Elsevier, pp. 187–266.

Rogers, L.L. (2010) Sensory methods for quality control. In: Kilcast, D. (ed.) *Sensory Analysis for Food and Beverage Quality Control*. Cambridge: Woodhead Publishing, pp. 51–74.

Staniforth, J. (2004) A question of consistency and quality. *New Food*, **2**, 20–22.

TSO (2014) *Food Information Regulations 2014*. SI 2014 No. 1855. London: TSO.

Williams, A.A. (1975) Development of a vocabulary and profile assessment method for evaluating flavor contribution of cider and perry aroma constituents. *Journal of the Science of Food & Agriculture*, **26**, 567–582.

Williams, A.A. & Langron, S.P. (1984) The use of free-choice profiling for the evaluation of commercial ports. *Journal of the Science of Food & Agriculture*, **35**, 558–568.

Word, K.M. & Gress, H.S. (1981) Selection, training and motivation of sensory panel judges. *Technical Quarterly, Master Brewers Association of the Americas*, **18**(1), 22–25.

Zook, A. & Wessman, C. (1997) The selection and use of judges for descriptive panels. In: Gacula, M.C. Jr (ed.) *Descriptive Sensory Analysis in Practice*. Connecticut: Food & Nutrition Press, pp. 35–50.

CHAPTER 19

Application of Descriptive Analysis to Non-Food Products

Anne Churchill and Ruth Greenaway

19.1 Introduction

In sensory analysis of food products, the primary focus is on sensations in the mouth. With non-food products, other sensations not perceived in the mouth could also be important. Our senses are specialized to give us information about various aspects of our environment. Hearing and smell are considered to be early warning systems and can provide us with information about things happening a distance from us. Touch and taste inform us about what is happening on the body surface. Touch and smell are important in governing our emotions, and our visual sense is considered our most dominant sense, with nearly half of the adult cortex (outer layer of the brain) involved in some way or another in visual processing of information.

Different parts of the cerebral cortex are responsible for processing the sensory signals from our eyes, ears, skin, mouth and nose. Messages from each of these primary sensory areas are subsequently sent to other parts of the brain where signals from each of our senses are merged to form one multisensory experience. The integration of all this information provides a holistic perception of our environment and of the products we use. It is this fact that complicates the design of sensory experiments to assess non-food products. The protocol for the preparation of samples, the environment in which they are tested and the test design (including whether one, more than one or all the senses are used) all influence the results obtained.

The sensory analyst's role is to measure the sensory properties of a product. A hypothesis needs to be formed and tested and then conclusions drawn from the test data. In order to do this, the sensory analyst has the task of understanding the key points of assessment for a consumer, the key attributes that need to be assessed and then establishing a test protocol that provides stimuli that mimic those experienced by the consumer as closely as possible. However, at the same time, he/she must control variables not under test as much as possible in order to

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

reduce variability, provide a reliable and reproducible test and make it possible to interpret the results. Inevitably, the analyst is faced with a compromise and it is his/her job to balance reliable measurability against reality.

It is this crucial step of defining key points of assessment and balancing laboratory control with real-life situations that is one of the major differences between testing food and non-food products. The assessment of food inevitably involves the process of eating a product. With non-food products, the range of different types of possible activities involved in assessment of a product is considerably greater. Examples will be given throughout this chapter that illustrate this, including using a skin cream where the assessment of the product is important not only during application but also over time, or hair care products where assessment is important during washing the hair, while drying it and over time after the process is complete. Typically, there may be a number of different time points during use of a product when assessments are carried out and often the in-use situation needs to be simulated as the real-life situation is difficult or expensive to use regularly with sufficient repeats to provide a good experimental design.

This chapter explores the measurement of sensory properties of non-food products and the decisions and compromises that need to be made in order to design a test that enables measurement of the sensory stimuli relevant during use of these products. More often than not, sensory protocols are first designed and fine-tuned for measurement of food products but the full range of principles and approaches can be adapted for testing with non-food products. Many of the design and statistical approaches are exactly the same irrespective of whether the test is designed for measuring how a product tastes, smells, feels, sounds or appears. However, as discussed above, there are some additional considerations not normally necessary in food analysis that will be highlighted in this chapter. The examples used in this chapter will not be fully comprehensive in covering all types of non-food products but the train of thought followed and discussed should help to resolve many situations encountered in other product areas.

19.2 Why is Descriptive Analysis of Non-Food Products Important?

The creative team or new product development team understands the current market, the production details and the technical advantages or disadvantages of one design over another in intimate detail but it is the end consumer's holistic perception of the product and its utility, appropriateness or subjectively preferred design features that will determine its success on the market. The descriptive sensory panel will provide an objective description of the product attributes that are not biased by any intimate knowledge of the product design features and as such more closely represent a consumer perception of the product. This provides a bridge of understanding between the product designer and the end consumer, giving valuable insights into how the consumer will perceive the product.

Common applications of descriptive analysis include the following.

- *Providing directions in research, product design or formulation*: for instance, a target consumer response may have been defined, such as a shampoo that leaves hair feeling more manageable or a toilet paper that feels softer, and confirmation is needed that new technical advances or product developments are providing perceivable benefits.
- *Market mapping or competitive assessment*: although knowledge may be available about the comparative success, or market share, of similar products on the market, defining the reasons for this success may not be so straightforward. Descriptive sensory analysis of the various market products provides insights into the key attributes defining their similarities and differences. For instance, descriptive analysis of toilet paper may define the smoothness of the surface, the force needed to compress the paper, the strength of the paper, its thickness, its softness and other parameters. When comparing the relative success of the same products on the market, it may become apparent that one or two of these attributes are related closely to the overall success of the product. However, although potentially important, these attributes are probably not the only drivers of a product's success and should be considered alongside factors such as advertising, market promotions, price and consumer loyalty.
- *Modification of current products*: material or ingredient substitution or other product modification may be necessary for legislative, safety, cost or environmental reasons. Where a loyal customer base is already in place, it is usually preferable for any changes to be imperceptible or of no consequence. If a trained sensory panel cannot detect differences it is unlikely that an untrained consumer will be able to. For instance, changes to a product base such as a skin cream may affect the overall perception of the fragrance of the product. The perfumery house supplying the fragrance will modify it to make sure that the overall product still smells the same. The success of this modification will be confirmed if a trained sensory panel completes a discrimination test and is unable to detect a difference. If differences from the original product can be perceived, the panellists will be able to describe the nature of those differences.
- *Quality control*: routine quality checks on production lines, usually in comparison with a standard control and settings defining acceptable deviations from the standard, may utilize a trained panel. Current thinking (Muñoz 2013) is that a limited number of key parameters should be selected and measured by a trained panel of assessors.
- *Claim substantiation*: the American Society for Testing and Materials (ASTM) provides a guide for use of sensory analysis to support claims (ASTM 2016) but in addition, professional legal advice and support should always be used. Every care should be taken to ensure that the test design is fair and stands up to the scrutiny of sensory and legal professionals. Well-designed and tightly controlled sensory descriptive tests following industry standards and best practice can be used as objective data to support product claims.

19.3 General Considerations in Applying Descriptive Analysis to Non-Food Products

When designing any sensory measurement test, there are five key factors to be considered.

- Test objective
- Test design
- Number of samples tested per session
- Panellists
- Statistical analysis

19.3.1 Test Objective

What kind of sensory test? It is worth devoting time and thought to getting the question right in the first place. Sometimes what may seem obvious is not as simple as it first appears. Is a full descriptive test what is really required or could the question be more simply answered by measuring one attribute alone? Measuring a whole string of product attributes may seem to give a lot of information but if the attributes are not discriminating between products, they are not of value or if they are all highly correlated with each other then one attribute will tell the same story as 20. On the other hand, if a full product description is asked for and the attributes given do not cover all product characteristics, panellists may be forced to 'dump' or use another attribute to allow differences detected to be described. It is necessary to spend time researching what the real question is and think about the best test to use in order to answer the question.

19.3.1.1 Assessment Protocol

Decisions need to be made about how to prepare and present the samples ready for assessment and to determine the points or times of assessment. A series of questions need to be considered.

- *What do we know about how the product is used?* Cultural differences or differently learnt habits may define entirely different modes of product usage which in turn may dramatically affect how the product is perceived. For instance, washing dishes in countries where water is scarce follows an entirely different routine to that used in countries where water is freely available. This difference in routine could strongly influence the perception of the product performance.
- *Do we have consumer insights or on-pack instructions?* Focus groups and consumer usage and attitude surveys provide valuable information and may highlight that consumers do not follow pack instructions. This is often the case with the amount of laundry detergent or laundry liquid that consumers use in their washing machines. There is evidence that consumers believe overdosing provides a cleaning benefit.

- *Is there just one point of assessment or several?* There may be one point of product usage that drives all consumer perceptions of the product or there may be several that are equally important. For instance, when using a surface cleaning product, the ease of use and the feel of removing stickiness or dust and dirt from a surface may be most important, but after use the fresh smell of the product in the room may also enhance the cleaning perception.
- *Do we know which point of assessment is the most important?* For example, as a tube of hand cream is first opened and the product is seen for the first time; as it is squeezed out of the tube and the way that it flows is observed; or as it is rubbed into the skin and the ease with which it is rubbed in is perceived; or half an hour after it is applied and the suppleness of the skin can be felt. If the relative importance of each of these stages is known, it may be possible to measure just one or two in order to gain a good understanding of perception of the product.
- *Is timing crucial or can the product be left standing for some time without it changing?* For instance, the properties of a wooden chair will not change if it is left standing in an assessment room for half an hour, but a hot bowl of diluted washing-up liquid will.
- *Can we simulate the real in-use process?* Can the in-use habits that have been identified be adapted to a laboratory environment? If not, what compromises are acceptable? It is necessary to understand how a change in circumstances will change the perception of the product. For example, it is possible to simulate hair washing on a small hair switch in the laboratory but does the absence of the steam and warmth of a shower environment considerably change the perception of the performance of the product?

19.3.2 Test Design

A relatively quick search of the scientific literature soon reveals a substantial source of information that teaches how easily bias can be introduced into a sensory test. Standard sensory experimental design considerations must be applied in non-food sensory tests in exactly the same way as would be appropriate for food tests. This is often more complicated for non-food products as different stages of use and often long assessment periods (e.g. assessments at intervals over a period of several months in the case of long-lasting fragrance on washed cloth) may be involved. Examples of these considerations are given in the product-specific section later in this chapter, but they could include the following.

- Ensure that all samples have consistent characteristics or that differences are masked; colour, size, smell and shape should be standardized unless these are the characteristics under assessment.
- All samples must be prepared in exactly the same way, standardizing age or preparation method.
- Control the testing environment so that panellists are comfortable and free from distractions during assessments, and also to ensure privacy and confidentiality (e.g. when showering in a test environment).

- The environment may also need to be controlled for temperature, humidity and air flow, depending on the product under assessment. Lighting quality and uniformity are important particularly when visual assessments are made, while background odour is important when odour assessments are key.
- Methods of handling, smelling, viewing or listening to products must be defined in detail so that all panellists assess the products using exactly the same approach. This will limit variability in the data but also defines a protocol that needs to be clearly described when reporting the resulting data.
- Train panellists so that they are familiar with the tests conducted and are not distracted by personal preferences.
- Avoid order effects such as carry-over or halo, where the perception of one sample may be modified by the characteristics of the previous sample, by using predetermined randomized sample presentation orders which are balanced and complete as far as possible.
- Avoid contrast or convergence effects where perceived differences between samples can be reduced if some very extreme samples are included or differences may be exaggerated if the samples in a test set are very similar.
- Use accepted statistical tests to confirm that any differences detected by the sensory panel are significant when adjusted for the variability and interactions in the data.

19.3.3 Number of Samples Tested Per Session

There are a significant number of product characteristics that limit the number of samples that can reasonably be tested in one sensory session. Some of these are discussed in the product-specific section, but could include the following.

- Odour assessments of particularly harsh odours (such as cleaning products containing bleach) or very faint odours or odours that panellists become adapted to very quickly. In all of these cases, the experience and testing trials carried out by the sensory scientist prior to panel testing will be crucial to determine how much rest panellists need between samples and how many samples they can reasonably assess in one session. Statistical tests should be carried out following a test to determine whether reliability is reduced over the time of the panel session.
- Where handling of products is required (e.g. to assess the feel of fabric), skin sensitivity may decrease over time or there may be a build-up of deposits on the skin that influence perception. There will be a limit to how many times in a session the skin can be cleansed or rested in order to maintain levels of perception.
- Where products are applied to the skin, there is also a limit to how many products can be assessed in one session. More than a single use of one skin area is not usually advisable.

19.3.4 Panellists

Selection and training of panellists for non-food products follows the same course as for food products. They are selected for their interest, commitment and availability as well as for their sensitivity, team play and creativity. A screening questionnaire is recommended to determine availability and interest; it may also include questions that test the applicants' ability to use descriptive words and rating scales.

Where product smell is important, common anosmia should be screened for (such as ability to smell musk or woody notes) and any nasal deficiencies or allergies. It is also important to screen for skin allergies or other common sensitivities that may be relevant to the product testing. Where panellists need specific personal characteristics such as hair type or skin type (for testing skin creams or shampoos), then these questions may also need to be included. Applicants should then be screened on a number of tests that reflect the type of test that will be included in future sensory assessments and test their ability to discriminate and rank order different types of odour. Levels of difficulty should vary so that volunteers are engaged and find the tests interesting (not too intimidating) but so that those who are less capable will be identified and screened out.

Where colour or visual appearance is important, any deficiencies in colour detection should be screened for and screening may include description of complex or ambiguous visual stimuli. Similarly where touch or handle characteristics are key, screening should give applicants the opportunity to handle relevant samples and their ability to manipulate the products should be observed and used as screening criteria.

During the recruitment phase, candidates should be informed of the time commitments for training, the duration of panels and the need for a high level of personal commitment. This could include potential changes to personal habits such as not wearing perfume if assessing fragrance or not using hand moisturizing cream if assessing the feel of fabrics or paper.

Training to eliminate subjectivity is about developing skills to discriminate and reproduce assessments that are coherent with the rest of the panel and also about developing familiarity with unusual sensory experiences, and potentially very rigidly defined handling or assessment methods. For instance, in the fragrance industry, an important part of panellists' screening and training is an introduction to malodours. Panellists must be able to discriminate, rank and score perceived intensity of unpleasant odours as well as pleasant odours objectively. For the assessment of fabric handle or the tactile properties of hair or skin creams, panellists need to be introduced to very strict protocols for handling the products. Very precise protocols are necessary to define exactly how the product is manipulated, including applying equal degrees of movement and pressure, using the same fingers or area of skin each time. Practice and feedback in every case (smell, touch, sound or vision) enables consistent, reproducible judgements.

During training, panellists should be introduced to the general concept of descriptive testing. Attributes should be discussed and where possible standards assessed and compared until panellists are clear about the meaning of the descriptive language and can use it consistently.

Informed written consent must be obtained from all assessors prior to their participation. This should cover the basic elements of informed consent including the nature of the experiment (sample information, associated risks), ability to withdraw from the study at any time and confidentiality. A useful checklist is given by Kemp et al. (2009).

19.3.5 Statistical Analysis

The rules applied to experimental design and statistical analysis in sensory analysis are the same whether we are considering food or non-food products. Standard sensory analysis texts should be referred to.

19.4 Development of Descriptive Language

Absolutely crucial to any sensory descriptive test is the development of a descriptive language, or sensory attributes that are meaningful in consumer terms, clearly defined for the assessors and also actionable for the product developers or research team.

Pharmaceutical and cosmetic chemists are well aware of the physicochemical properties of cosmetic creams and lotions, experts from the textile industry are familiar with the parameters used for measuring the physical, chemical and mechanical properties of fabrics, and similarly experts in all industries will understand the technical language meaningful to others in their field.

Consumers can recognize differences between products but they often have difficulty describing the attributes which differentiate two products from each other and this information may be difficult for the product developer to understand (Close et al. 1982). The trained sensory panel fills the gap between product development and consumers. It has the potential to provide quantified descriptions of products which help in understanding why consumers prefer one product to another and guide formulators in their development work.

The language used in a descriptive test must be meaningful to those who use the data (e.g. product developers) and, assuming that it is important to understand how consumers react to the product, it must cover the range of attributes that determine how consumers differentiate between products. Probably most important is that it must be well defined so that the sensory panel can use it in a consistent and reproducible way. The approach taken by Meilgaard et al. (1991) is that panellists should be given a sound understanding of the real chemical and physiological properties of a product and this along with given attributes and standard references which anchor the attribute scales provides reproducible 'absolute' scales. An alternative approach provides panellists with a series of relevant products to assess and then identify, define and agree attributes and reference standards (Aust et al. 1987). This involves the panel working together to develop a test vocabulary.

An approach described by Moskowitz (1982) used a repertory grid method (Kelly 1955) to stimulate panellists to compare pairs of products and explain how the samples were similar and how they were different from each other. This enabled panellists to generate terms that describe the products and could subsequently be used as descriptive attributes. This procedure leads to a wide variety of attributes and needs a structured approach to rationalizing the language to use in the final descriptive vocabulary.

The assumption being made in the approaches described above is that the small group of panellists will discriminate between products in a way that is representative of the way in which consumers, who will ultimately determine the success or failure of the product, would also discriminate.

An adaptation of these approaches was used by Churchill et al. (2009) to explore the textural language used by consumers to describe the feel of shampoo products and of their hair after use of the products. The same approach was also used by Givaudan to generate textural language to describe skin creams and is described here in detail to provide an example of this approach to language generation.

19.4.1 Language Development Case Study: Skin Creams

- *Objective*: to develop a tactile language for description of skin creams and lotions.
- *Sample population*: the test was carried out in the UK with a group of naive female consumers who were regular users of skin creams (i.e. used them at least three times a week).
- *Stimuli*: to identify a 'comprehensive' textural vocabulary, a wide range of products that exhibited as many different textural characteristics as possible were used. Ten products were used for this purpose.
- *Approach*: 75 women were recruited to use the products at home in pairs, using first one product and then the other on different days (fully rotated), and were provided with a diary to complete for each product. Descriptions and records were spontaneous (unprompted). In order to achieve an assessment on perceptions across the whole sensory experience, respondents were asked to record their impressions of each product, at three different usage occasions:
  1 feeling on hands when first removed from the container
  2 feeling on hands and skin when rubbing in
  3 feeling of skin after rubbing in.
- *Interviewing technique*: after each pair had been placed with the respondent for 2–3 days, an interviewer returned to complete the interview. An interviewing technique, based on Kelly's theory of personal constructs, was used to access the vocabulary relevant to consumers (Kelly 1955). This technique is based on the method of paired comparisons, where respondents compare and contrast products, deciding which they prefer and then discussing why. This enabled them to rationalize the reasons for choosing one product over another and helped them to better articulate positive and negative descriptive language.

- *Analysis*: the textural vocabulary used by the consumers was recorded and tabulated in order to identify the commonly used descriptions at each usage occasion and to group those words that were synonymous. This analysis allowed selection of vocabulary that covered the range of attributes important to consumers in describing the differences and similarities between products; by studying the context in which the attributes were used, descriptions of the meanings could also be extracted.
- *Outcome*: the resulting vocabulary is directly useful for future quantitative consumer tests. As it was generated spontaneously by consumers, it is clearly language that is meaningful to them. The definitions provided explain the descriptive language and provide examples of market products that illustrate the characteristics discussed. This provides a descriptive language for the sensory panel that can be discussed and clarified using the information given. Finally, these descriptions and further clarification by the trained panel will provide the product development experts with an interpretation of the attributes that will guide their work.

Table 19.1 gives the 18 most used descriptors and Table 19.2 gives the consumer comments relevant to four of these descriptors. Table 19.2 would be the start point for a sensory panel when developing a descriptive language.

Table 19.1 Consumer attributes used to describe textural attributes of skin creams.

	Descriptor	Feeling on hands when first removed from container	Feeling on hands and skin when rubbing in	Feeling on skin after rubbing in
1	Soft	✓✓	✓✓✓	✓✓✓✓
2	Smooth	✓✓	✓✓✓	✓✓✓
3	Greasy	✓✓	✓✓	✓✓
4	Silky	✓✓	✓✓	✓✓✓
5	Oily	✓✓	✓✓	✓✓
6	Creamy	✓✓✓	✓✓	✓
7	Sticky	✓✓	✓✓	✓✓
8	Cool	✓✓	✓✓	✓
9	Absorbent	✓✓	✓✓	✓
10	Light	✓✓	✓	✓
11	Thick	✓✓	✓	×
12	Runny	✓✓	✓	×
13	Watery	✓✓	✓	×
14	Tacky	×	✓	✓✓
15	Wet	✓	✓	✓
16	Slimy	✓	✓	✓
17	Rich	✓	✓	×
18	Velvety	✓	✓	✓

Key: The number of ticks indicates the frequency of use of that term (✓= low, ✓✓✓= high usage, X = not relevant for this stage of assessment).

Table 19.2 Examples of consumer 'definitions' of descriptive vocabulary.

Descriptor	Examples of Consumer description
Sticky	It felt sticky and tacky like cream I use for nappy rash
	It felt sticky like kids glue
Greasy	It had an oily feel, wet and greasy like tuna oil.
	It left my skin feeling greasy and oily as though it had a layer of oil on top of it.
	It felt like Vaseline, slimy, greasy, oily.
Smooth	It felt like velvet when rubbed over skin, velvet smooth and silky.
	It was smooth and gentle, felt velvety and smooth. Gave an impression of silky smooth skin.
Soft	My skin felt peachy soft, nourished, like rose petals. It made my skin feel like a baby's, very soft
	It felt like whipped cream, soft light and fluffy.

19.5 Considerations Relevant to Specific Non-Food Products

19.5.1 Cosmetics (Skin Cream, Make-Up, e.g. Lipstick, Mascara): Tactile Properties

The assessment of personal products brings special considerations specifically related to the fact that simulation of real in-use conditions requires the assessor to use the product on their own skin, face or hair. Although the detailed methods of application or manipulation and assessment of the product can be clearly defined, pre-treatment of the substrate can be standardized and panellists can be trained to carry out specific procedures, the variability from person to person determined by skin or hair type and individual metabolism cannot be controlled. The sensory analyst will need to determine whether a specific target group (e.g. age, skin or hair type or condition) should be selected for tests or whether a larger group of subjects is necessary in order to randomize for individual differences (the size of group needed will be determined by sample size calculations).

The details discussed below take the procedures for skin creams and lotions as an example.

19.5.1.1 Pre-treatment

Samples must be preconditioned before use. This should consist of storing the samples in stable temperature and humidity conditions, the same as those present in the test room, until the samples equilibrate to those conditions.

Skin pre-treatment should remove any previously used products from the surface of the skin to be used for application of the test product, and standardize the condition of the skin surface. This could be under supervised conditions carried out by trained operators or the panellists may pre-treat their own skin at home before the evaluation session, following strict procedures. A typical treatment would include a 1-minute wash with mild (non-perfumed and non-moisturizing) soap and a 15-minute dry-out period.

Prior to product application, the areas to be used or test sites should be marked on the panellists; often the inner face of the forearm is used. Using an appropriate skin marker (for instance, an eye brow pencil), an appropriate sized circle is marked, such as 2 × 2 inch (51 mm) diameter. The recommendation given in ASTM standard E1490 (ASTM 2011) is that these circles should be located 2 inches (51 mm) above the wrist and 2 inches (51 mm) below the elbow, and overlap, contaminating adjacent areas, must be avoided.

19.5.1.2 Language and Product Handling

The type of product under assessment and the project objectives will have an influence on the language developed and used for product evaluation. Details of language utilized for skin cream and lotion assessments are given in ASTM standard E1490 (ASTM 2011), by Wortel and Wiechers (2000) and in Meilgaard et al. (1991). These approaches all include language that considers the appearance of the product, its performance on skin before rubbing (pick-up), its performance on skin during rubbing (rub-out) and skin after-feel (immediate and over time).

During training, the panellists should be introduced to the need for strictly controlled procedures for the manipulation and application of samples and the detailed understanding of all of the sensory attributes at each stage. These could include the following.

- *Appearance*: the attributes of a measured amount of product delivered to a petri dish or onto the skin surface and assessed using vision, including integrity of shape and gloss (light reflected).
- *Pick-up*: assessment of a measured amount of product applied using a syringe or pipette ($0.63\,cm^3$) between the thumb and forefinger to assess firmness (force required to compress the product between thumb and forefinger), stickiness (force required to separate fingers), cohesiveness (amount sample strings rather than breaks when fingers are separated) and amount of peaking (degree to which product makes stiff peaks on finger tips).
- *Rub-out*: assessment of a measured amount of product applied to the skin ($0.63\,cm^3$) using a gentle circular motion, stroking at a rate of two strokes per second (rate controlled using a metronome), including wetness, spreadability, thickness, oiliness, waxiness, greasiness, absorbency.
- *After-feel*: assessing appearance and skin-feel including gloss (light reflected), stickiness (degree to which fingers adhere to the product), slipperiness (ease of moving fingers across the skin), amount of residue, type of residue (which could include oily, waxy, greasy, silicone (dry/slick), powdery and chalky).

Examples of reference standards and scale values for skin-feel attributes are provided in ASTM (2011).

19.5.1.3 Special Considerations

Skin temperature and humidity both of the applied skin surface and the finger surfaces used for manipulation will affect the sensitivity of sensory measurements and must be controlled. ASTM advice is that skin temperature measures should be made at the start of a test, and again if any change in room temperature and humidity is experienced, recorded and reported.

19.5.1.4 Testing Environment

A consistent light source is recommended for each assessor for evaluation of visual characteristics such as shine. It is important that all assessors receive the same amount of light on the test product or tested site (e.g. skin area) and the same angle and distance of light source.

Ambient temperature and humidity should be maintained at a constant and comfortable level.

19.5.2 Underarm Products (Antiperspirants and Deodorants, Roll-Ons, Sticks and Sprays)

Special consideration needs to be given to measuring the performance of underarm products as the specific environment in which they are used is unique (no other part of the body can be used as a proxy). The effect of previous use of antiperspirant products carries over on the skin for several days even with regular washing, individual skin microbiology varies from one individual to the next, recent consumption of highly flavoured foods modifies the odour of perspiration and individual metabolism will modify the amount of sweat produced and the temperature of the underarm area.

The key performance measure may be underarm odour, the appearance of the pack and ease of application of the product or the feel of the product on the skin. Comments made here refer to the measurement of underarm odour.

19.5.2.1 Pretreatment

In order to remove the carry-over effect of previously used underarm products, it is preferable that volunteer testers refrain from using antiperspirant or deodorant products for 2 weeks prior to testing. This does cause significant time and cost considerations in the testing programme so may often be compromised in practice. During this period, unfragranced soap should be provided where possible.

At the start of the test, the underarm area should be washed with unperfumed soap, rinsed and dried and then the product should be applied in a controlled and uniform manner. It is preferable for all volunteers in a trial to be treated by a trained technical assistant to reduce variability from subject to subject but if this is not possible, volunteers must be given clear instructions on how to apply the product (particularly controlling how much is applied).

During the trial, the volunteer subjects must follow strict instructions not to go swimming, not to wash under their underarms, not to apply any products to this area and not to eat any strong-flavoured foods.

19.5.2.2 Assessment

Product assessment may either take the form of self-assessment (i.e. the volunteers assess their own underarm area) or a trained panel assesses the underarms of a team of recruited volunteers. If a series of assessments at regular intervals are required, it often becomes very difficult to organize panellist assessments of volunteers unless both remain in the same building for the total period of the test, so self-assessment may be preferable.

Panellist assessment of volunteers' underarms may cause embarrassment because of the intimate closeness of panellist to volunteer. This situation can be alleviated by providing clear protocols that stipulate how the volunteers' arms are lifted and held and the angle that panellists use to approach the volunteer. This formalizes the procedure, avoids any physical contact and removes much of the embarrassment.

19.5.2.3 Language

If intensity of underarm malodour (body odour) is the key measurement of interest, panellists should be trained to scale perceived intensity of a range of dilutions of chemicals that have characteristic odours often similar to underarm malodour. A series of aqueous isovaleric acid dilutions from 0.013 mL/L to 3.57 mL/L are suitable training material. Panellists should be instructed to concentrate specifically on the underarm malodour when making their measurements and not to be distracted by any other odours present. Taking this approach cuts down on the variability in the data introduced by giving panellists too much to do in a complex situation.

19.5.3 Personal Washing Products (Soap, Shower Gel)

Personal washing products require consideration of many of the items discussed in the previous two sections and in addition, the issue of fragrance bloom during use (perception of the fragrance in the area surrounding where it is being used), dilution of the product during use and rinsing after use.

Fragrance bloom during use is difficult to measure using accurate and controlled protocols as, in real life, it requires a subject to be lathering their hands or body probably in a steamy atmosphere in a shower. Simulated conditions can be considered, such as diluting and stirring the product in a beaker of warm water and assessing the headspace above the solution, and also, to be more realistic, creating a steamy atmosphere around the beaker. This may provide a reasonable proxy for some of the product experience but does not fully replace it.

Language use will depend on the objectives of the study but is likely to include product and skin-feel characteristics and fragrance attributes during and after use and potentially over a period of time after use.

19.5.4 Hair Care Products (Shampoo, Conditioner, Styling Products, Colorants)

The assessment of hair products, as with other cosmetic and personal care products discussed above, is complicated by the fact that hair types differ from person to person. With hair, however, it is possible to make a choice and either evaluate on real heads of hair or use hair switches. The switches are small uniform bundles of commercially blended, usually virgin hair which can be treated by panellists or trained technicians in order to evaluate appearance, feel, application, foaming and wet and dry hair attributes.

When assessing on real heads of hair, half-head testing is often used. This involves pre-treating the whole head with unperfumed shampoo in order to remove any previous product residue and then half the head is treated with one product and the other half with a second product for comparison. This approach enables a statistical design that compares the performance of the products within each individual subject and the data are then tested statistically across the whole population.

American Testing Standard E2082-12 (ASTM 2012) gives a comparative overview of the advantages and disadvantages of using hair switches or real heads. These include the consideration that the switches enable a high level of control but are an artificial substrate as there is no scalp and sebum involvement, whereas a real head is highly variable from person to person but a completely real substrate. The switches allow rapid sample turnround and relatively low cost for testing once the switches have been purchased (as they can be cleaned and reused), but salon facilities and fully trained salon operatives are needed for real head testing, which is costly; in addition, the number of samples that can be tested at a time are more limited when using real heads so testing is a lengthier process.

19.5.4.1 Pre-treatment

Hair switches should all be from the same bundle or batch. Each switch should be cut to the same length and be a standardized weight. They should all be mounted at one end on a mesh and secured. Depending on the test, the hair should be washed with a standard non-conditioning shampoo or detergent solution (e.g. 5% TEA lauryl sulfate or 15% sodium lauryl sulfate) and then left to dry at room temperature. If the hair has been used for a previous test, it may also need to be presoaked for up to 48 hours in a 50/50 ethanol/water mix before washing in unfragranced shampoo.

If evaluation of the product requires measurement of touch characteristics, the panellists' hands should be treated with a 5% aqueous solution of a surfactant such as TEA lauryl sulfate and dried before the start of the test. If this is not necessary, thin protective gloves should be worn during handling of the hair, and removed and replaced with a new pair between each sample assessment.

19.5.4.2 Language and Product Handling: Hair Switches

- *Appearance*: from a standard amount of product dispensed onto a petri dish, evaluation could include visual firmness, integrity of shape, intensity of colour, brightness of colour, transparency, gloss and pearlescence.
- *Pick-up characteristics*: a standard amount of product is placed on the panellist's thumb; they then compresses it between the thumb and forefinger three times, opening and closing to a predetermined distance. The evaluation attributes could be firmness, stickiness and stringiness.
- *Application and foam characteristics*: a hair switch is wetted and a standard amount of product is measured onto the hair. The procedure for handling the switch and massaging the product into the hair must be fully standardized so that each panellist performs the actions in exactly the same way. Evaluation attributes could include ease of spreading, amount of foam, cushion of foam, wetness of foam, bubble size and ease of rinsing.
- *Wet hair characteristics*: the hair switch should be towel dried or patted dry with laboratory-grade paper towel. Again using standard techniques, the ease of combing and slipperiness of the hair and ease of detangling can be evaluated while combing the hair down the full length of the switch.
- *Dry hair characteristics*: the techniques for drying the hair will vary according to the objectives of the test but could involve hanging to air dry or blow drying using a hair drier. Ease of detangling, slipperiness, force to comb, residue and pliability can all be evaluated while combing the hair, starting at the top and moving down, and then rubbing the hair between thumb and forefinger in a downward motion.

19.5.4.3 Language and Product Handling: Half and Whole Heads

Because of the large variation between subjects, a minimum sample size of 20 volunteers should be used and these people should be pre-screened to fit within the criteria of the target sample population.

- *Appearance and pick-up*: initial appearance and pick-up characteristics can be evaluated in exactly the same way as for hair switches.
- *In-use evaluation*: once the hair is thoroughly wetted, it needs to be parted down the middle from the forehead to the nape of the neck. A measured amount of the shampoo product should be delivered from a syringe directly to the hair on one side of the head. The head is then shampooed using an equal number of strokes and equal amounts of pressure for each head. Evaluation could include ease of spreading and speed to foam or volume of foam, bubble size and wetness of foam. The head can then be rinsed by turning the subject's head to one side and thoroughly rinsing from the forehead to the back of the ear using a standard flow and standard amount of time of flowing water.
- *Wet hair evaluation*: this is essentially similar to the hair switch evaluation but on a full head of hair. A standard combing technique is used and the hair is evaluated for ease of combing and ease of detangling while combing down the full length of the hair. By running the hands and forefingers through the hair, slipperiness and residue can also be evaluated.
- *Dry hair evaluation*: a standard system for drying must be used and once completely dry, the dry hair evaluation can be carried out. Combing the

hair from the centre to the ends of the hair enables an assessment of ease of detangling, force to comb and static. Then running the hands and forefingers through the hair enables assessment of smoothness, residue, pliability, body/fullness and sheen/lustre.
- *Post evaluation*: post evaluations may occur at time intervals after the hair has been treated, for instance 6, 12 and 24 hours later.

19.5.5 Fragrances

Descriptive profiling of fragrances will include considerations relevant to the product that the fragrance is in. This could be a prestige fine fragrance, a cosmetic product, a household cleaning product, an air freshener, laundry product or a whole range of other products from new cars to babies' nappies (diapers). The other sections within this chapter will help with factors that should be considered when deciding how to handle the products.

Assessment of fragrances requires considerations specific to the sense of smell. Particular attention must be paid to the air extraction of the panellists' booths during and after sensory tests to make sure that residual odours do not influence subsequent assessments. For effective design of facilities, the knowledge of an experienced engineer should be used. Adjustment of air pressure in different areas of the facility will prevent odour contamination of assessment areas, and fume cupboards should be used for preparation of particularly smelly samples.

Panellists should be trained not to over-smell or use deep sniffing. They can also be advised to take breaks between samples and to smell a paper tissue or clean skin area to clear their nose between samples. Lengthy smelling sessions must be controlled in order to avoid fatigue. The exact protocols can be developed through trial and error and experience and by checking test reliability and panel performance under different conditions. Feedback sessions also provide a useful opportunity for panellists to discuss their experiences and highlight any problems. As a rule of thumb, when smelling a series of fragrances or fragranced products, a rest of 2 minutes is necessary between smelling each sample.

The descriptive language for fragrances can be extremely detailed, including attributes that cover different odour types such as floral, fruity, green, citrus, herbal, spicy, sweet, woody, marine and chemical or medicinal types of odours. At Givaudan, a training box of 80 odour standards is used as a reference. These standards will usually be agreed with specialist perfumers. An alternative route is to use the test products or a series of relevant market products as stimuli for the panellists to compare, describe and discuss in order to generate relevant language.

19.5.6 Laundry Products (Detergents, Fabric Softeners)

In order to assess the performance of fabric detergents and fabric conditioners or softeners, the sensory analyst needs to wash or treat cloth. The design of experiments in this case is complicated by variables introduced by the washing process and the cloth selected as well as environmental factors (such as water hardness) and the details must be defined and agreed with the project team in order to standardize the approach.

19.5.6.1 Washing Process

This could include treating fabric in the washing machine, rinsing, line drying or machine drying in a home laundry with the products to be tested, or treating fabric using hand washing, rinsing and line drying with the products to be tested. The details agreed by the project team should be reported in detail as they will affect the outcome of any sensory test.

Factors affecting product performance include the following.

- Washing machine specifications, washing cycle type, including length, water volume, water temperatures, spin speeds and water hardness levels.
- Fabric load composition, taking into account the total weight of the load, which affects the movement of the fabric in the wash and also the ratio of product to fabric, and the composition of the fabric included in the wash. It is advisable to use a full load (usually between 2.3 and 2.7 kg dry weight depending on the machine used), making up the load of test cloths with a ballast of mixed fabric types.
- The type of cloth to be used in the evaluation; this will include construction (e.g. 100% cotton, loop terry cloth) as well as fibre density.
- Detergent and/or fabric conditioner dosage.
- Level and type of soil, although protocols without any soil are also common.

19.5.6.2 Cloth Pre-treatment

Includes pre-treatment of all cloth to strip it of any mill conditioners and processing auxiliaries, or any previous test products if the cloth has been used before.

Recommendations given by the ASTM (2013) for evaluating softness state that cloth should be prepared the day before any evaluation and conditioned by leaving in a constant temperature room for 24 hours prior to evaluation at 18–24 °C (65–75 °F) and 40–50% relative humidity. Cloths left for longer have a tendency to lose their 'fluff' over time.

19.5.6.3 Language

The language used will depend on the objective of the test and could include visual appearance (e.g. whiteness), handle (see textile assessments below) and/or odour (which could include residual perception of malodours as well as fragrance).

19.5.6.4 Assessment Considerations

As mentioned above, the types of assessment could include feel, appearance or smell.

- All evaluations of the feel of fabric will require detailed specifications about how the fabric should be handled (see also section 19.5.11 concerning

the assessment of textiles). For instance, softness may be evaluated by cupping the cloth gently in both hands and passing the thumbs over the surface of the cloth as it rests over the fingers. This will only be clearly understood by demonstration and practice with all the panellists as a group.
- Panellists should wash and dry their hands before handling the test fabrics. They may need to rewash their hands to remove any softener or oily build-up during assessments that might interfere with the test.
- Evaluation of visual appearance must take place under standardized lighting and background surfaces. The ASTM (2013) recommends towelling swatches placed on a black background under fluorescent lighting for assessment of whiteness.
- Evaluation of the fragrance on the cloth must be carried out under controlled conditions free of any background odour, preferably in booths with good extraction facilities and controlled temperature and humidity. Variations in temperature and humidity may affect the fragrance release from the cloth.

19.5.6.5 Product Evaluation

Assessment of laundry products provides a good example of a situation where timing of the evaluation is important. When assessing the ability of the fabric fragrance to fill a room during drying, sufficient time must be allowed for the fragrance to fill the room before the assessment is made. There will be a period of time during which the fragrance builds and then stabilizes. The character will remain stable for a while but then change as the cloth dries. After drying, during storage for perhaps 1 day, 1 week or several months, the odour character may change over time so it is important to carefully define and control assessment time intervals and standardize across product evaluations.

19.5.7 Cleaning Products (Spray Cleaners, Dishwashing Products (Liquid, Dishwasher Tablets))

The whole experience of using household cleaners offers a number of different assessment points and aspects of the product that could be the focus of an evaluation.

- The visual appearance of the product, including packaging shape and colour.
- The feel of the product and ease of handling. This could include the feel of the surface of the pack and the action of a trigger spray or the feel of the pack when squeezed to dispense the product.
- The appearance and characteristics of the product as it is applied to a surface; this would include flow and shine of a cream or distribution of a spray.
- The feel of the product during use, such as ease of spread.
- The efficiency of the product in removing soil.
- The appearance of a surface after cleaning.
- The odour of the product before, during and after use.

The objective and scope of the sensory test will determine which of these different points of assessment should be included in a test, and the sensory methods adapted accordingly.

19.5.7.1 Pre-treatment

Assessment of the product's ability to remove soil or its performance characteristics on a surface will require preparation of a standard surface for testing. Typically, this might be a non-porous surface such as a ceramic tile which can be thoroughly cleaned before and between uses.

The cleaning process might typically include:

- a wash in warm water with unscented hand dish wash liquid
- rinse with deionized water
- dry with laboratory-grade paper or cheese cloth
- rinse with acetone
- wipe dry

ASTM E2346 (2009) recommends placing the cleaned surface into a live stream of air. By doing this, the effectiveness of the cleaning process can be confirmed as any areas not thoroughly cleaned will take on a white, highly reflective appearance. After cleaning, the tiles should be placed on a vertical rack.

The pre-treatment may also include application of a standard soil. Industry standard soils are available or specific soils of interest (e.g. grease or food soils) may be applied in a pre-determined standard and uniform way. This process requires skill and judgement to be sure that a relevant amount of soil is applied and the application is reproducible from sample to sample.

When evaluating trigger sprays, the products should all be tested from identical packs with the same type of trigger and nozzle. Noticeable differences in the amount of product released, the distribution of the spray and the nature of the spray are delivered by different types of pack.

19.5.7.2 Product Handling

Knowledge of consumer habits and/or on-pack instructions will help to define the instructions given to panellists about how to handle the product and when and how to evaluate performance. The distance of the pack from the surface, the timing of a spray or number of complete full depressions of the trigger will all affect delivery of the product to the surface. Details such as the timing of the action and whether or not to wipe and rinse must all be exactly agreed and understood by the panellists.

The surface to be treated should ideally be presented to the panellists in the manner in which it would usually be found (e.g. a kitchen surface will be horizontal but a mirror would be vertical).

19.5.7.3 Language and Evaluation

For this type of product, it could be the surface being treated or the product itself that is evaluated, or the effect of the product on the surrounding environment (such as the release of fragrance). The evaluation procedure and language should be selected and agreed to suit the objectives of the test.

Each panellist should work alone and ideally be given the space and manoeuvrability to handle the products and make the evaluations.

19.5.7.4 Experimental Design

Sample randomization will be constrained by the number of cleaning surfaces available. Ideally, the surfaces will be removable so that after each panellist assessment, a new surface can be introduced. This enables assessment of several products in one session.

19.5.8 Toilet Cleaners

Many of the comments made above are also relevant for toilet cleaners but additional factors may need to be considered such as the distribution of the product around the bowl and removal after rinsing. Another consideration is the mechanism of a toilet flush or consumer habits in cleaning different types of toilets, particularly where water is scarce, which will influence the performance of the product so this should be carefully researched prior to starting any testing.

19.5.9 Air Fresheners (Stick-Ups/Gels, Sprays, Plug-Ins, Candles)

There is a vast array of different types of air freshening products and the types of appliances and their mode of action is constantly being updated to meet current consumer needs or introduce novel applications. This means that methods of testing need also to be regularly reviewed and updated.

There are two main performance criteria for air fresheners that are commonly evaluated: filling a space with a pleasant odour and cleaning or removing unpleasant odours. These are difficult to measure without appropriate rooms that can be used for testing and then efficiently cleaned of any residual odour after use. Manufacturers often build a series of identical small rooms with air extraction facilities but in the absence of these, a number of identical hotel rooms may be hired for the purpose.

19.5.10 Textiles: Fabric Hand

The term fabric 'hand' or 'handle' has been defined as the quality of a fabric or yarn assessed by the reaction obtained from the sense of touch or the total sum of the sensations expressed when a textile fabric is touched by touching and flexing with the fingers (Bishop 1996). Other sensorial attributes may also be important such as visual or acoustic perceptions or potentially also odour perceptions.

Unfortunately, fabric hand has been described differently by different researchers over the years and no one clear approach to assessing fabric hand has been defined. Bakar (2004) and Naima et al. (2013) both give reviews of the various approaches and the range of different language used by researchers which is a useful starting point for the sensory analyst. The following guidelines and considerations are very general and should be followed up with more specific refinements according to the fabric type and proposed end use.

19.5.10.1 Fabric Handling

Specific patterns or strokes of the hand can be used to assess particular details of perception (defined by Moody et al. 2001).

- The touch-stroke method, which involves all fingers stroking the fabric surface, is used to evaluate surface details (texture) and thermal feel related to temperature.
- The rotating cupped method tests for comfort relative to stiffness, weight, resilience, temperature and texture.
- The multiple pinch action involves fabric being rotated between the thumb and one or two other fingers of the same hand to feel the stiffness, texture, structure, friction property, stretch and temperature.
- The two hand rotation seeks to evaluate sheerness and stretch.

Detailed panellist training is necessary to ensure uniform approaches across the panel to the use of the methods selected. Naima et al. (2013) give a detailed description of the progress of their panel training to use these methods.

19.5.10.2 Language

The language used for fabric assessment will be determined by the type of fabric in question and should be refined by discussion with the panellists and the technology or development team. A good example of this process is given by Naima et al. (2013) who started with a list of 56 attributes describing a wide range of fabrics and refined them eventually to 16 using a process of sorting and then principal component analysis. Discussion with the panellists then defined positive and negative references to help define the attributes. These 16 attributes were grouped into three categories: surface handle, characterizing the textile fabric surface (cold-heat, moist-dry, smooth-grooved, tender, silky, sleek, slippery, hairy, compact); physical handle, characterizing the physical properties of the textile fabric (thin-thick, light-heavy, falling); and dynamic handle, describing the dynamic properties of the textile fabric which require the application of light force on the fabric to determine them (supple-stiff, flexible, elastic, wrinkly).

19.5.10.3 Special Considerations

Panellists' hands must be washed and dried thoroughly before assessment sessions as touch sensations are stronger with high moisture content on the skin (Kamalha et al. 2013) and variability may be introduced by the presence of cosmetic products (Naima et al. 2013). Naima et al. (2013) also suggest that evaluation time should be limited to 30 minutes as the hands will become less sensitive if the test is too long.

19.5.10.4 Fabric Pre-treatment

Many fabrics are inherently variable because they are designed with stretch characteristics. Pensé-Lhéritier et al. (2006) worked with knitted fabrics that stretched easily and pre-treated them to provide a relaxed state which is the most reproducible state. This pre-treatment was carried out on wet fabric with mechanical vibrations under water steam for 5 minutes. Other pre-treatments may be necessary for different fabrics for different reasons (e.g. stiffness of cotton or silk) to ensure a uniform state that is reproducible over a series of tests.

It is also necessary to store cloth at a standard temperature and humidity for 24 hours prior to testing to standardize the perception of surface feel. Naima et al. (2013) used a temperature of 20 °C ± 2 °C and relative humidity of 65% ± 2% to store a wide range of different types of fabric samples prior to assessment.

19.5.10.5 Testing Environment

The tests should be carried out in a room where the temperature and humidity are constant and felt comfortable by the subjects, avoiding clammy, hot or particularly dry and cold hands.

19.5.11 Disposable Razors and Shaving Foam/Gel

The feeling of the shaving foam and/or razor on skin during shaving as well as the after-feel of skin following shaving is important. Properties such as 'thickness' may be a positive attribute when the shaving foam is initially applied (the consumer can feel the product is present and the thickness provides some protection during shaving). However, after shaving (on interaction with water), a consumer is likely to want the product to be lower in thickness so it is easily rinsed away. For this reason, attributes related to shaving foam should include various time points to capture the overall properties of the product. Stages of rating may include during application, immediately after application, interaction with water, following rinsing and following drying.

Attributes may include the following.

- Properties of the shaving gel: smoothness, thickness
- Feeling of skin during shaving: smooth, slimy
- Feeling of skin after shaving: irritation, softness, dryness
- Properties of the razor: closeness of shave, ease of use

Hair length is also an important consideration, depending on the application. For example, for products targeted at women, panellists may be required to grow hair to be 0.5 cm long prior to testing. For this reason, testing may be limited to every couple of days for underarm shaving or more often for legs as panellists could use a set area per test.

19.5.12 Paper Products

Paper products include a variety of different items:

- paper used for writing/printing
- paper products used in personal hygiene and cleaning such as toilet paper, paper towels, kitchen roll and tissues
- functional paper products such as paper plates, paper cups, paper bags, napkins, boxes.

This section describes methods employed for testing paper products used in hygiene and cleaning.

Toilet rolls, paper towels, tissues and kitchen roll are all held in the hand when used, so the feel properties are important from a consumer's perspective. Key attributes for these products include softness (especially for tissues/toilet roll as they have more interaction with skin that may cause irritation if the product is too rough) and rigidity (for kitchen roll and paper towels in particular as they need to remain intact during cleaning).

Batch-to-batch variability is a major issue for manufacturers so products are often tested using descriptive panels to monitor quality between batches. Samples should be stored in the same environment (ideally in cool, dry conditions) prior to testing and, where possible, packaging should not be opened as this can affect the moisture content of the products. After packaging has been opened, samples should therefore be stored in air-tight containers away from a humid environment.

Assessment areas may include the following.

- Visual properties including: visual texture (presence of layers or embossed pattern) and colour.
- Textural properties including: softness, pliability, thickness, smoothness, cushiony and absorbency (particularly for paper towels and kitchen roll).
- Odour: some products contain a 'balm' and therefore may be rated for attributes describing the aroma such as menthol and minty.

19.5.13 Automotive

The automotive industry has many different types of consumer, from commercial van drivers to families. The characteristics of the vehicle to be tested will vary depending on the targeted consumer. Also the variety of factors that affect performance and consumer acceptability cover a wide range of specialisms in sensory science; for example, both sound and odour are important and both require specialist panels. This section will introduce some of the sensory attributes related to automobiles, in particular focusing on those related to cars, although the areas discussed can be translated to other types of vehicle. The following key areas are covered: odour, sound and comfort.

19.5.13.1 Odour

Odour – 'new car smell' – is very important to consumers and has even been incorporated into air fresheners. The new car smell is generally made up of the different interior trim parts and materials used within the car and is more commonly tested using GC and GC-MS analysis, and more recently the 'electronic nose' (Morvan et al. 2000).

19.5.13.1.1 Special Considerations

Ideally, panellists must have good sense of smell and not have anosmia to any odours; this is particularly important as many odours used in 'new car smell' may be chemical and therefore less commonly recognized by the general population.

19.5.13.1.2 Methodology

Descriptive testing of car odours may be carried out in fragrance booths with good extraction facilities. In this case, materials from the car would be brought into the booths and rated for different odour qualities such as overall intensity, chemical odour and rubber odour. However, in some cases in-context testing would be more realistic and may be used where the overall new car smell is important. For example, the odour released from leather seats will interact with the other odour compounds within the car in a different way to odour emitted from fabric seats.

To carry out in-context testing, the panellist would sit within the car to rate the aroma for different properties, perhaps using a tablet computer or paper questionnaire to record their responses. When carrying out this testing, it would be important to rate the odour perceived in different conditions – changing, for example, the air-conditioning/heater settings and position of the windows.

A more detailed summary of sensory descriptive odour analysis of car cabins has been reported by Verriele et al. (2012). This research uses the concept of the ‘field of odours’ odour characterization methodology in which chemical standards representing different fragrances are used to aid training of the panel.

19.5.13.2 Sound

Sound plays a large role in a consumer’s perception of the quality and performance of a car. For example, the sound of the engine as well as the sound of a closing door can determine whether a car is perceived to be luxury or cheap (Fastl 2005; Lyon 2003). The sound of a car horn also evokes images of the size of the car in a consumer’s mind; a deeper sounding horn matches a consumer’s perception of a larger vehicle whereas a high-pitched tone is associated with a smaller car.

The sound of a car (perhaps neglecting engine noise) is not necessarily thought to be important by the consumer; however, if neglected by the manufacturer, it can have a massive impact on the overall perception of the product. For example, a car with no engine noise may suggest to the consumer that something is wrong or the car has limited performance. Consequently, manufacturers employ large teams to tune the exhaust notes of their cars so that they sound ‘beefy and raucous’ or ‘smooth and refined’ to reflect the image of the car or appeal to different consumers. Electric cars have given rise to some unusual problems for manufacturers; for example, the absence of sound may create danger for pedestrians who rely on the sound of approaching vehicles when crossing the road (Cerrato 2009). This has sparked discussion over alternative sounds or even recorded engine sounds to be played in built-up areas so that the car is noticed by pedestrians.

The presence of sound when closing a car door is also important; if there was silence on closing the door, you may think it hasn’t closed properly and the car is not secure. It is therefore crucial that car manufacturers match the correct sounds to consumer expectations.

19.5.13.2.1 Special Considerations

Panellists involved in assessing sound quality must have exceptional hearing and be screened to test their sensitivity to different sounds. Suitable candidates for a panel measuring sound may include musicians and recording engineers due to their enhanced sound sensitivity.

19.5.13.2.2 Methodology

In practice, the majority of testing would take place in a lab/sensory booth wearing headphones through which sounds are played or testing may take place in a soundproof booth into which the actual car may be placed and, for example, the engine switched on for the overall sound to be rated. Panellists would listen and rate attributes relating to sound quality such as loudness, pitch, wholeness of sound (is it a thin or thick sound?), fluctuation in sound, echoey, brightness, booming, tinny, roaring and sharpness.

In cases where recordings are used, it is important that the same conditions are in place for every recording. For example, in the descriptive analysis of diesel engine sound, the microphone used in recording the sounds must be the same distance from the bonnet of each car recorded (e.g.1 m), and all cars must be in the same environment whilst the recording takes place (either outside or inside, depending on the objectives of the study) (Poirson et al. 2010). For studies testing the acoustic properties experienced during driving, the HMS III Head Acoustics GmbH artificial head and torso system may be used to make recordings (Brizon & Medeiros 2012). Panellists should also wear the same type of headphones to ensure they all receive the same sound quality recording.

The following list describes some of the key features of automobiles that create sounds important in sensory research.

- Engine (when stationary and in motion)
- Air-conditioning/fan heating system
- Windscreen wipers
- Door opening and closing
- Windows opening and closing
- Locking and unlocking
- Alarm
- Car horn
- Tyre road noise during driving
- Sound system – quality of speakers

Although descriptive analysis is used in the car industry, the competitive environment means that innovation cycles are often shortened and therefore the need for faster sensory methods is heightened. For this reason, many car makers are looking to use alternative, more rapid sensory evaluation methods such as free choice profiling and flash profiling (Poirson et al. 2010).

19.5.13.3 Comfort

The overall comfort of a car's interior is also of high importance, especially as people are spending more time travelling long distances in cars (Gameiro da Silva 2002). Areas of descriptive testing relating to comfort in the automobile industry include:

- feeling of fabric used in the car (Giboreau et al. 2001)
- drivability of the car (Liu et al. 2013)
- car seat comfort (Kyung et al. 2008)
- vibrations (Hostens & Ramon 2003).

19.5.14 Pharmaceuticals

Pharmaceutical products are generally tested in clinical trials for their efficacy. However, the flavour of pharmaceutical products is often undesirable (drug products often have a bitter taste that is difficult to mask) which presents a huge problem to the pharmaceutical industry. If a manufacturer wants its product to succeed on the market, it needs to be palatable to the consumer. This is of particular importance in paediatric medicine where children may refuse to take medicine if it tastes unpleasant. Sensory panels are therefore used to test the aroma and flavour of pharmaceutical products either by testing the carrier (tablet/liquid base) or the product including the drug. For the latter method, the product may be held in mouth then expectorated to minimize drug consumption by the panellist.

Assessment points include the following.

- *Appearance*: the visual properties of the product in pack as well as the dispensed product on a spoon (liquid) or in hand (capsule); attributes could include translucency, shininess, colour, size of pill, thickness of liquid.
- *Aroma*: more commonly used for liquid products although some capsules when presented loose in a bottle may have an aroma. Aroma attributes may include menthol, overall intensity, burning sensation, chalky.
- *Flavour*: holding the product in mouth for a set amount of time then expectorating and evaluating the following key areas: overall intensity, sweet, sour, salt, bitter, metallic, cooling, hot, spicy, burning, anaesthetic, astringent, medicinal, minty/menthol, warming, sharp, alcohol, painful, irritating, stinging, dry, peppery and paper (Anand et al. 2007).
- For non-toxic products such as cough sweets and lozenges, dissolvability and ease of swallowing may also be tested.

19.5.14.1 Special Considerations

As with most sensory methodologies, panellists must be healthy volunteers. However, due to the nature of this testing, they must also undergo a strict medical screening process to ensure that they are in optimum health and that any products tested will not interfere with their current condition. Pharmaceutical sensory tests are carried out under more controlled conditions than a standard sensory test and the ethics behind each test is rigorously investigated prior to testing. The Code of Ethics of the World Medical Association (Declaration of Helsinki) lays down ethical guidelines for medical research involving human subjects (WMA 2013). Due to the potential risk involved, recruiting for 'pharmaceutical sensory panels' can be difficult and maintaining panellists' motivation for testing these products can also be a challenge (Anand et al. 2007). The flavor profile* method is commonly used for testing pharmaceutical products as it involves fewer assessors (4–6), thus overcoming the issues with recruitment, and it focuses on aroma, flavour and aftertaste, the key areas of interest required for pharmaceutical analysis (Lorenz et al. 2009).

Legal issues as well as potential toxicity limit the number of samples that can be tested at any one time. Lorenz et al. (2009), for example, presented panellists with two products in a 1-hour session with a maximum of four products to be tested in a day. However, a more toxic product would require tighter restrictions and may require a number of days break between testing to remove any potential drug from the panellists' systems.

The limited number of samples that can be tested in a pharmaceutical sensory panel session has led to the development of the 'electronic tongue' – 'an instrument that can be trained to screen taste attributes of formulations in a rapid timeframe, when used in conjunction with human taste assessment data' (Guhmann et al. 2012; Lorenz et al. 2009). Studies using electronic tongues have shown the results to be useful for identifying differences between samples. However, despite their ability, they are still not as sensitive as humans and therefore the sensory panel is still key in the testing of pharmaceutical aroma and flavour (Eckert et al. 2013). The electronic tongue is likely to be more widely used in future as a screening tool to reduce the number of samples that need to be tested by the human sensory panel (Lorenz et al. 2009).

19.6 Summary

This chapter has focused on the measurement of the sensory properties of non-food products with particular reference to the handling of the products and highlighting the need to consider all the senses and all the assessment points throughout the process of using the product. Specific expertise beyond the sensory laboratory may need to be consulted in order to fully develop relevant protocols and define assessment points. This may include acquiring consumer usage and attitude information or the expertise of the technical, product development or applications laboratories. Attention to detail must be given to make sure that panellists are fully trained and consistently use the specified handling protocols during assessments. These requirements should not distract, however,

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

from the fundamental need to develop relevant and meaningful language for the panellists to use and to follow the accepted sensory principles of good experimental design and statistical analysis of results.

19.7 Future Developments

Traditionally non-food sensory scientists have followed tightly controlled and rigorous experimental designs and sensory approaches. There is now an awareness of how important it is to consider less controlled 'real-life' experimental approaches in order to understand how consumers might perceive products in a real-use situation. These have been discussed at global sensory conferences both for food and more recently also non-food sensory studies (Churchill 2014). Examples have been provided to illustrate how good sensory data can be collated by experienced scientists under less controlled, more realistic conditions. This area is important to continue as a focus for development of descriptive sensory analysis techniques for the future.

References

Anand, V., Kataria, M., Kukkar, V., Saharan, V. & Choudhury, P.K. (2007) The latest trends in the taste assessment of pharmaceuticals. *Drug Discovery Today*, **12**(5/6), 257–265.

ASTM (2009) *E2346-09. Standard Guide for Sensory Evaluation of Household Hard Surface-Cleaning Products with Emphasis on Spray Triggers*. West Conshohocken: ASTM International.

ASTM (2011) *E1490-11. Standard Guide for Two Descriptive Analysis Approaches for Skin Creams and Lotions*. West Conshohocken: ASTM International.

ASTM (2012) *E2082-12. Standard Guide for Descriptive Analysis of Shampoo Performance*. West Conshohocken: ASTM International.

ASTM (2013) *D5237-05. Standard Guide for Evaluating Fabric Softeners*. West Conshohocken: ASTM International.

ASTM (2016) *E1958-16a. Standard Guide for Sensory Claim Substantiation*. West Conshohocken: ASTM International.

Aust, L.B., Oddo, L.P., Wild, J.E., Mills, O.H. & Deupree, J.S. (1987) The descriptive analysis of skin care products by a trained panel of judges, *Journal of the Society of Cosmetic Chemists*, **38**, 443–449.

Bakar, B.A. (2004) Subjective and objective evaluation of fabric handle characteristics. Master of Science thesis, Department of Textiles, University of Leeds.

Bishop, D.P. (1996) Fabrics: sensory and mechanical properties. *Textile Progress*, **26**, 1–62.

Brizon, C.J.S. & Medeiros, E.B. (2012) Combining subjective and objective assessments to improve acoustic comfort evaluation of motor cars. *Applied Acoustics*, **73**, 913–920.

Cerrato, G. (2009) Automotive sound quality - powertrain, road and wind noise. *Sound & Vibration*, 16–24.

Churchill, A. (2014) Developments in the Sensory Measurement of Fragranced Products. Presented at SenseAsia, 11–13 May, SingEx, Singapore.

Churchill, A., Meyners, M., Griffiths, L. & Bailey, P. (2009) The cross-modal effect of fragrance in shampoo: modifying the perceived feelof both product and hair during and after washing. *Food Quality & Preference*, **20**, 320–328.

Close, J., Blank, R., Gelinas, A. & Penkin, N. (1982) Sensory evaluation: a scientific aid for R&D chemists. *Cosmetic Technology*, 42–45.

Eckert, C., Pein, M., Reimann, J. & Breitkreutz, J. (2013) Taste evaluation of multicomponent mixtures using a human taste panel, electronic taste sensing systems and HPLC. *Sensors & Actuators B*, **182**, 294–299.

Fastl, H. (2005) Psychoacoustics and sound quality. In: Blauert, J. (ed.) *Communication Acoustics*. Berlin: Springer.

Gameiro da Silva, M.C. (2002) Measurement of comfort in vehicles. *Measurement Science & Technology*, **13**, R41–R60.

Giboreau, A., Navarro, S., Faye, P. & Dumortier, J. (2001) Sensory evaluation of automotive fabrics: the contribution of categorization tasks and non-verbal information to set-up a descriptive method of tactile properties. *Food Quality & Preference*, **12**, 311–322.

Guhmann, M., Preis, M., Gerber, F, Pöllinger, N., Breitkreuz, J. & Weitschies, W. (2012) Development of oral taste masked diclofenac formulations using a taste sensing system. *International Journal of Pharmaceutics*, **438**, 81–90.

Hostens, I. & Ramon, H. (2003) Descriptive analysis of combine cabin vibrations and their effect on the human body. *Journal of Sound & Vibration*, **266**, 453–464.

Kamalha, E., Zeng, Y., Mwasiagi, J.I. & Kyatuheire, S. (2013) The comfort dimension: a review of perception in clothing. *Journal of Sensory Studies*, **28**, 423–444.

Kelly, G.A.. (1955) *The Psychology of Personal Constructs*. New York: Norton.

Kemp, S.E., Hollowood, T. & Hort, J. (2009) *Sensory Evaluation: A Practical Handbook*. Oxford: Wiley-Blackwell.

Kyung, G. & Nussbaum, M.A. (2008) Driver sitting comfort and discomfort (part II): relationships with and prediction from interface pressure. *International Journal of Industrial Ergonomics*, **38**, 526–538.

Kyung, G., Nussbaum, M.A. & Babski-Reeves, K. (2008) Driver sitting comfort and discomfort (part I): use of subjective ratings in discriminating car seats and correspondence among ratings. *International Journal of Industrial Ergonomics*, **38**, 516–525.

Liu, P., Zhang, T. & Zhao, X. (2013) Vehicle drivability evaluation and pedal-acceleration response analysis. *Advances in Information Sciences & Service Sciences*, **5**(10), 506–513.

Lorenz, J.K., Reo, J.P., Hendl, O., Worthington, J.H. & Petrossian, V.D. (2009) Evaluation of a taste sensor instrument (electronic tongue) for use in formulation development. *International Journal of Pharmaceutics*, **367**, 65–72.

Lyon, R.H. (2003) Product sound quality – from perception to design. *Sound & Vibrations*, 18–22.

Meilgaard, M., Civille, G.V. & Carr, B.T. (1991) *Sensory Evaluation Techniques*. Boca Raton: CRC Press.

Moody, W., Morgan, R., Dillon, P., Baber, C. & Wing, A. (2001) Factors Underlying Fabric Perception. Presented at the 1st EuroHaptics Conference, Birmingham.

Morvan, M., Talou, T., Gaset, A. & Beziau, J.F. (2000) Electronic-nose systems for control quality applications in automotive industry. *Sensors & Actuators B*, **69**, 384–388.

Moskowitz, H.R. (1982) Sensory analysis of thickness. *Cosmetics & Toiletries*, **97**, 35–45.

Munôz, A. (2103) Current Status of and New Developments in the Field of QC/Sensory Evaluation. Presented at the 10th Pangborn Sensory Science Symposium, August, Rio de Janeiro.

Naima, H.A., Mehdi, S. & Morched, C.H. (2013) Training and control of performance of a tactile sensory panel. *Journal of Applied Sciences*, **13**(3), 366–376.

Poirson, E., Petiot, J.F. & Richard, F. (2010) A method for perceptual evaluation of products by naïve subjects: application to car engine sounds. *International Journal of Industrial Ergonomics*, **40**(5), 504–516.

Pensé-Lhéritier, A.M., Guilabert, C., Bueno, M.A., Sahnoun, M. & Renner, M. (2006) Sensory evaluation of the touch of a great number of fabrics. *Food Quality & Preference*, **17**(6), 482–488.

Verriele, M., Plaisance, H., Vandenbilcke, V., Locoge, N., Jaubert, J.N. & Meunier, G. (2012) Odor evaluation and discrimination of car cabin and its components: application of the 'Field of odors' approach in a sensory descriptive analysis. *Journal of Sensory Studies*, **27**, 102–110.

WMA (2013) WMA Declaration of Helsinki – Ethical Principles for Medical Research Involving Human Subjects. *Journal of the American Medical Association*, **310**(20), 2191–2194.

Wortel, V.A.L. & Wiechers, J.W. (2000) Skin sensory performance of individual personal care ingredients and marketed personal care products. *Food Quality & Preference*, **11**, 121–127.

SECTION 4
Summary

CHAPTER 20
Comparison of Descriptive Analysis Methods

Alejandra M. Muñoz, Sarah E. Kemp, Tracey Hollowood and Joanne Hort

20.1 Introduction

Descriptive analysis has undergone a remarkable growth from its inception in the 1940s with the development of the flavor profile* method, to the present time and the development of non-traditional/rapid descriptive methods.

The most well-known descriptive methods are covered in this book. Each chapter presents a detailed discussion of each method's characteristics, its implementation and use (i.e. selection/screening of assessors, training approach (if applicable), data collection, data analysis, applications, etc.). Thus the reader is encouraged to review the chapter(s) of interest for comprehensive information on each method, its philosophy, implementation and applications.

Other texts have covered descriptive methodologies. For example, the ASTM *Manual on Descriptive Analysis Testing for Sensory Evaluation* (ASTM 1992) presents a summary of several traditional descriptive methods and compares key elements of these methods. More recently, Dehlholm et al. (2012), Valentin et al. (2012) and Varela and Ares (2012) summarize newer methodologies. Chapter 1 of this book also gives an overview of descriptive methods.

Over the years, numerous studies have compared results produced by different descriptive methods with the aim of validating and/or identifying differences, benefits and disadvantages. Examples of studies comparing newly developed methods against older traditional methods include Williams and Arnold (1985), Risvik et al. (1997), Dairou and Sieffermann (2002), Delarue and Sieffermann (2004), Moussaoui and Varela (2010), Albert et al. (2011), Dehlholm et al. (2012) and da Silva et al. (2013). Examples of studies comparing newer

* 'Flavor profile' is a formal name in common usage using American English spelling and is therefore cited in this manner.

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.

methods against each other include Dehlholm et al. (2012), Valentin et al. (2012) and Varela and Ares (2012).

There have also been many studies comparing different aspects of descriptive methodology, such as the execution of the same study with trained assessors and untrained consumers (for example, see Albert et al. 2011; Ares et al. 2015; Bárcenas et al. 2004; Bruzzone et al. 2012; Lelièvre et al. 2008; Moussaoui & Varela 2010; Worch et al. 2010).

This chapter's main objective is to present a focused discussion on the key features, strengths/weaknesses and main differences/similarities among the methods covered in this book. The intention is to provide the reader with sufficient information to assess, compare and contrast all methods at once. In addition, this information can assist practitioners in choosing the most appropriate descriptive philosophy and method for their situation/company.

For thorough and specific information on a particular method of interest, the reader is encouraged to review the corresponding chapter in this book. In addition, an overview of less well-known descriptive methods is given in the introductory chapter. Temporal descriptive methods fall beyond the scope of this book, and thus are not covered in this chapter.

20.2 Relevant Comments on the Presentation of the Methods' Key Characteristics and Their Comparison

Prior to presenting the specific comparison and discussion of methods (sections 20.3 and 20.4), several observations are needed to aid in the review of these sections.

- Several key characteristics were chosen to provide an overall overview/comparison of all the methods presented in this book in tabular form. The chosen parameters provide a broad overview of the main differences and similarities among the methods. This comparison is presented in section 20.3 and Tables 20.1 and 20.2.
- A comparison based on several methodological characteristics is presented in section 20.4 and Tables 20.3–20.20. In this section, methods are more thoroughly contrasted. Methods that are similar in a characteristic/comparison point are grouped/clustered and discussed as a group.

The clustering or grouping of methods varies depending on the characteristic/discussion point. The main methodological characteristics chosen for comparison in this summary chapter are as follows.

- Philosophy
  - Traditional versus non-traditional/rapid
  - Technical versus consumer-based approaches
- Type of results/output
- Overall needs for programme implementation

- Assessors
    - Type of assessors and selection/screening
    - Recommended number
- Panel leader's role
- Training characteristics
    - Training versus participation/no training
    - Nature/level and length of training
- Characteristics of terminology/lexicon
    - Generation of terminology/lexicon
    - Type of terminology
    - Singular versus complex terms/attributes used by technical methods
- Training references
    - Qualitative
    - Quantitative
- Data collection/scoring approach
- Data analysis
- Advantage/strengths and disadvantages/weaknesses
- Applications

Quantitative descriptive analysis (QDA) as described in this chapter is the original and proprietary method of Tragon (as described in Chapter 8), rather than the generic QDA-type methodology that is commonly termed QDA in Europe.

Ranking and rank-rating methodology can be viewed purely as a scaling approach or as a full methodology in its own right. In this chapter, it will be classified as a traditional-type descriptive methodology, comprising the stages of assessor selection, assessor training with a panel leader, including attribute generation, familiarization/training on intensity references (if used), validation, data generation and statistical analysis. Since ranking and rank-rating methodology can be used in combination with any traditional descriptive philosophy, some comparison tables show the method in more than one group/philosophy.

Check-all-that-apply (CATA), also referred to as tick-all-that-apply (TATA), is an old method that has long been used in market research and has included sensory attributes. However, it has only recently received interest as a technique for rapid descriptive analysis in the field of sensory science. It will therefore be treated in this chapter as a new/rapid method for descriptive analysis.

20.3 Overall Comparison of Descriptive Methods

Tables 20.1 and 20.2 show the comparison of traditional methods and new/rapid methods, respectively, presented in this book based on key features. These tables show how the two distinct categories of descriptive methods (traditional and non-traditional/rapid) differ. In addition, the reader is referred to Table 20.19, which summarizes each method's main advantages and disadvantages.

Information on time required for data collection is not presented in Tables 20.1 and 20.2, since it is difficult to compare this information across

Table 20.1 Overall comparison of key features of traditional descriptive analysis methods.

Method	Assessors	Type of sensory language and references used	Scale and measurement	Panel training	Data analysis
Consensus method	• Minimum 4–6 • Screened • Trained	• Technical • Comprehensive • Uses references	• Typically 1–15 • May use intensity references • Consensus results	• Yes • Extensive • >100 hours (4–12 months) • Panel leader who might also be one of the panellists • Checks on performance	• None
Flavor profile method (FP)/ Texture profile method (TP)	• 6–10 • Screened • Trained	• Technical • Comprehensive • Uses references	• FP scales (e.g. 4, 7 points) and consensus results • TP rating scales Might use consensus or collect individual measurements	• Yes • Extensive • 4–6 months (100 hours) • Panel leader • Checks on performance	• None for consensus data • Quantitative data from TP use analysis as for other profile-based methods
Spectrum™ Method	• 12–20 • Screened • Trained	• Technical • Comprehensive • Uses references	• Universal, absolute numerical 150-point scale • Intensity references • Replication	• Yes • Extensive • 3–4 months per modality, e.g. flavour • Panel leader • Validated performance	• ANOVA • Multiple range tests (e.g. Duncan, Tukey, LSD) • Correlation • Multivariate analyses
Quantitative flavour profiling (QFP)	• 12–16 • Screened • Trained	• Technical • Comprehensive • Uses references	• Unstructured line scale • Anchor products • Replication	• Yes • Extensive • 6 months • Panel leader • Validated performance	• ANOVA • Multiple range tests (e.g. Duncan, Tukey, LSD) • Correlation • Multivariate analyses

A5daptive Profile Method®	• Number depends on customized method (typically 10–15) • Screened • Trained	• Technical • Comprehensive • Uses references	• Rating method dependent on customized methodology (i.e. individual measurements or consensus) • 10-point scale, expanded/adapted for higher intensities • Universal or product-specific rating • Intensity references • Replication	• Yes • Extensive • Time depends on customized method • Panel leader/moderator • Validated performance	• ANOVA • Multiple range tests (e.g. Duncan, Tukey, LSD) • Correlation • Multivariate analyses
Quantitative descriptive analysis (QDA)	• 10–12 • Screened • Trained	• Consumer based • Study specific • Uses only qualitative references	• Unstructured or semi-structured line scale • Product/study-specific scale based on range of sample intensities • Replication	• Yes • Shorter than profiling type • Min. 8–12 hours (2 weeks) • Panel leader is moderator • Validated performance	• ANOVA • Multiple range tests (e.g. Duncan, Tukey, LSD) • Correlation • Multivariate analyses
Ranking and rank-rating descriptive analysis	• Minimum 12 • Screened • Trained or naive assessors can be used	• Technical or consumer based • Limited number of attributes • Might use references	• Ranking giving ordinal data • Rating data for rank-rating with a scale • All products must be presented simultaneously	• Yes • Shorter • Can include panel leader/moderator and validation of performance	• Rank data: Friedman and R-Index analyses • Rating data: as above

ANOVA, analysis of variance; LSD, least significant difference.

Table 20.2 Overview comparison of key features of new/rapid methods.

Method [modified]*	Assessors	Type of sensory language and references used	Measurement and type of data	Panel training	Data analysis
Check/tick-all-that-apply (CATA, TATA) [pick K-attributes]	• 50–100 • No requirement for screening or training	• Limited list of predetermined attributes • Attributes typically not defined • No references	• Selection of attributes that apply to a product • Nominal data • Frequencies • No quantitative intensity data	• None	• Frequency matrix • CA • MFA • MCA
Free choice profiling (FCP)	• 20–30 • No requirement for screening or training	• Idiosyncratic, consumer • No references	• Rating	• None	• GPA • PCA
Flash profiling (FP) [rank descriptive analysis (RDA), individual vocabulary profiling (IVP)]	• 4–12 trained • 20–50 naive • No requirement for screening or training	• Idiosyncratic, consumer • Individual attribute list reviewed against global list • No references	• Ranking giving ordinal data • Replication • All products must be presented simultaneously	• None	• PCA • GPA • MFA • STASIS • DISTASIS
Sorting [labelled, directed, descending/ ascending hierarchical]	• Minimum 20 • No requirement for screening or training	• None • No references • Post-sorting naming of groups may be carried out	• Sorting/grouping products according to their similarities • All products must be presented simultaneously	• None	• Similarity matrix • MDS • MCA • DISTATIS • SORTCC

Projective mapping [ultra FP, Sorted Napping®, polarized projective mapping]	• Minimum 20 but up to 100 has been reported • No requirement for screening or training	• None • No references • Post-sorting naming of groups may be carried out	• Placing products on sheet of paper according to similarities and dissimilarities • Spatial co-ordinates are recorded • All products must be presented simultaneously	• None	• Matrix of X and Y co-ordinates • PCA • MFA • INDSCAL • DISTATIS
Polarized sensory positioning (PSP) [pivot profile, triad-PSP, polarized projective mapping]	• 10–15 trained • 30–60 naive • No requirement for screening or training	• None • References used for poles	• Rating dissimilarity between sample and three reference samples • Typically unstructured 10 cm scales (anchored from exactly the same to extremely different) although any type of intensity scale can be used	• None	• PCA • MDS • ANOVA based on mixed models • GPA • STATIS • MFA • CA for triad-PSP

*Modified versions of rapid methods are in square brackets.

ANOVA, analysis of variance; CA, correspondence analysis; GPA, generalized Procrustes analysis; INDSCAL, individual difference scaling; MCA, multiple correspondence analysis; MDS, multidimensional scaling; MFA, multifactor analysis ; PCA, principal component analysis; SORT CC, common components and specific weights analysis; STATIS, Structuration des Tableaux A Trois Indices de la Statistique. (DISTATIS, handles multiple distance matrices).

methods. The time required varies depending on the objective of the study, the number and type of products, and the number and type of attributes. For example, traditional methods can be completed in a minimum of 1–2 sessions for situations involving a short evaluation time but more sessions might be needed for products with a longer lasting experience, such as toothpaste, chewing gum, etc., or where a product is to be assessed in detail. The consensus method may take slightly longer for each sample than other traditional methods, as scores for each attribute need to be discussed and agreed upon. Rapid methods can be completed in one session, with the exception of free choice profiling (FCP) and flash profiling, which require a minimum of two sessions.

In general, the information in these tables shows that the time/resources available for training, individual preferences on type of sensory language/ information obtained and overall philosophy are key factors to select the most appropriate descriptive method for a company. For example, if a practitioner has limitations on time and resources, seeks to obtain only general product information (i.e. does not require detailed and accurate attribute information and high discrimination) and has the statistical tools required by specific methods, a non-traditional/rapid descriptive method might be the best selection in this case. Conversely, if a practitioner requires detailed and accurate product information with the ability to detect small product differences (high discrimination) and has the resources and the time to complete a training programme, one of the traditional descriptive methods might be the best option.

The discussion that follows provides additional details and information in the key characteristics chosen for comparison for all methods presented in this book.

20.4 Comparison of Methods Based on Key Characteristics

20.4.1 Comparison Based on Philosophies

Two philosophical concepts are presented that differentiate descriptive methods:

- traditional versus non-traditional/rapid approaches
- technical versus semi-technical/consumer-based approaches.

20.4.1.1 Traditional versus Non-Traditional/Rapid Descriptive Philosophies

Traditional descriptive methods implement a training programme to develop a panel capable of providing detailed qualitative and quantitative product information. The trained panel develops group terminology/lexicons and, in most cases, a ballot to evaluate the intensity of product attributes.

Non-traditional/rapid descriptive methods do not implement a training programme nor provide as detailed product information as that generated by traditional methods. Therefore, non-traditional/rapid descriptive methods require less time to implement and thus are less expensive. In rapid methods, assessors either:

- are allowed to choose their own terminology (or evaluation criteria) without any group consensus/standardization of terminology, or
- holistically focus on the products' similarities and differences to sort or group products.

In general, no specific quantitative information on the products' attributes is obtained by rapid methods. They provide broad product information and the products' differences/similarities are depicted in multidimensional groups or maps.

The descriptive methods presented in this book fall in these two categories, as shown in Table 20.3.

As discussed in section 20.3, the main advantage of non-traditional/rapid methods is a short and less expensive programme implementation, since no training is completed. However, the main weaknesses of these methods, as a result of not completing a training programme, are non-detailed product information, lower discrimination, need for advanced statistical analyses and sometimes a cumbersome process in data interpretation.

Conversely, traditional methods' main advantages are the ability to provide more detailed product information, robust/reliable data, and thus higher discrimination. The main disadvantage is that the training programme implementation makes these methods more expensive and involved.

20.4.1.2 Technical versus Semi-Technical/Consumer-Based Descriptive Methods

Another philosophical difference among descriptive methods is in the type of sensory language and information generated.

Table 20.3 Comparison of methods based on philosophies.

Traditional methods	Non-traditional/rapid methods
Flavor and texture profile methods	CATA
Consensus	Free-choice profiling
QDA	Flash profiling
QFP	Projective mapping and sorting
Spectrum Method	Polarized sensory positioning
A^{5}daptive Profile Method	
Ranking/rank-rating	

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Technical descriptive methods strive to develop technical terminology or lexicons to provide detailed product information and guidance to the user. Qualitative references, which are often ingredients and chemical references, are used to exemplify attributes and achieve a technical panel vocabulary. In several methods, the panel leader has an active participation in the language development to ensure that it is technical and appropriate.

Only some of the traditional descriptive methods are considered technical, and typically their training programmes are long. On the other hand, semi-technical/consumer-based descriptive methods philosophically believe that the sensory language should consist of the words/terms that the assessors are comfortable using. In some methods (e.g. QDA), qualitative references might be used by the panel as guidelines in the selection of terms, but ultimately the assessors are allowed to use the language they choose. As a result, the lexicon often might include consumer terms and might not be as technical.

Also all non-traditional/rapid methods presented in this book are considered semi-technical/consumer-based descriptive methods since participants choose their own evaluation criteria/own words, or are provided with a list of consumer terms, or products are evaluated holistically without focusing on selected product characteristics/attributes. Thus, no technical and specific information is obtained in non-traditional/rapid methods, unless trained assessors participate.

Table 20.4 shows the classification of the methods presented in this book based on this criterion.

20.4.2 Comparison Based on Type of Results/Output

One of the most important aspects sensory practitioners should consider in selecting the most appropriate method for their situation is the type of results provided by each method.

Table 20.5 shows that four different types of descriptive results are obtained depending on the method.

Table 20.4 Comparison of methods based on their technical nature.

Technical-based methods	Consumer-based/semi-technical methods
Flavor and texture profile methods	QDA
Consensus	Ranking/rank-rating*
QFP	CATA*
Spectrum Method	Free choice profiling
A^5daptive Profile Method	Flash profiling
	Projective mapping and sorting
	Polarized sensory positioning*

*More technical language is obtained in these methods when trained assessors participate.
CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Table 20.5 Comparison of methods based on type of results/output.

Panel-generated attributes and consensus intensities	Panel-generated attributes and individual intensities	Panel-generated attributes and ranks or frequencies	Multidimensional maps and other configurations
Flavor and texture profile methods Consensus A[5]daptive Profile Method (*when consensus is used*)	Texture profile method (*with individual scores*) QDA QFP Spectrum Method Ranking/rank-rating (*rating component*) A[5]daptive Profile Method	Ranking/rank-rating (*ranking component*) CATA	Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

All traditional descriptive methods provide attribute information. The difference among them, and outlined in this classification, is in the type of quantitative results obtained.

- Several traditional descriptive methods (e.g. the flavor profile method) use consensus to gather attribute intensities. In these methods, no statistical analysis is completed and each attribute's intensity is reported as one/consensus intensity.
- Most traditional descriptive methods collect and analyse individual scores/intensities.
- A traditional descriptive method can also use ranking to evaluate differences among products in the attribute's strength.

Non-traditional/rapid methods (e.g. free choice profiling, sorting) distinguish themselves from traditional approaches in the type of results obtained. In general, these methods' results are reported as multidimensional maps and other configurations, which allow spatial/visual representation of the differences and similarities among samples. The type of map and configuration obtained is method-dependent. Each of the rapid methods' chapters explains how these maps are generated and the information provided.

20.4.3 Comparison Based on Overall Needs for Programme Implementation

A key parameter to assess in choosing a descriptive method is the programme's implementation needs.

Table 20.6 shows the main needs for each method. The methods with more requirements are more expensive to implement. In general, the traditional descriptive approaches (listed first in the table) tend to be the more expensive methods to implement given their needs.

Table 20.6 Comparison of methods based on overall needs for programme implementation.

Method	Need					
	Panel training	Panel leader	References		Advanced statistical analyses	Other needs
			Qualitative	Quantitative		
Flavor profile (FP) and texture profile (TP) methods	Y	Y	Y	Y	None (FP) No (TP – when individual data are collected)	
Consensus	Y	Y	Y	Y	None	
QDA	Y	Y	Y	No	No*	
QFP	Y	Y	Y	Occasionally	No*	Input from flavourists
Spectrum Method	Y	Y	Y	Y	No*	
A^{5}daptive Profile Method	Y	Y	Y	Y	No*	Preliminary assessment and programme customized design
Ranking/rank-rating	Y	Y	Y	Y(optional)	No*	
CATA	No	No	No	No	No	
Free choice profiling	No	No	No	No	Y	Different assessors typically recruited for each study
Flash profiling	No	No	No	No	Y	
Projective mapping and sorting	Orientation	No	No	No	Y	
Polarized sensory positioning	Orientation	No	Y (poles)	No	Y	

*Multivariate statistical analysis is also used by these methods in specific projects and when applicable. However, for routine evaluations, only common univariate statistical analyses are used.

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

20.4.4 Comparison based on Assessors' Characteristics

20.4.4.1 Type of Assessors and Selection/Screening

Table 20.7 gives a comparison of methods based on the type of assessors needed and whether a screening process is implemented or not.

Methods requiring a training programme incorporate a selection/screening process to choose the most appropriate participants to be trained. Thus, a structured selection/screening process is followed by all traditional descriptive methods. These methods recruit, select and train employees or consumers/residents from the community if they meet the selection criteria and have passed the acuity and other screening tests.

Non-traditional descriptive programmes do not incorporate a screening process, mainly recruit their participants based on availability and interest, and do not train participants. However, several researchers report having used trained assessors in non-traditional methods.

20.4.4.2 Recommended Number of Assessors

Each method provides a recommendation on number of assessors to train (if applicable) and use in product evaluation, as described in the corresponding chapter.

Table 20.8 classifies methods based on recommended number of assessors used in product evaluations (Note: More assessors than indicated in the table would be recruited, screened and trained to have a larger pool of assessors).

As with other characteristics, this element is also method dependent. Rapid/non-traditional methods require a large number of assessors because no training or standardization of techniques is completed. Therefore, a large number of assessors is needed to compensate for the large variability among them. Conversely, all traditional methods require a smaller number of assessors in the descriptive evaluation of products because the assessors' variability is decreased by the training and standardization.

Table 20.7 Comparison of methods based on type of assessors and necessary screening.

Trained assessors Selection/screening needed	Untrained assessors No screening needed
Flavor and texture profile methods	CATA
Consensus	Free choice profiling
QDA*	Flash profiling
QFP*	Projective mapping and sorting
Spectrum Method	Polarized sensory positioning
A^{5}daptive Profile Method*	(*trained assessors can be used*)
Ranking/rank-rating*	

*Screening completed on specific products or attributes when the training is category/product specific.
CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Table 20.8 Comparison of methods based on number of assessors.

Less than 8	12–20	Over 20
Flavor profile method Consensus	Texture profile method (*less when consensus is used*) QDA QFP Spectrum Method A^5daptive Profile Method Rank-rating	CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning (*less with trained assessors*)

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

The original flavor profile method and the consensus method use a small number of assessors in product evaluations. These two methods are also technical approaches, and thus use a smaller panel. However, the main reason for including such a small number of assessors is because of the discussions and consensus process intrinsic to these methods. A very small panel of well-trained assessors is needed to be able to effectively reach consensus.

20.4.5 Comparison Based on Panel Leader's Role

Non-traditional descriptive methods do not require a panel leader, mainly because no training is completed. Thus, only a sensory professional or staff member is needed to administer the programme and tests. This professional interacts with assessors in their recruitment and when giving them directions for the task/test.

Traditional methods, on the other hand, require a panel leader. This professional might be involved in most panel activities, such as programme design and planning, panel recruitment and screening/selection, training and monitoring programmes, and project design and execution.

The role of the panel leader in the training programme differs among traditional descriptive methods. In most traditional methods, the panel leader (or another expert/professional) is in charge of providing the panel with all the technical knowledge required by the method and sensory dimension being learned (e.g. flavour or skinfeel), definitions, elements of the product evaluation protocol, quantitative references (if applicable), etc. Therefore, the panel leader (or the expert who teaches a specific topic to the panel in methods such as the original flavor profile method or quantitative flavour profiling (QFP)) takes an active participation in the training. Furthermore, in highly technical methods, such as in the texture profile method, Spectrum™ Method and A^5daptive Profile Method®, the panel leader has an active role, guiding the panel in making decisions regarding the selection of sensory terms and evaluation protocols and ensuring the development of the most technical and adequate lexicons.

Conversely, the panel leader's role in the QDA method is very different. He/she only acts as a moderator/facilitator to guide panel discussions in the development of descriptive protocols and sensory vocabulary. Therefore, all decisions are made by the panel, including the terms and evaluation procedures. Thus the language might not be as technical as it is in other traditional methods.

Table 20.9 compares the panel leader's role for different methods.

20.4.6 Comparison Based on Training Characteristics

20.4.6.1 Training versus Participation/No Training

One of the main differences between the traditional and non-traditional/rapid methods is the completion of a training programme.

Rapid methods are less expensive and time-consuming because they do not implement a training programme. In contrast, all traditional methods require a training programme, making them more expensive to implement (see Table 20.10 for a comparison).

Table 20.9 Comparison of methods based on panel leader.

Panel leader facilitates discussions	Panel leader guides discussions and teaches	No panel leader. Staff member provides instructions/supervises
QDA Ranking/rank-rating	Flavor profile method (*teaching also provided by other professionals*) Consensus Texture profile method QFP (*mainly teaching standardized language*) Spectrum Method A^5daptive Profile Method	CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning (*less with trained assessors*)

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Table 20.10 Comparison of methods based on whether training is required.

Panel training required	No panel training
Flavor and texture profile methods	CATA
Consensus	Free choice profiling
QDA	Flash profiling
QFP	Projective mapping and sorting
Spectrum Method	Polarized sensory positioning
A^5daptive Profile Method	
Ranking/rank-rating	

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

The differences in philosophy and methodology among the traditional descriptive methods influence their training programme's type and depth, as explained in section 20.4.6.2.

20.4.6.2 Nature/Level and Length of Training

Traditional descriptive methods, which require a training programme, differ in the nature/depth of their training programmes (Table 20.11). Several methods (e.g. original and modified profile methods), which are considered the most technical approaches, require a lengthy in-depth training. This type of programme is necessary to be able to cover all these methods' technical concepts. Through their in-depth training programmes, traditional descriptive methods' goal is to achieve assessors' proficiency in both the qualitative and quantitative components. In addition, in several traditional descriptive methods, the use of quantitative/intensity references adds time and another layer of training. Some traditional methods which do not use quantitative references, such as QDA, require shorter training programmes.

20.4.7 Comparison Based on Type of Sensory Terminology

20.4.7.1 Generation of Sensory Terminology/Lexicon

Descriptive methods fall into four categories based on the type of sensory terminology generated: no generated terminology, self-defined terminology, terminology provided to panel, and group/consensus terminology (Table 20.12).

20.4.7.1.1 No Generated Terminology

Several of the rapid/non-traditional descriptive methods presented in this book do not generate or use any sensory terms in product evaluations. For example, in the projective mapping and sorting methods, assessors sort or position products based on overall/holistic differences and similarities, without requiring the selection or specification of product attributes/terms. However, groupings may be named afterwards.

20.4.7.1.2 Generated Terminology

Besides the methods mentioned in section 20.2.7.1.1, all other descriptive methods require the generation of a sensory language for the evaluation of the products' sensory characteristics. These methods differ in the way the terminology is generated, as follows.

Table 20.11 Comparison of methods based on nature of training.

Focused and shorter training	In-depth/longer training
QDA	Flavor and texture profile methods
QFP (*moderate*)	Consensus
Ranking/rank-rating	Spectrum Method
	A^5daptive Profile Method (*shorter for a product-specific programme*)

QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Table 20.12 Comparison of methods based on generation of sensory terminology.

No terminology (no specified attributes)	Individual terminology (own/self-defined terms – no group terminology)	Terminology provided to assessors	Group/consensus terminology
Projective mapping and sorting Polarized sensory positioning	Free choice profiling Flash profiling	CATA (*list of attributes provided*)	Flavor and texture profile methods Consensus QDA QFP (*based on standardized language developed by flavourists*) Spectrum Method A^{5}daptive Profile Method Ranking/rank-rating (*attributes might be selected by sensory professional*)

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

- A few rapid/non-traditional descriptive methods, such as free choice and flash profiling, do not develop or use group/consensus sensory language. Instead, assessors develop and use their own self-defined language.
- In CATA the terminology is provided to assessors.
- All traditional descriptive methods, which form the fourth category in this classification, develop and use group/consensus sensory language. During the training programme of traditional descriptive methods, assessors develop the knowledge and skill to verbalize perceptions, generate individual terminology and then participate in a discussion to reach consensus on the group lexicon to be used in product evaluation. Traditional methods differ in the philosophy and type of group language developed, as explained in these methods' corresponding chapters.

20.4.7.2 Type of Sensory Terminology

This section furthers the discussion of concepts presented in section 20.4.1.2.

This discussion and the classification in Table 20.13 only apply to those methods generating panel/group sensory terminology, which can be technical or semi-technical/consumer based, as follows.

20.4.7.2.1 Technical Sensory Terminology

This language type is developed with a well-trained panel, which uses an extensive array of qualitative references and is given guidance during lexicon development. The goal of the technical methods listed in Table 20.13 is to develop panels capable of generating a detailed and technical sensory language for product evaluation.

20.4.7.2.2 Semi-Technical/Consumer-Based Sensory Terminology

The QDA method develops semi-technical/consumer-based terminology because the panel leader acts only as a facilitator/moderator and assessors are allowed to use the terms they feel most comfortable with. QDA panels do

Table 20.13 Comparison of methods based on type of sensory terminology.

Technical terminology	Semi-technical, consumer-based terminology	Consumer terminology
Flavor and texture profile methods Consensus QFP Spectrum Method A5daptive Profile Method Rank/rank-rating (*if profile-based approach used*)	QDA Rank/rank-rating	Free choice profiling Flash profiling CATA (*unless carrying out with trained panel*)

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

not participate in in-depth terminology development sessions such as the ones technical methods conduct, where an extensive array of references is reviewed and lengthy group discussions are held. Therefore, semi-technical/consumer-based terms are generated and used by QDA panels.

Free choice and flash profiling methods also develop and use consumer-based language, because each assessor is allowed to use his/her own terms. No discussion is conducted on the appropriateness of the terms or to reach consensus/standardized language among assessors. In general, the terms generated by these two rapid methods tend to be general/non-technical and thus are considered consumer-based terms.

20.4.7.3 Singular versus Complex Terms/Attributes Used by Technical Descriptive Methods

This classification (Table 20.14) only applies to traditional descriptive methods generating technical terminology.

In traditional descriptive analysis, the sensory language necessary to characterize the products' sensory properties is developed and used. Technical traditional methods strive to develop the most technical/detailed language.

Table 20.14 Comparison of methods based on complexity of attributes.

Technical methods mainly using singular sensory terms	Technical methods using singular and complex sensory terms (e.g. balance, blend, complexity)
Texture profile method Consensus QDA QFP Spectrum Method	Flavor profile method A5daptive Profile Method

QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Two technical methods, the original flavor profile method and the A^{5}daptive Profile Method, take the approach that in flavour and fragrance evaluation, there is the need to evaluate complex/holistic attributes in addition to the singular and very detailed technical terms. Sometimes, these complex/holistic attributes are the only ones distinguishing products and offering information that cannot be captured by singular terms alone (see Chapter 7). Examples of these complex/holistic terms include balance, blendedness, fullness, etc.

The evaluation of these terms is complex. The training exercises focused on developing this knowledge and skill are only introduced after a trained panel has had sufficient experience in product evaluation.

20.4.8 Comparison Based on Use of Training References

20.4.8.1 Qualitative References

All traditional descriptive methods use qualitative references in their training programmes to aid in the development of the sensory language. Methods differ in the philosophy and type of qualitative references used. In general, technical methods such as the original flavor and texture profile methods, Spectrum Method and A^{5}daptive Profile Method use an extensive array of qualitative references for panels to achieve the desired highly technical language.

Polarized sensory positioning is the only rapid descriptive method that uses qualitative references. These references are the poles that must be presented to assessors to complete product evaluations. Table 20.15 categorizes the descriptive methods based on the use of qualitative references.

20.4.8.2 Quantitative References

This discussion only applies to traditional descriptive methods. Rapid methods do not use quantitative references, given their philosophy and since no training or standardization is completed.

Table 20.15 Comparison of methods based on use of qualitative references.

Use of qualitative references	No use of qualitative references
Flavor and texture profile methods	CATA
Consensus	Free choice profiling
QDA	Flash profiling
QFP	Projective mapping and sorting
Spectrum Method	
A^{5}daptive Profile Method	
Ranking/rank-rating	
Polarized sensory positioning (*poles*)	
CATA (*if used with a trained panel*)	

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Traditional descriptive methods fall into two categories based on their philosophy and views regarding the use of quantitative references: Some traditional methods use quantitative references extensively, while others do not. Table 20.16 shows the difference between the two groups of methods based on the use of intensity references.

As explained in Chapter 7, one of the noteworthy contributions of the flavor and texture profile methods to the development of descriptive analysis was the use of quantitative/intensity references to rate intensities. Several traditional methods (e.g. Spectrum Method and A^{5}daptive Profile Method) adapted this philosophy and use quantitative references.

The use of quantitative/intensity references is controversial, since this practice has advantages and disadvantages, as explained in Chapter 7. Therefore, some traditional methods, such as QDA, do not use these references. In this method, assessors may instead be familiarized with the range of attribute intensities of the product set prior to evaluation.

20.4.9 Comparison Based on Type of Data Collection/ Scoring Approach

A very important consideration in choosing the most appropriate descriptive method for a user is the type of sensory information obtained. Methods differ vastly in philosophy and methodology, and thus in the type of results delivered (Table 20.17). Therefore, a practitioner should carefully assess the characteristics of the data generated by each method, to ensure that the chosen method provides the desired information.

For example, a practitioner should understand the benefits that consensus results offer relative to the fact that no individual data are obtained and thus no statistical analysis is performed. Similarly, a practitioner should understand the type of broad/general product information that is obtained by many of the rapid/non-traditional methods, compared to the detailed attribute information provided by traditional methods.

Table 20.16 Comparison of methods based on use of quantitative references.

Use of quantitative references	No use of quantitative references
Flavor and texture profile methods	QDA
Consensus	QFP (*uses occasionally*)
Spectrum Method	CATA
A^{5}daptive Profile Method	Free choice profiling
	Flash profiling
	Projective mapping and sorting
	Ranking/rank-rating (*optional use*)
	Polarized sensory positioning

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

Table 20.17 Comparison of methods based on type of data collection/scoring.

Type of data collection/ scoring	Method(s)	Comments
Consensus ratings of panel/ group attributes or group discussion of differences/ similarities	Flavor profile method Texture profile method (*might use individual scores*) Consensus A^{5}daptive Profile Method (*when consensus is used*)	No statistical analysis completed
Rating of panel/group attributes using chosen/ specific scale by each method	Texture profile method (*with individual scores*) QDA QFP Spectrum Method A^{5}daptive Profile Method Rank-rating (*rating data*)	Panel data are statistically analysed. The results can be semi-technical/ consumer based to very technical and precise
Ranking of panel/group attributes, or ranking combined with rating of panel/group attributes	Ranking/rank-rating	Data from studies using different sample sets can be difficult to combine
Rating of own/self-defined attributes	Free choice profiling	More difficult to interpret, as data are idiosyncratic
Ranking of own/self-defined attributes	Flash profiling	More difficult to interpret, as data are idiosyncratic
Selecting attributes from pre-defined list	CATA	Data produced is nominal only
Sorting of products in mutually exclusive groups based on perceived product similarities	Sorting	Data are holistic and lack detail
Positioning of products on a sheet of paper according to the products' similarities or dissimilarities	Projective mapping	Data are holistic and lack detail
Comparison of samples based on general differences (no indication of attributes) with a set of fixed references, called poles	Polarized sensory positioning	Two types: (1) Indication of the overall degree of difference between each sample and the poles (2) Triadic approach: Indication as to which of the poles a sample is most similar to and most different from Data are holistic and lack detail

CATA, check-all-that-apply; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling.

The type of information obtained is based on the method's philosophy and methodology. In addition, it should be noted that different resources, costs and timing are required by each method to deliver its unique sensory information. Thus it is essential to consider the resources needed by each descriptive method when making a selection based on information desired.

20.4.10 Comparison Based on Data Analyses completed

Methods using consensus in the data collection phase do not require any statistical data analysis. All other methods use specific statistical approaches to analyse their data.

Traditional methods use a variety of uni- and multivariate approaches in the analysis of their data. Rapid/non-traditional methods require more advanced and specific data analyses. Therefore, practitioners selecting rapid methods must take into consideration the statistical tools needed to complete the data analyses required by these methods.

Details on the data analyses and their requirements can be found in each method's corresponding chapter.

Table 20.18 summarizes the main analyses used by each method.

20.4.11 Comparison of Advantages/Strengths and Disadvantages/Weaknesses

Several of the methods' advantages and disadvantages have already been mentioned in previous sections. This section addresses strengths and weaknesses more comprehensively and a summary is presented in Table 20.19. In order to put these advantages and disadvantages in the right perspective, they should be assessed vis-à-vis information previously given on the methods' characteristics.

The main reported strength of the rapid methods is their speed. However, their speed of data collection is on a par with traditional methods once these methods' training phase is completed. In fact, because rapid methods recruit new assessors for every project, additional time is needed to complete a project, compared to the shorter time to complete a project by a standing trained panel from traditional methods. The main disadvantage of newer/rapid methods is their less detailed data, partly due to the use of untrained consumers. It is, however, possible to carry out these methods with trained assessors, which gives more reliable results.

The main reported strength of traditional methods is the quality of the data obtained, specifically more detailed and robust. The main disadvantage of these methods is the longer time and higher cost to implement. However, as indicated above, once the panel is trained in any of the traditional methods, it can undertake many studies, which shortens the time for test execution.

Table 20.18 Comparison of methods based on data analyses.

Type of analysis	Methods	Comments
No data analysis	**Those using consensus:** Flavor profile method Texture profile method (*analyses data with individual scores*) Consensus A^{5}daptive Profile Method (*when consensus is used*)	
Correspondence analysis, multi-block	CATA	
Univariate data analyses (*routine evaluations*) and a variety of multivariate data analyses depending on project	Texture profile method (*with individual scores*) QDA QFP Spectrum Method A^{5}daptive Profile Method Ranking/rank-rating (*rating data*)	Examples of univariate analyses: t-test, ANOVA, correlation, etc. Examples of multivariate analyses: PCA, factor analysis, cluster analysis, etc.
Friedman, R-Index analysis	Ranking/rank-rating (*ranking data*)	
PCA and GPA	Free choice and flash profiling	
PCA, DISTATIS, MFA, INDSCAL, etc.	Projective mapping Flash profiling	
MDS	Sorting	Other analyses used might include: MCA, DISTATIS, SORTCC, etc.
A combination of the following: ANOVA, PCA, MDS, MFA, GPA, CA, STATIS	Polarized sensory positioning (*poles*)	

ANOVA, analysis of variance; CA, correspondence analysis; CATA, check-all-that-apply; GPA, generalized Procrustes analysis; INDSCAL, individual difference scaling; MCA, multiple correspondence analysis; MDS, multidimensional scaling; MFA, multifactor analysis; PCA, principal component analysis; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling; SORTCC, Common Components and Specific Weights Analysis; STATIS, Structuration des Tableaux A Trois Indices de la Statistique (DISTATIS, handles multiple distance matrices).

20.4.12 Comparison Based on Applications

Whereas traditional methodologies can be used for a wide range of applications, newer methodologies have more limited applications and have not yet been tested in a wide range of situations. However, taken as a whole, descriptive analysis is a flexible tool for a variety of uses. Table 20.20 gives a summary of how each of the descriptive methods covered in this book are best applied.

Table 20.19 Comparison of methods based on advantages/strengths and disadvantages/weaknesses.

Method	Strengths/advantages	Weaknesses/disadvantages
Consensus method	• Flexible and immediately adaptable • Cross-pollination of ideas • Inclusion of time sequence information • Reduced data collection time	• Potential for bias due to dominant assessors • Individual variation between panellists cannot be accounted for • No statistical analysis
Flavor profile method	• Qualitatively and quantitatively strong • High discrimination • Group discussion after evaluation allows further exploration of attributes and differences often missed in individual evaluations • Provides order of appearance • Stable over time • Easier for R&D to understand	• Long training • Potential for bias due to dominant assessors • Individual variation between assessors cannot be accounted for • No statistical analysis
Texture profile method	• Qualitatively and quantitatively strong • High discrimination • Stable over time • Easier for R&D to understand • Data can be analysed statistically when numerical or unstructured scales are used	• Long training • Does not cover full range of modalities • Statistical analysis only possible if scales used
QDA	• Shorter than profiling techniques, with quicker lexicon development • No requirement to calibrate assessors on intensity • No requirement for comprehensive reference library • Use of product-specific intensity scale might give greater discrimination • More robust data through use of replication • Good discrimination • Consumer/semi-technical language • Easier for marketing to understand	• Less technical language • Longer data collection due to requirement for replications • Not as easy for R&D to understand/use due to the semi-technical language used • Use of attribute/product-specific scales may make it more difficult to compare data across studies, attributes and products
Spectrum Method	• Qualitatively and quantitatively strong • More robust data through use of replication • High discrimination • Statistical analysis possible • Stable over time • Can combine data over studies • Easier for R&D to understand	• Long training • Time to develop and use qualitative and quantitative references • Use of universal scales may give lower discrimination for similar products (same category) • Longer data collection due to requirement for replication • Not as easy for marketing to understand • Expensive

Table 20.19 (Continued)

Method	Strengths/advantages	Weaknesses/disadvantages
QFP	• As for Spectrum Method • Universal flavour language • Precise and comprehensive for flavour	• As for Spectrum Method • Proprietary to company • Highly technical language may be difficult to relate to consumer perceptions • Restricted to flavour
A[5]daptive Profile Method	• Adapts approach, lexicon type, scale and references to objectives and unique category/products and variables • High discrimination • Qualitatively and quantitatively strong • Can include comprehensive/holistic product evaluations (e.g., when applicable/appropriate, holistic/complex attributes or in-use experience/ evaluations) • Data from different panels/locations and time intervals can be compared • Easier for R&D to understand	• Long training for universal panels • Time required to develop customized references
Ranking and rank-rating descriptive analysis	• High discrimination • Focused	• Limited number of attributes and samples may be included in studies • Limitations on number and type of products (see disadvantages on flash profiling – same weaknesses apply for this method) • Cannot combine data from different studies
CATA/TATA	• Simple • Quick • No need for alignment on attribute name and intensity	• Large number of assessors needed • Only produces nominal data • Typically recruits new assessors for each study • Researcher needs to select a limited number of attributes, which might be problematic • Assessors may not interpret descriptors in same way
Free choice profiling	• Faster and cheaper than traditional methods (no training programme) • No need to define attributes • No need for alignment on attributes and intensity • Produces product description • Can compare different consumer groups directly	• Typically recruits new assessors for each study • Less detailed and accurate • Assessor may not use descriptors in same way • Lower discrimination • Complex statistical analysis • Cumbersome and difficult data interpretation • More difficult to obtain direction for R&D

(Continued)

Table 20.19 (Continued)

Method	Strengths/advantages	Weaknesses/disadvantages
Flash profiling	• As for free choice profiling • Variability due to use of scale is eliminated	• As free choice profiling • Limitations on number and type of product due to memory load, fatigue, high carry-over and simultaneous presentation (e.g. difficult with hot products, frozen products, personal care products)
Sorting	• Rapid • Much less time-consuming than traditional methods and newer methods that use direct (dis)similarity scaling, such as PSP • Less expensive • No need to specify attributes • Very simple task • Can be used with trained and untrained assessors • Since attributes are not specified method can be used with assessors less able to verbalize, e.g. naive consumers and children	• Limitations on number and type of products (see disadvantages on flash profiling – same weaknesses apply for this method) • Typically need to recruit new assessors for each study • Lack of detailed product information • Post-sorting naming/description idiosyncratic; might contain non-sensory terms and data are difficult to analyse and interpret • Low discrimination and lack of detail • Cannot combine studies • Complex statistical analysis • Cumbersome and difficult interpretation • Limited guidance for R&D
Projective mapping	• Rapid • Much less time-consuming than traditional methods and newer methods that use direct (dis)similarity scaling, such as PSP • Less expensive • No need to specify attributes	• As for sorting • Difficult to perform for assessors who are less adept at handling spatial information • Constrains assessors to using two dimensions to discriminate between the products
Polarized sensory positioning	• Rapid • Less expensive • No need to define attributes • Results from different studies using same poles can be combined	• Identification of poles • Needs stable references • Typically recruits new assessors for each study • Low discrimination and lack of detail • Complex statistical analysis • Cumbersome and difficult data interpretation • Limited guidance for R&D

CATA, check-all-that-apply; PSP, polarized sensory positioning; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling; TATA, tick-all-that-apply.

Table 20.20 Comparison of methods based on applications.

Application	More suitable	Less/not suitable
Detailed product description (e.g. for target definition)	Consensus Flavor and texture profile methods QDA QFP Spectrum Method A^5daptive Profile Method	Ranking/rank-rating CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning
Detailed comparison of product characteristics and discriminative (i.e. identifying product differences)	Consensus Flavor and texture profile methods QDA QFP Spectrum Method A^5daptive Profile Method Ranking/rank-rating	CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning
Product/category mapping (e.g. for market overview, opportunity definition)	All methods	
New product development (NPD) guidance	All methods (*Quicker in new methods*)	
Late-stage NPD – more detail required	All traditional methods	All new/rapid methods
Product optimization	All traditional methods	All new/rapid methods. NB: All rapid methods are unsuitable to measure very small differences
Protection of competitive advantage, e.g. patent protection	All traditional methods	All new/rapid methods
QA/QC where one sample available at a time	Consensus Flavor and texture profile methods QDA QFP Spectrum Method A^5daptive Profile Method CATA	Ranking/rank-rating Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning
QA/QC where multiple samples available at a time	As above QA/QC entry plus Ranking/rank-rating	Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning
Use of consumer language	Free choice profiling Flash profiling CATA QDA (*to some extent*)	All other methods
Product mapping with consumer data	All new/rapid methods	All traditional methods
Product mapping with detailed descriptive data	All traditional methods	All new/rapid methods

(*Continued*)

Table 20.20 (Continued)

Application	More suitable	Less/not suitable
Linking to instrumental physicochemical data and/or consumer data for modelling (e.g. preference mapping, RSM)	Consensus Flavor and texture profile methods QDA QFP Spectrum Method A^{5}daptive Profile Method Ranking/rank-rating	CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning
Advertising claim substantiation	All traditional methods	CATA Free choice profiling Flash profiling Projective mapping and sorting Polarized sensory positioning

CATA, check-all-that-apply; QA, quality assurance; QC, quality control; QDA, quantitative descriptive analysis; QFP, quantitative flavour profiling; RSM, response surface methodology.

20.5 Summary

Descriptive analysis is a very useful, successful and flexible tool in sensory evaluation, with many philosophies and methodological nuances. It has undergone continuous evolution in response to demands from research and industry.

Descriptive methods have developed different philosophies and approaches, and yield different types of data/information. Thus, today's users have a wide range of options, and careful consideration needs to be given to each of the descriptive approaches' key elements in selecting a method. This will enable users to best meet their particular objective, either by selecting an existing method or by developing their own customized technique.

20.6 Future Developments

Judging by the recently renewed interest in descriptive analysis, it seems clear that methods will continue to grow, driven by research, commercial needs and possible limitations. Speed combined with detail is highly desirable and it seems likely this area will be given more attention in the future. As ever, there is a strong focus on better prediction of consumer perception, which is leading to more in-context testing of products. In addition, developments in technology, such as mobile phone apps for real-time data collection and improved virtual reality devices, are contributing to the development and use of enhanced descriptive practices.

Descriptive analysis will continue to adapt and improve its capabilities to give users a wider range of choices to meet their specific needs.

References

Albert, A., Varela, P., Salvador, A., Hough, G. & Fiszman, S. (2011) Overcoming the issues in the sensory description of hot served food with a complex texture. Application of QDA®, flash profiling and projective mapping using panels with different degrees of training. *Food Quality & Preference*, **22**, 463–473.

Ares, G., Antúnez, L., Bruzzone, F. et al. (2015) Comparison of sensory product profiles generated by trained assessors and consumers using CATA questions: four case studies with complex and/or similar samples. *Food Quality & Preference*, **45**, 75–86.

ASTM (1992) *Manual on Descriptive Analysis Testing for Sensory Evaluation*. Philadelphia: American Society for Testing and Materials.

Bárcenas, P., Pérez Elortondo, F.J. & Albisu, M. (2004) Projective mapping in sensory analysis of ewes milk cheeses: a study on consumers and trained panel performance. *Food Research International*, **37**, 723–729.

Bruzzone, F., Ares, G. & Gimenez, A. (2012) Consumers' texture perception of milk desserts. II – Comparison with trained assessors' data. *Journal of Texture Studies*, **43**, 214–226.

Dairou, V. & Sieffermann, J.M. (2002) A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile. *Journal of Food Science*, **27**, 826–834.

da Silva, R.C.S.N., Minim, V.P.R., Carneiro, J.D.S., Nascimento, M., Lucia, S.M.D. & Minim, L.A. (2013) Qualitative sensory description using the Optimized Descriptive Profile: comparison with conventional and alternative methods for evaluation of chocolate. *Food Quality & Preference*, **30**, 169–179.

Dehlholm, C., Brockhoff, P.B., Meinert, L., Aaslyng, M.D. & Bredie, W.L.P. (2012) Rapid descriptive sensory methods – comparison of Free Multiple Sorting, Partial Napping, Napping, Flash Profiling and conventional profiling. *Food Quality & Preference*, **26**, 267–277.

Delarue, J. & Sieffermann, J.M. (2004) Sensory mapping using flash profile. Comparison with conventional descriptive method for the evaluation of the flavour of fruit dairy products. *Food Quality & Preference*, **15**, 383–392.

Lelièvre, M., Chollet, S., Abdi, H. & Valentin, D. (2008) What is the validity of the sorting task for describing beers? A study using trained and untrained assessors. *Food Quality & Preference*, **19**, 697–703.

Moussaoui, K.A. & Varela, P. (2010) Exploring consumer product profiling techniques and their linkage to a quantitative descriptive analysis. *Food Quality & Preference*, **21**, 1088–1099.

Risvik, E., McEwan, J.A. & Rodbotten, M. (1997) Evaluation of sensory profiling and projective mapping data. *Food Quality & Preference*, **8**, 63–71.

Valentin, D., Chollet, S., Lelievre, M. & Abdi, H. (2012) Quick and dirty but still pretty good: a review of new descriptive methods in food science. *International Journal of Food Science & Technology*, **47**, 1563–1578.

Varela, P. & Ares, G., (2012) Sensory profiling, the blurred line between sensory and consumer science. A review of novel method for product characterization. *Food Research International*, **48**, 893–908.

Williams, A.A. & Arnold, G.M. (1985) A comparison of the aromas of six coffees characterised by conventional profiling, free-choice profiling and similarity scaling methods. *Journal of the Science of Food & Agriculture*, **36**, 204–214.

Worch, T., Lê, S. & Punter, P. (2010) How reliable are the consumers? Comparison of sensory profiles from consumers and experts. *Food Quality & Preference*, **21**, 309–318.

Index

Descriptive Analysis in Sensory Evaluation, First Edition. Edited by Sarah E. Kemp, Joanne Hort and Tracey Hollowood.