Food Quality and Preference 71 (2019) 463–474

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Common roasting defects in coffee: Aroma composition, sensory characterization and consumer perception☆

Davide Giacalone[a,*], Tina Kreuzfeldt Degn[b], Ni Yang[c], Chujiao Liu[c], Ian Fisk[c], Morten Münchow[b,d,e]

[a] University of Southern Denmark, Odense, Denmark
[b] University of Copenhagen, Frederiksberg, Denmark
[c] University of Nottingham, Sutton Bonington, United Kingdom
[d] CoffeeMind Aps, Valby, Denmark
[e] Specialty Coffee Association, Essex, United Kingdom

ARTICLE INFO

Keywords:
Coffee
Clean cup
Food quality
Roasting defects
Consumers vs experts

ABSTRACT

The demand for high quality and specialty coffee is increasing worldwide. In order to meet these demands, a more uniform and standardized quality assessment of coffee is essential. The aim of this study was to make a sensory scientific and chemical characterization of common roasting defects in coffee, and to investigate their potential relevance for consumers' acceptance of coffee. To this end, six time-temperature roasting profiles based on a single origin *Arabica* bean were developed: one 'normal', representing a reference coffee free of defects, and five common roast defects ('dark', 'light', 'scorched', 'baked' and 'underdeveloped'. The coffee samples obtained from these beans were evaluated by means of (1) aroma analysis by Gas Chromatography-Mass Spectrometry (GC–MS), (2) sensory descriptive analysis (DA) by trained assessors, and (3) hedonic and sensory evaluation by consumers using a Check-All-That-Apply (CATA) questionnaire. Multivariate analyses of aroma, DA, and CATA data produced similar sample spaces, showing a clear opposition of the light roast to the dark and scorched roasts), with the normal roast having average values of key aroma compounds. The DA data confirmed this indications and showed the normal roast to have a balanced sensory profile compared to the other defects. Importantly, the normal roast was also significantly preferred in the consumer test ($N = 83$), and significantly associated to positive CATA attributes 'Harmonic', 'Pleasant', and 'Balanced'. Taken overall, the results provide a solid basis for understanding chemical and sensory markers associated with common roasting defects, which coffee professionals may use internally in both quality control and product development applications.

1. Introduction

1.1. Quality grading in the coffee industry vs. sensory analysis

With more than 2 billion cups consumed around the globe on an everyday basis, coffee is the most important beverage commodity traded in world markets (Nair, 2010; Ponte, 2002). Coffee consumption rates have increased 1–2% per year worldwide during the last decades, and the demand for specialty and high quality coffee has experienced the sharpest increase over the last years (Bhumiratana, Adhikari, & Chambers, 2011). Coffee quality is determined by numerous factors, such as the origin, post harvesting process and roasting of the coffee beans, different grinding and brewing methods, and serving conditions (Agresti, Franca, Oliveira, & Augusti, 2008; Baggenstoss, Poisson, Kaegi, Perren, & Escher, 2008; Brown & Diller, 2008; Lee & O'Mahony, 2002; Steen, Waehrens, Petersen, Münchow, & Bredie, 2017). In the coffee industry, several quality grading methods are used to classify the coffee at different stages of the production leading to a large number of classification systems related to plant type, origin, process treatment, defect count or bean size (Ribeiro, Augusto, Salva, Thomaziello, & Ferreira, 2009). Such methods, however, do not necessarily relate much to the eventual sensory quality of the brews. Therefore, sensory evaluation is a crucial tool to determine the drinking quality of the coffee.

In the coffee industry, sensory quality grading of brewed coffee, usually referred to as 'cupping', is conducted by expert 'cuppers' (Di Donfrancesco, Gutierrez Guzman, & Chambers, 2014; Feria-Morales, 2002). Typically, the procedure consists of tasting three to ten cups of the same coffee, prepared according to brewing conditions standardized

☆ Presented at SenseAsia.

* Corresponding author.
E-mail address: dg@iti.sdu.dk (D. Giacalone).

https://doi.org/10.1016/j.foodqual.2018.03.009
Received 3 July 2017; Received in revised form 12 March 2018; Accepted 12 March 2018
Available online 14 March 2018

with regard to temperature, contact time, water to coffee ratio, water quality and brewing method (ISO, 2008; SCAA, 2009). The cupping score sheet includes important flavor attributes for coffee, ranging from 0 to 10. In the current version, these are Fragrance/Aroma, Flavor, Aftertaste, Acidity, Body, Balance, Uniformity, Clean Cup, Sweetness, Defects, and Overall. However, unlike assessors in sensory descriptive analysis, cuppers do not rate the intensity but rather give a subjective appraisal of the individual attributes. For example, a high grade in "Acidity" would indicate how well the sourness of the coffee fits within the context of that particular coffee, regardless of absolute intensity. This blend of hedonic and analytical assessment marks a very important difference with scientific sensory analysis.

Generally speaking, expert cupping is more anchored in the product grading tradition than it is in proper sensory analysis. Indeed, in spite of their widespread application, from a scientific point of view current cupping procedures can be criticized on several grounds. Firstly, while sensory science methods rely of a larger pool of assessors to ensure robustness in the results, the coffee branch mostly relies on few expert tasters with years of experience. Oftentimes, only one or two tasters are responsible for the quality grading of a large number of coffee samples, sometimes amounting to more than 200 cups per day. Furthermore, the tasting is often not blind, meaning that the expert cuppers will typically have information about the coffee variety, supplier, etc. (Feria-Morales, 2002). Finally, until recently[1] there was no consensus regarding the sensory vocabulary or the use of particular scales, which still vary quite substantially depending on the country of origin of the coffee, and even on the individual company performing the cupping (Feria-Morales, 2002). Accordingly, two previous studies (Feria-Morales, 2002; Di Donfrancesco et al., 2014) have reported a poor correlations between results from 'cupping' (sensory evaluation by coffee experts) and descriptive sensory analysis with trained panelists, leading the authors to the conclusions that these two approaches are not interchangeable.

Another notable difference from sensory evaluation is that the quality judgments in cupping combine an overall quality scale (presumably reflecting consumer dislikes) with diagnostic information about defects, whereas in mainstream sensory evaluation these two functions (descriptive and consumer) would be typically separated in two distinct tests with different respondents (Lawless & Heymann, 2010). Assuming that the opinion of a single (or a few) expert can effectively predict consumer preferences is extremely questionable: in fact, particularly for coffee, recent evidence indicates that quality evaluations performed by coffee experts do not necessarily correspond to consumer preferences (Giacalone, Fosgaard, Steen, & Münchow, 2016).

A final problematic aspect with cupping protocols is the use of holistic quality attributes that rely substantially more on the experts' product knowledge and expectations regarding what is desirable in a coffee (similar to typicality judgments for wine), rather than on clearly defined sensory properties.

1.2. Motivation for the present study

One quality attribute that has recently gained attention is the concept of 'clean cup' or 'cleanliness', which has been used in the scientific literature as a sensory attribute for coffee (Ribeiro, Ferreira, & Salva, 2011; Ribeiro, Augusto, Salva, & Ferreira, 2012), and which is now included in the most important cupping protocols (ISO, 2008; SCAA, 2009). The attribute is not related to sanitary aspects (despite what the name might suggest), but is instead used as a quality attribute related to the absence of flaws/defects, which is purportedly associated to consumer preferences.

Situated within this context, the aim of this study was to understand the compositional and sensory basis of common roasting defects in coffee, as well as their relation with consumers' perception and preferences. Although defects in coffee may arise from different sources (indeed, concepts like 'clean cup' are most often associated with quality control of green coffee by experts (Feria-Morales, 2002)), we chose to focus on defects related to the roasting process resulting in off-flavours in the coffee brew, as previous research has shown that coffee's distinct aroma profile is very closely related to the time-temperature profiles used during the roasting (Baggenstoss et al., 2008; Fisk, Kettle, Hofmeister, Virdie, & Kenny, 2012; Masi, Dinnella, Barnabà, Navarini, & Monteleone, 2013; Yang et al., 2016).

Specifically, the chosen strategy was to focus on six distinct roasting profiles, obtained by varying time and temperature in the roasting process (see Section 2.1). One of them was roasted to represent a standard roast free of defects, according to recommendations of the Specialty Coffee Association of Europe (Münchow, 2016). The remaining five represented instead roast defects commonly found in the marketplace.

Moreover, this study extends a previous investigation in which the aroma volatile composition of coffee brewed from these six roasting profiles was documented (Yang et al., 2016). The goal of this earlier work was to investigate the formation of aroma compounds in these different time–temperature profiles, in order to identify marker compounds associated with each defect. Due to the complexity of aroma interactions, it is however uncertain whether those chemical changes correspond to perceptually relevant differences in the coffee. Thus, in the present paper, we continue this line of work by presenting the following new data and analyses:

1. A perceptual characterization of the same coffee samples by sensory descriptive analysis, in order to document the sensory properties associated with each roasting profile, as well as to look at the differentiation between the Normal roast and the defects;
2. An exploration of the relationship between the instrumental and sensory data, in order to evaluate the degree to which the aroma composition is predictive of the perceptual quality of the coffee;
3. A consumer test focusing on consumer perception and liking of coffee brewed from the different roasting profiles, carried out to understand whether absence of defects bears any correspondence with actual consumer preferences for coffee.

2. Materials and methods

2.1. Roasting profiles

The coffee used in the study was a single-origin washed Kenyan *Arabica* from the wet mill Ndaroini, from crop year 2012/2013 and 2013/2014. The beans were roasted using a Probat drum roaster (Probat-Werke, Germany) modified to include additional temperature sensors to monitor bean temperature. Due to the limited batch size of the Probat roaster (1 kg), the coffee was roasted on two separate occasions: one batch for the sensory evaluation, and one batch for the consumer and aroma analysis. The coffee beans samples were individually packed in odor-free air-tight package, and kept in a cold storage at 5 °C.

Six different roasting profiles were obtained by varying start temperature and roasting time. Five of the roasting profiles were created to obtain common roasting defects, whereas the last served as a control ('Normal') roast. These roasting profiles were developed by a panel of six coffee experts from the Specialty Coffee Association of Europe (SCAE), headed by the last author, to be part of SCAE roasting certification system, which provides a systematic framework for evaluation of roasting defects (Münchow, 2016). They were designed by modulating the roasting process on three different dimensions: roasting degree, time before 'first crack' (when a popping sound is first heard during roast), and time after first crack, which represent the roasting phases were the beans undergo significant the most significant chemical and

[1] Shortly after this study was conducted, a standardized vocabulary for coffee evaluation had just been released based on a comprehensive work carried out at Kansas State University (https://worldcoffeeresearch.org/work/sensory-lexicon/).

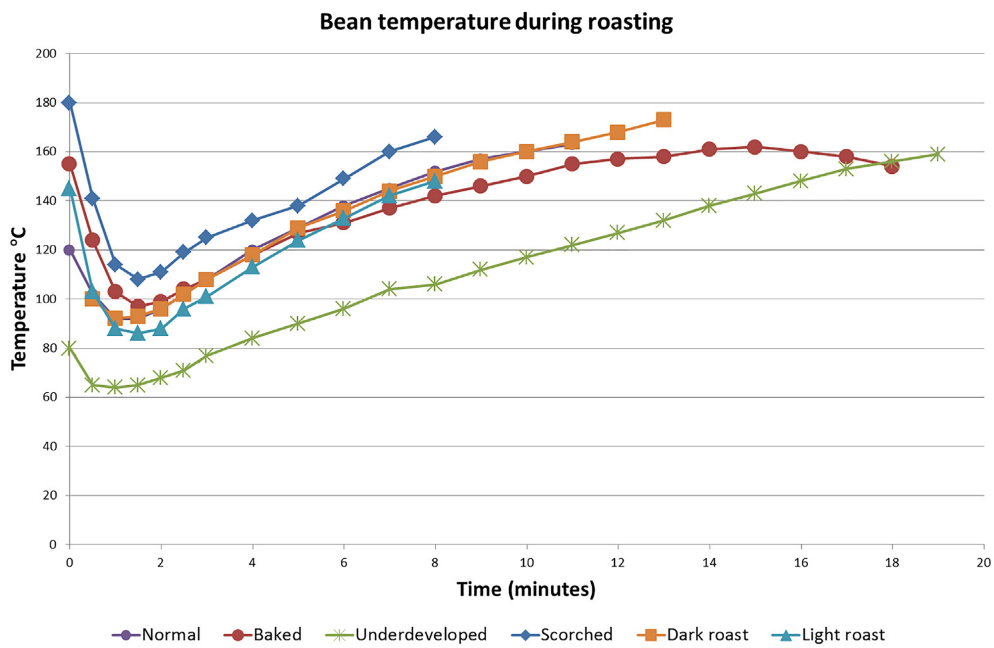

Fig. 1. Plot of temperature development over time for the six roasting profiles.

Table 1
Roasting conditions for the six roasting profiles.

Roasting profile	Starting temperature[a] (°C)	Developing time[b] (min)	Total Roasting Time (min)	Agtron[c]
Normal	210	02:40	11:25	74.4
Light	210	00:10	08:40	116.6
Scorched	275	01:50	07:40	66.0
Dark	220	04:45	13:45	45.7
Baked	230	06:20	18:00	68.3
Underdeveloped	135	02:30	20:20	74.9

[a] Air temperature when the beans entered the roaster.
[b] Time from 'first crack' to the end of the roast.
[c] Spectrophotometric measure indicating the color of the roast (smaller numbers indicate darker roasts).

physical changes – see Schenker, Handschin, Frey, Perren, and Escher (1999, 2000) for an overview. A visual representation of the variation in time-temperature profiles is given in Fig. 1, whereas detailed roasting conditions are reported in Table 1.

The main characteristics of the six roast profiles are the following:

- *Normal*. A reference coffee roast with time-temperature profile according to roasting guidelines of the Specialty Coffee Association of Europe (Münchow, 2016) with respect to initial temperature, developing time and total roasting time (Table 1). The coffee attained the highest 'Clean Cup' grade (10) by an experienced coffee roaster (author MM).
- *Light*. This roast defect has a temperature curve similar to the normal roast, but the roasting process was stopped about 4 min earlier, resulting in a shorter development time (Table 1). This prevented full aroma development from occurring. Accordingly, Yang et al. (2016) found a reduction in most volatile compounds for this sample compared to the Normal roast, with the exception of the heterocyclic compound indole (flowery, mothball-like), which was proposed as chemical marker for this defect.
- *Scorched*. The roasting process for this defect closely resembles that of the Normal roast profile, but it was quicker and at a higher temperature (Fig. 1). This high temperature-short time combination was found to cause a major change in aroma composition compared to the Normal roast. In particular, higher levels of the compounds 4-Ethyl-2-methoxyphenol, pyridine, phenol and difurfuryl ether (Yang et al., 2016). According to the known properties of these compounds, the coffee brewed from this roast could expectedly be described as smoky, burnt, roasted, bitter and astringent.
- *Baked*. The Baked roast had a temperature curve that start at a higher initial temperature in the bean compared to the Normal roast, and its roasting time lasted about 6 min longer (Table 1). The resulting aroma profile revealed a slight increase in most aroma compounds compared to the Normal roast, with the largest increase found for the compounds maltol (caramel-like), difurfuryl ether (roasted), and pyridine (roasted, burnt) (Yang et al., 2016).
- *Underdeveloped*. In this defect, the coffee was roasted at a much lower initial temperature (135 °C) and for 8 min longer than the Normal roast. In the authors'intention, the stalling of the temperature curve at the beginning of the roast should have prevented the development of many of the characterizing coffee aromas. This should have resulted in a flat, slightly sour coffee. Nevertheless, Yang et al. (2016) found that, despite the lower initial temperature, the relative abundance of most compounds was comparable to that of the Normal roast. It is thus expected that these two samples would be close from a perceptual point of view. The main difference with the Normal roast was the higher concentration of the compound 2,5-dimethylfuran (ether-like odor) (Yang et al., 2016).
- *Dark*. Finally, the Dark defect was roasted with a temperature curve similar to the normal roast, but for 2 min longer. As for the Scorched roast, this resulted in a general increase in aroma compounds compared to the Normal roast, most notably in the phenolic compounds 4-Ethyl-2-methoxyphenol and phenol (Yang et al., 2016). This would expectedly results in a coffee brew that could described as smoky or burnt.

All in all, the sample space obtained can be sees as reflecting a consensus representation among coffee professionals of common roasting defects, whereas the Normal reference would be regarded as "clean" (free of defects). Admittedly, the definition of the six roasting profiles took as point of departure current roasting practises in the European market (especially Northern Europe), and may not necessarily apply to other geographical regions where e.g. darker or lighter roasts may be more common.

2.2. Brewing

Sample preparation for the GC–MS analysis is described in Yang et al. (2016). This section describes brewing procedures using in relation to the sensory and consumer tests.

The packaged coffee beans were ground the day of serving using an electronic coffee grinder (KG 49, Delonghi, Austria), approximately three hours prior to tasting. The coffee was brewed using French press brewers (3 Cup Black Cafetiere, Argos, UK) by adding 50 g (± 0.5 g) of coarse ground coffee to 900 g (± 5 g) water. The hot water (approximately 95 °C) was poured over the grounds and the plunger was pressed down after 4 min and then decanted. 100 ml coffee was poured into each porcelain cup and the coffee settled in the cups in Thermaks cabinets at 22 °C to a temperature of 60 °C (± 1 °C) at which it was served. For the consumer test the coffee was held in thermos prior to serving for no more than 60 min before 100 ml was poured into each porcelain cup and settled to a temperature of 60 °C (± 1°) at which it was served. The thermos was labelled with sample number and the same flask was used only for that sample throughout the entire test period. From the literature various serving temperatures for coffee are suggested. There seems to be consensus of a serving temperature in the range of 80–85 °C among established coffee authorities and producers (Merrild, n.d.; National Coffee Association of America, n.d.), whereas several different consumer studies reveals that most consumers prefer a serving temperature between 60 and 70 °C (Borchgrevink, Susskind, & Tarras, 1999; Lee & O'Mahony, 2002). The temperature of 60 °C was chosen as it is low enough not to induce scalding hazards (Brown & Diller, 2008) and also represents the same temperature as the coffee would normally be consumed by the consumer.

2.3. Aroma composition analysis (GC–MS)

Analyses of volatile aroma compounds was conducted using a trace 1300 Gas Chromatograph-Mass Spectrometer (Thermo Fisher Scientific, Hemel Hemptead, UK). Volatiles were identified by comparing their mass spectrum with that of authentic compounds and/or with spectra in reference libraries. Concentrations was calculated with use from the internal standard and expressed in ppb. We refer the reader to Yang et al. (2016) for the detailed protocol used for the GC–MS analysis.

2.4. Sensory descriptive analysis

A descriptive analysis was carried out based on the principles of the Quantitative Descriptive Analysis. Nine assessors were recruited from the sensory panel at the University of Copenhagen. All assessors had been screened for sensory acuity and availability prior to inclusion in the sensory panel and were experienced in sensory evaluation of food prior to the study. The profiling took place in the sensory laboratory of University of Copenhagen standardized after ISO guidelines (ISO 8589:2007) and following good sensory practice (Lawless & Heymann, 2010).

The panel was instructed to evaluate the samples by the cupping method, where the coffee is aspirated into the mouth from a spoon (SCAA, 2009). Assessors were instructed to cleanse the mouth with plain white toast bread, milk and tepid water before the first and between each sample. All samples was served warm at a temperature of 60 °C (± 1 °C) in porcelain cups, blind labelled and with a three digit code. The profiling was carried out over four consecutive days (two days of training and two days of evaluation). The assessors initially generated their own attributes, and were later supplemented with a list of potential attributes and references to help the panel reach consensus on the meaning on the attributes. The final set of attributes and the associated references are reported in Table 2. The coffee samples were rated on a 15 cm unstructured line scale using the FIZZ software (Biosystemes). The coffee was evaluated in individual sensory booths using a randomized block design for the serving order, whereby each assessors evaluated each sample three times.

Table 2
Final set of attributes developed for the DA with corresponding scale anchors and reference material.

Modality	Attribute	Scale	Reference material
Overall	Intensity	A little → A lot	
	Complexity	A little → A lot	
Taste	Acidic	A little → A lot	0.50 g/L solution of citric acid
	Bitter	A little → A lot	Tepid strongly brewed dark roasted coffee
	Sweet	Nothing → A lot	7.3 g/L solution of sucrose
Flavor	Burnt	A little → A lot	Dark roasted toast bread
	Tobacco	Nothing → A lot	Roasted Red Orkil tobacco
	Licorice	Nothing → A lot	Karlsens licorice granulate
	Chocolate	Nothing → A lot	Amma 100% chocolate
	Dark Berries	Nothing → A lot	Elderberry juice, black currant juice and water (ratio 1:0.5:4)
	Roasted Ryebread	Nothing → A lot	Roasted 100% ryebread
	Nutty	Nothing → A lot	Roasted hazel nuts
	Caramel	Nothing → A lot	Dark syrup
	Citrus	Nothing → A lot	Thin slices of lemon and lime
Mouthfeel	Astringent	A little → A lot	
Aftertaste	Acidic	A little → A lot	
	Bitter	A little → A lot	
	Burnt	A little → A lot	

2.5. Consumer test

Eighty-three regular coffee consumers (40 males and 43 women, aged 18–70) participated in the test on a voluntary basis. Consumers were served the six samples monadically. The serving order was randomized across consumers following a balanced block design.

Unlike the trained panel, the consumers did not receive any specific instructions other than to drink the coffee as they would normally do. For each sample, they were first asked to rate the overall liking on a 9-point hedonic scale, and then to complete a check-all-that-apply (CATA) question. The latter consisted of 30 attributes, including both sensory and hedonic terms (the full list is visible in Table 6). The order in which the CATA attributes appeared on the ballots was randomized both *between* and *within* assessors to minimize possible order biases (Ares & Jaeger, 2013). At the end of the test, participants were asked to fill in a questionnaire with background information concerning their demographic and coffee habits (Table 3). The evaluations took place at the Department of Food Science, University of Copenhagen, in a well-lit air-conditioned room at a temperature around 22–24 °C. On average, consumers used approximately 30 min to complete the test.

Table 3
Information on the consumer sample who participated in the study ($N = 83$).

Background variable	N
Gender	
Males	40
Females	43
Age	
19–29	49
30–49	23
50 +	11
Coffee drinking frequency	
> 5 cups a day	7
3 to 5 cups a day	30
1 to 2 cups a day	28
1 to 6 cups a day	13
< 1 cup a week	5

2.6. Data analysis

All analyses were performed in R (R Core Team, 2014) using either native functions or functions from the packages FactoMineR (Lê, Josse, & Husson, 2008) and RVAide Memoire (Hervé, 2015). For analyses of inferential nature, the usual $\alpha = 0.05$ level for statistical significance was considered.

2.6.1. Sensory descriptive analysis

Differences in mean ratings between the samples in each of the sensory attributes were assessed by ANalysis Of VAriance (ANOVA) using a mixed model with sample and replicate as fixed effects, and assessors as random. When significant fixed effects were found, the ANOVA was followed by post hoc comparison by Tukey's Honestly Significant Difference test. To enable a visual exploration of the sensory results, Principal Component Analysis (PCA) was performed on the significant sensory attributes using data averaged across both replications and assessors. The data were mean-centered and scaled columnwise (i.e., values were multiplied by the inverse of the standard deviation for that attribute) prior to the computation of the PCA model.

2.6.2. Relationships between sensory and instrumental aroma measurements

In order to explore relationships between the aroma composition and the sensory data, a Multiple Factor Analysis (MFA) (Husson et al., 2005) was conducted using two inputs matrices: one containing aroma compounds concentrations, and one containing sensory attributes. Both datasets contained data averaged across samples and only included compounds and sensory attributes that significantly discriminated between the samples assessed by ANOVA[2].

2.6.3. Consumer data: liking and CATA evaluations

A mixed model ANOVA was performed to uncover differences in mean liking ratings between the samples. The model included sample as fixed effect and consumer as random, and was followed by pairwise comparisons by Tukey Honestly Significant Difference (HSD). The CATA responses where rendered as a dichotomous data where a value of 1 indicated that an attribute had been checked and a value of 0 indicated the opposite. Differences between samples with respect to frequency of mention on each individual CATA attribute were assessed using Cochran's Q Test, as customary for this type of data (Meyners, Castura, & Carr, 2013). To visualize the frequency of associations of samples with the CATA attributes a correspondence analysis was performed on the contingency table.

3. Results

3.1. Sensory descriptive analysis

Table 4 shows the results of the ANOVA analyses on the DA data. All but two attributes (*nutty* and *caramel*) were found to significantly discriminate between the samples.

The PCA scores and loadings plot for the model using averaged DA data are shown in Figs. 2 and 3, respectively. The first two model dimension accounted for high proportion of the variance (over 95%) in the sensory profiles, indicating a clear variance structure in the data.

The first PCA dimension mainly differentiated between the Light on one end, and the Dark and Scorched Roast on the other (Fig. 2). The Dark and Scorched sample were associated to the attributes *Intensity*, *Bitter*, *Bitter (Aftertaste)*, *Astringent*, *Burnt*, *Burnt (Aftertaste)*, *Licorice*, and *Tobacco* - many of these attributes can be linked to a higher degree of roast, lending face validity to this opposition. Conversely, the Light sample was rated significantly lower in these attributes (Table 4), and was instead primarily associated with the attributes *Citrus*, *Sweet*, and *Roasted Ryebread* (Figs. 2 and 3). The association of these sensory attributes with the Light sample would suggest that these flavor notes may already be present in the bean, or formed very early in the roasting process. The Light sample was the most singled out in the first dimension Fig. 2. It generally obtained lower mean rating than the other samples in all remaining attributes, which would suggest that most of the sensory variation is due to the roasting process. The remaining three samples (Normal, Baked, and Underdeveloped) were not well described by the first dimension. Their position close to center of the plot, as well as inspection of Table 4, indicates that these samples generally received ratings close to the grand mean of the attributes described by the first dimension.

These three samples were better discriminated by the second model dimension, which mostly described variation in the attributes *Complexity*, *Acidic*, *Acidic (Aftertaste)*, and *Dark Berries*. The Normal roast had the largest positive score on this dimension and, accordingly, received the highest mean ratings in these attributes. However, pairwise comparisons (Table 4) indicated that the difference with the other two samples positively loaded on this dimension (Baked and Underdeveloped) was not statistically significant. The second dimension also highlighted differences between the Scorched and the Dark sample (Fig. 2), which had nearly identical position on the first dimension. Their distance on the second dimension was due to slight differences on the attributes *Acidity*, *Acidity (Aftertaste)*, *Chocolate*, and *Dark Berries*, where the Dark roast had lower mean ratings. The differences in acidity could be attributed to additional acid degradation associated with the longer roasting time for the Dark sample. Generally speaking, the DA results corresponded with the definitions of the roasting profiles from Section 2.1.

3.2. Sensory-instrumental relationships

The GC-MS results have previously been reported in Yang et al. (2016), to which we refer the reader for in-depth analysis on the aroma profiles of individual roast defects. In this section, we will focus on exploring instrumental-sensory correlations modeled by MFA.

The main outputs of this analysis are shown in Figs. 4 and 5, whereas a full numerical account of the contribution of each variable to the MFA model is given in the appendix to this paper. As in the previous PCA model, two dimensions accounted for over 85% of the original variance. The product space obtained is shown in Fig. 4 which also included partial points obtained by considering the two input matrices separately. The plot shows that the aroma and sensory data produced nearly identical product spaces. The only noteworthy difference concerned the distance between the samples Dark and Scorched, which the panel perceived as very close perceptually, whereas in the aroma data they are quite strongly differentiated on the first dimension (Fig. 4, see also Fig. 2).

The first MFA dimension again related to the opposition of the Light vs. the Scorched and Dark roasts. The MFA loadings plot (Fig. 5) indicates that this was due to a general increase in aroma compound concentrations associated with the Scorched roast, which was according to expectations (see Section 2.1). The vast majority of the aroma compounds appear bundled in a tight cluster – including mostly pyrazines, aldehydes, alcohols, sulphides, pyrrols, and furans – positively correlated with the first MFA dimension. As we have seen, from a sensory point of view these resulted in an increase in the intensity of several attributes related to the higher degree of roast. The sensory attributes *Burnt*, *Astringent* and *Burnt (Aftertaste)* also correlated highly with the first dimensions, which could be due to high concentrations of pyridine and furfuryl alcohol.

The second MFA dimensions separated the Dark roast, and to a lesser extent the Normal roast, from the Light and the Scorched roasts. This direction was mainly associated with variation in the concentration of organic acids (acetic acid, butanoic acid, hexanoic acid) and, correspondingly, variation in acidity (Fig. 5).

In spite of a general agreement regarding relative sample differences, some differences between the datasets are observable; for

[2] ANOVA results for the GC-MS data are shown in Yang et al. (2016).

Table 4
Mean ratings (15 cm unstructred scale) for each sensory attributes for each of the six roasting profiles. The last two columns show the F value for the sample effect from the corresponding ANOVA model and the associated p value ($n.s.$ = not significant; $^*p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$). Within rows, means not sharing superscript letters are significantly different ($p < 0.05$), following pairwise comparison by Tukey HSD test. Attributes are ranked by decreasing size of the F statistic, i.e. by most to least discriminating attribute.

Attribute	Normal	Baked	Dark	Light	Scorched	Underdev.	$F_{(5,150)}$	p
Intensity	9.2^b	9.3^b	11.5^a	3.8^c	11.9^a	8.2^b	35	***
Burnt	9.1^b	8.9^b	12^a	3.9^c	11.8^a	8^b	32.6	***
Bitter	9.6^b	10.3^b	12.2^a	4.8^c	12.4^a	9.1^b	31.3	***
Burnt (Aftertaste)	7.8^b	8.4^b	12.2^a	3^c	11.2^a	6.7^b	31.1	***
Tobacco	8.2^b	7.4^{bc}	10^a	4^d	9.9^a	6.9^c	20.9	***
Bitter (Aftertaste)	8.8^b	9^b	11.3^a	4.1^c	11.1^a	7.7^b	18.3	***
Citrus	5.7^b	4.4^{bc}	1.9^c	8.8^a	3^c	5.2^{bc}	16.6	***
Licorice	5.4^b	5.9^{ab}	6.7^a	2.3^c	6.6^a	5^b	10.6	***
Astringent	9.2^{ab}	9.1^{ab}	10.9^a	5.7^c	10.6^a	8.3^b	8.4	***
Sweet	3.8^b	3^{bc}	2.3^c	5^a	2.2^c	3.7^b	5.2	***
Chocolate	6.2^{ab}	6.9^{ab}	6^{ab}	3.4^c	7.1^a	5.4^b	4.9	***
Acidic (Aftertaste)	10.3^a	7.9^b	5.7^c	8.6^{ab}	8.2^b	9.3^{ab}	4.7	***
Complexity	8.8^a	8.4^{ab}	6.5^b	5.2^c	7.3^{ab}	7.6^{ab}	4.6	***
Roasted Ryebread	6.9^{abc}	7.2^{ab}	5^c	8.6^a	5.9^{bc}	6.7^{abc}	3.5	**
Acidic	9.3^a	7.6^{ab}	6.5^b	8.7^a	7.9^{ab}	9.3^a	2.9	*
Dark Berries	5.7^a	6.2^a	3.7^b	5.3^{ab}	5^{ab}	5.9^a	2.6	*
Nutty	4	5.1	4	5.9	4.1	4.3	1.8	n.s.
Caramel	3.3	3.6	2.9	4.2	3	2.7	1.2	n.s.

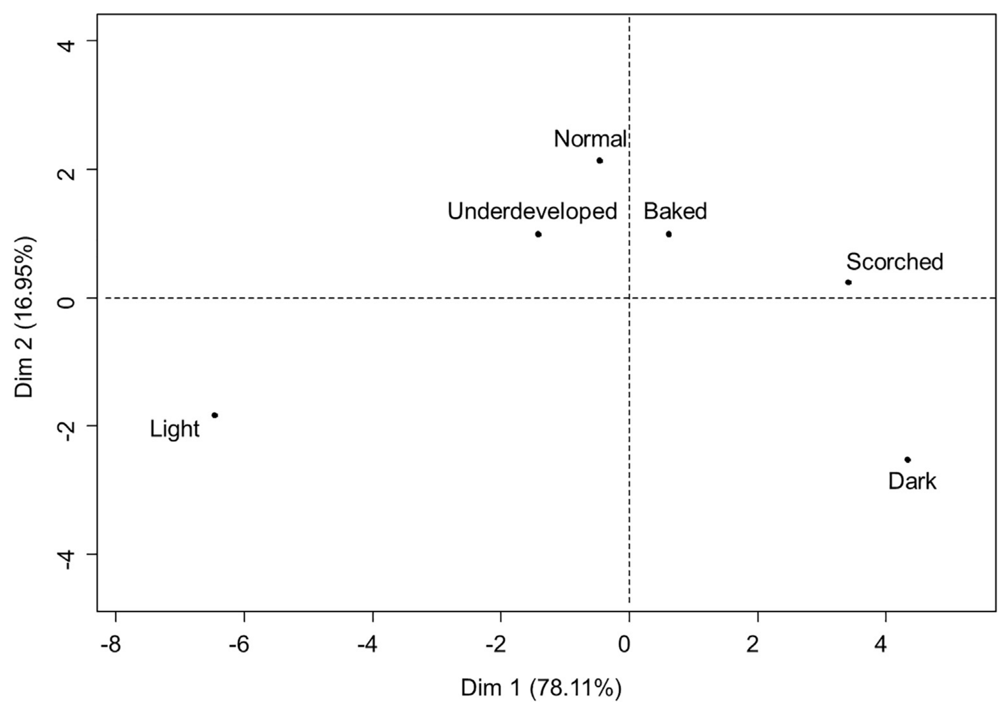

Fig. 2. Scores plot showing relative sensory differences between the samples on the first two PCA dimensions.

instance, we can see that aroma data do not seem to explain the sensory attributes on the left side of Fig. 5 and that there are no sensory attributes correlated to the compounds at the bottom of the same plot. This inconsistency could be due to different factors. For some of the volatiles on the bottom of the plot it may be due to their presence at subthreshold level and/or to limitations in the attribute list that did not include specific odors commonly associated with these volatiles. This is quite possibly the case for furfural (almond-like) and 2,3-pentanedione (buttery). For sensory attributes located on the left side of the plot, the fact that there are no associated volatiles associated might be due to suppression effects. Recall that the left end of the plot is defined by the sample ”Light” and mostly reflects the fact that this sample has the lowest concentration of nearly all aroma compounds. Lower concentrations of aroma compounds may have made some sensory attributes (sweetness and acidity in particular) more prominent in the Light sample, regardless of absolute values. For example, the sample Underdeveloped had the highest concentration of Acedic acid (41.64 ppm), much higher than both Light (20.72 ppm) and Normal (31.04 ppm), yet looking at Table 4 reveals that it was not different from those samples in terms of perceived acidity.

3.3. Consumer perception of the coffee samples

Mean hedonic ratings for the six roasts are reported in Table 5. ANOVA results revealed a significant main effect of sample on liking ($F_{(5,492)} = 7.7, p < .001$). As expected, the Normal roast obtained the highest liking ratings, whereas the Light roast was the least liked (Table 5). The range of the consumer liking ratings was not very large (Min: 4.2, Max: 6). However, it is worth noting that there was a statistically significant difference between the Normal roast and all other samples, except for Baked.

Table 6 reports the frequency of occurrence of each CATA attributes

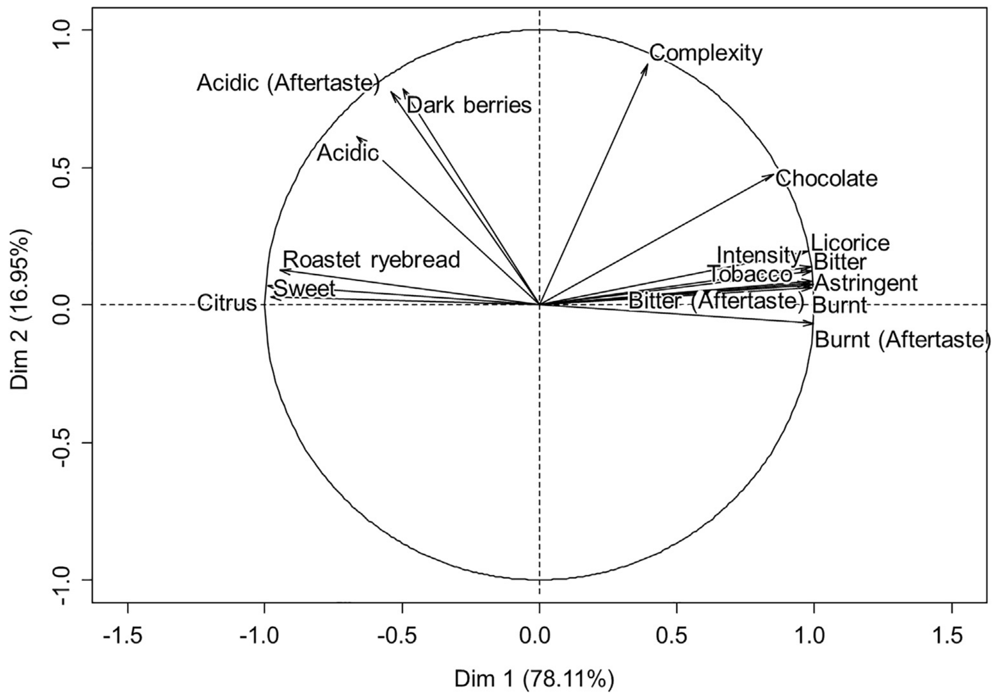

Fig. 3. Correlation of sensory attributes with the first two PCA dimensions.

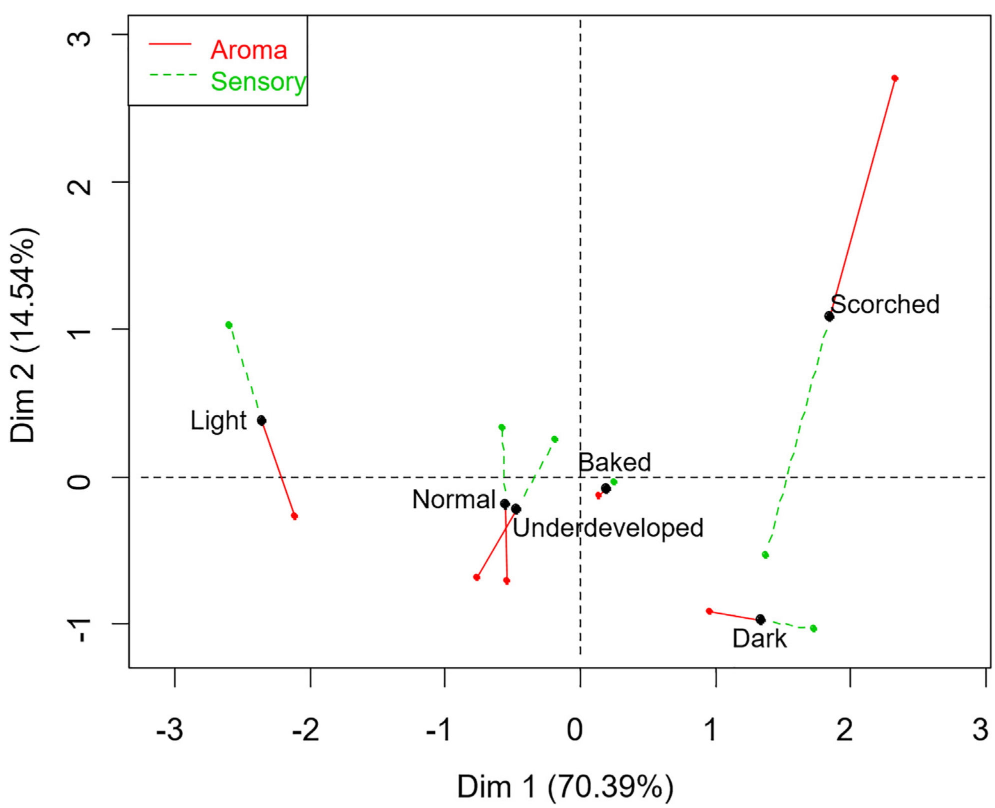

Fig. 4. Scores plot showing relative differences between the samples on the first two MFA dimensions. The model is based on both sensory and aroma data and also shows partial points obtained from the two datasets separately.

across the six samples. All terms were used at least once for each sample. Even the attribute with the lowest occurrence (*Grass*) was used 38 times, indicating that all the attributes were relevant to the consumers. Significant differences between the samples were found for 20 out the 30 CATA attributes. The three most discriminating CATA attributes were *Thin*, *Strong*, and *Mild*.

The associations between samples and CATA attributes are visually summarized in Fig. 6, which shows the bi-plot of the CA performed on the CATA contingency matrix (two dimensions retained, 93.55% of explained variance). Comparing this plot with Fig. 2, it is easy to see that the consumers generated a sensory space almost identical to that of

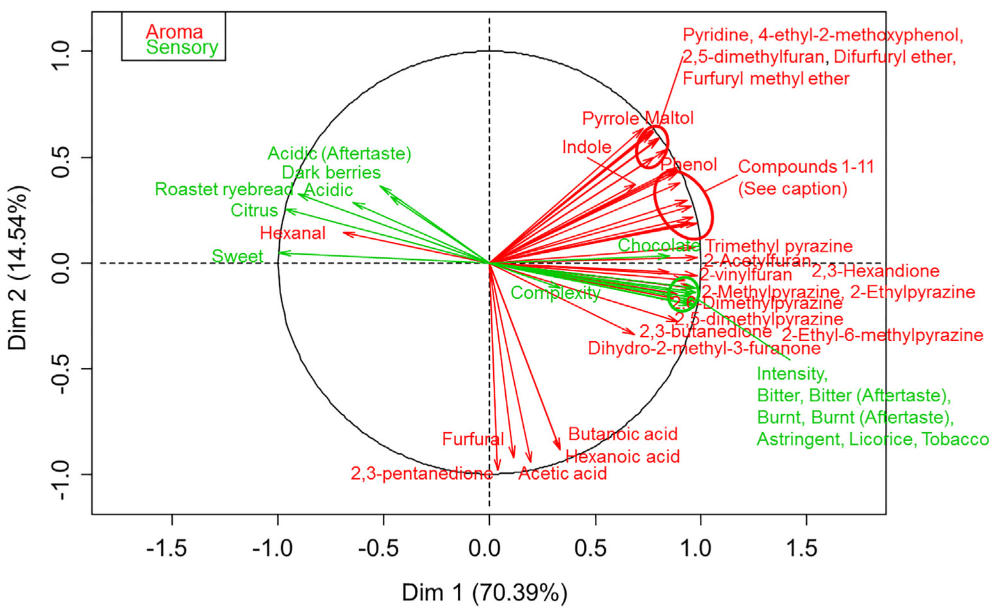

Fig. 5. Correlations of the sensory attributes and aroma compounds with the first two MFA dimensions. The unlabeled compounds are the following (ordered by size of correlation with Dim 1): (1) 1-Furfurylpyrrole; (2) Furfuryl alcohol; (3) 2-Methylbutanal; (4) 2,3-dimethyl-Pyrazine; (5) Dimethyl Trisulfide; (6) 3-Methylbutanal; (7) Octanoic acid; (8) 2-Furfuryl methyl disulfide; (9) 1-(1-H-pyrrol-2-yl) ethanone; (10) 3-Methylthiophene; (11) Dimethyl disulfide.

Table 5
Mean liking (9-pt hedonic scale) and standard deviation for the six samples ($N = 83$). Means not sharing superscript letters are significantly different (Tukey $p < 0.05$).

	Mean	St. Dev.
Normal	6.0^{a}	1.8
Baked	5.7^{ab}	1.9
Scorched	5.6^{bc}	2.2
Underdeveloped	5.4^{bc}	2.0
Dark	5.1^{c}	2.2
Light	4.2^{d}	2.1

the trained assessors[3].

Again, the first model dimension describes variation between the Light and the Dark samples. The Dark sample was again associated with attributes related to the darker degree of roast (e.g., *Tobacco*, *Burnt*, *Sharp*, *Long aftertaste*, *Bitter*). This sample was also perceived as the most intense (highest in attribute *Strong* and *Intense*, see Table 6) to such an extent that it is also described as *Unpleasant* by the consumers. The Scorched sample lies close to the Dark sample in the first CA dimension. As in the panel data, these two samples are associated with the same attributes, although they are better differentiated here due to the fact that the Scorched sample was generally perceived as less intense than in the Dark (Table 6). With respect to the differences between these two samples, the product space obtained from the CATA data are therefore in line with the indications of the aroma data. The Light sample lies in the opposite direction in the first dimension (Fig. 6), and appears as the most different from all others. Like the trained panelists, consumers perceived it as the sweetest and less intense tasting of all samples and, additionally, associated this samples with the attributes *Thin* and *Hey/straw* (Fig. 6 and Table 6).

Although accounting for only 9.5% of the data variance, the second CA dimension provided useful information on the differences of the Normal roast (Fig. 6). This sample was primarily associated with the attributes *Sweet* and *Caramel*, and with two holistic attributes with a positive valence, *Pleasant* and *Balanced.* The latter associations are interesting as they related to absence of defects, and confirm the indications of the hedonic ratings (Table 5). The Normal roast was also the most frequently associated with the attribute *Harmonic*, though in this case the differences were too close to reach statistical significance (Table 6).

The Underdeveloped and Baked roast were again poorly described by the model and showed sensory profiles quite similar to the Normal roast, especially in the first dimension (Fig. 6). However, Table 6 shows some significant differences between these two samples and the Normal roast. The Baked sample was perceived as significantly more *Bitter* and less *Sweet* than the Normal, whereas the Underdeveloped roast was perceived as significantly lower in the attributes *Caramel* and *Dark Berries*.

4. Discussion

All three datasets (aroma, sensory and consumer) provided consistent indications concerning the main direction of differences between the six samples. As previously reported in Yang et al. (2016), the results indicated a significant increase in aroma compound concentration – particularly pyrazines, aldehydes, alcohols, sulphides, pyrrols, and furans – associated with prolonged roasting time and temperature. This is well in line with literature accounts regarding the influence of roasting to aroma formation in coffee (Masi et al., 2013). The highest aroma concentrations were found in the samples Dark and, especially, Scorched (Table 1). This was clearly reflected in the corresponding sensory profiles for these samples which were highest in overall sensory intensity, and scored highest in attributes typically associated with the roasting process.

The Normal roast generally obtained values close to average with respect to sensory attribute intensity and aroma compounds concentration. The aroma compounds most strongly associated with the Normal roast were organic acids, which resulted in a higher perceived acidity in this sample compared to all others. This ultimately separated this sample from the high intensity roasts (Scorched in particular), where these acids are lost, and where Maillard compounds and lipid breakdown products abound (Yang et al., 2016). The Normal roast was also well differentiated from the Light defect, which was perceived as the sweetest and least intense of all samples, due to the fact that, as per our intentions with this sample, shortening the development time did not allow full aroma development to occur.

The Normal roast was instead not well differentiated from the defects Baked and Underdeveloped, which was expected as their aroma composition (particularly in the case of Underdeveloped) was relatively close to that of the Normal roast (Yang et al., 2016). The results are thus inconclusive with respect to the differences between these two samples and the Normal roast, although it is worth noting that the consumers perceived the

[3] The ranking of the samples on the two dimension is reversed compared to Fig. 2. However, this is accidental and irrelevant to the interpretation as the focus of both models is on *relative* differences between the samples.

Table 6
Contingency table showing the frequency of mention of each CATA attribute for each of the six roasting profiles. The last two columns report the test statistic for Cochran's Q test (Q) and the associated p value ($n.\ s.$ = not significant; $*p < 0.05$; $**p < 0.01$; $***p < 0.001$). Within rows, frequencies not sharing superscript letters are significantly different ($p < 0.05$), following pairwise comparison by Cochran's Q test. CATA attributes are ranked by decreasing size of the Q statistic, i.e. by most to least discriminating attribute.

	Normal	Baked	Dark	Light	Scorched	Underdev.	Col. Total	Q	p
Thin	17^{c}	14^{c}	2^{d}	58^{a}	10^{c}	27^{b}	128	123	***
Strong	18^{c}	26^{c}	55^{a}	1^{d}	41^{b}	24^{c}	165	97.9	***
Mild	28^{b}	23^{bc}	3^{d}	57^{a}	15^{c}	20^{bc}	146	93	***
Tobacco	12^{de}	26^{b}	43^{a}	7^{e}	22^{bc}	16^{cd}	126	65.8	***
Burnt	29^{b}	36^{b}	56^{a}	8^{c}	40^{b}	32^{b}	201	63.8	***
Long aftertaste	26^{c}	38^{b}	53^{a}	6^{d}	37^{b}	31^{bc}	191	62.5	***
Bland	8^{bc}	7^{bc}	3^{c}	32^{a}	6^{c}	16^{b}	72	54.8	***
Intense	17^{c}	23^{bc}	36^{a}	2^{d}	30^{ab}	15^{c}	123	52.9	***
Bitter	36^{b}	4^{d}	50^{a}	13^{c}	46^{ab}	39^{b}	225	47.6	***
Sharp	12^{c}	13^{c}	32^{a}	2^{d}	23^{ab}	14^{bc}	96	42.3	***
Rich	21^{b}	21^{b}	33^{a}	4^{c}	25^{ab}	20^{b}	124	31.1	***
Sweet	14^{a}	4^{bc}	1^{c}	17^{a}	9^{ab}	11^{a}	56	24.2	***
Hey/straw	8^{b}	11^{b}	15^{ab}	24^{a}	8^{b}	10^{b}	76	20.2	**
Balanced	29^{a}	16^{bc}	9^{c}	17^{bc}	22^{ab}	20^{ab}	113	16.1	**
Complex	9^{c}	18^{ab}	14^{abc}	9^{bc}	21^{a}	7^{c}	78	15.7	**
Unpleasant	7^{b}	7^{b}	20^{a}	11^{ab}	13^{ab}	6^{b}	64	15.5	**
Astringent	14^{a}	12^{a}	19^{a}	5^{b}	15^{a}	12^{a}	77	13	*
Caramel	11^{ab}	9^{ab}	5^{b}	11^{ab}	12^{a}	2^{c}	50	12.1	*
Pleasant	33^{a}	30^{a}	17^{b}	19^{b}	25^{ab}	23^{ab}	147	11.3	*
Dark berries	12^{ab}	7^{b}	5^{b}	10^{ab}	16^{a}	8^{b}	58	10.8	*
Licorice	7	7	9	2	5	10	40	9.4	n.s.
Earthy	14	21	21	12	16	11	95	8.7	n.s.
Harmonic	23	18	13	11	19	16	100	7.4	n.s.
Acidic	32	33	26	30	39	37	197	7.3	n.s
Chocolate	26	38	53	6	37	31	107	6.9	n.s.
Grass	14	4	1	17	9	11	38	5.5	n.s.
Nutty	21	20	16	19	25	15	116	5.3	n.s.
Citrus	14	11	10	17	15	11	78	4.3	n.s.
Delicate	14	10	7	12	13	9	65	4.1	n.s.
Roasted ryeb.	16	22	21	18	19	21	117	2.1	n.s.
Row Total	529	548	613	458	613	508			

Normal roast as significantly sweeter and less bitter than these two defects.

An interesting finding was that the trained panelists rated the Normal roast highest in the holistic attribute *Complexity*. In the sensory literature, flavor complexity has been defined as the total number of separate recognizable sensory qualities in a stimulus (Giacalone, Duerlund, Begh-Petersen, Bredie, & Frst, 2014), and this definition was also used in the training of the panel for this study. Looking at the PCA for the sensory panel data (Fig. 3), it would appear as though

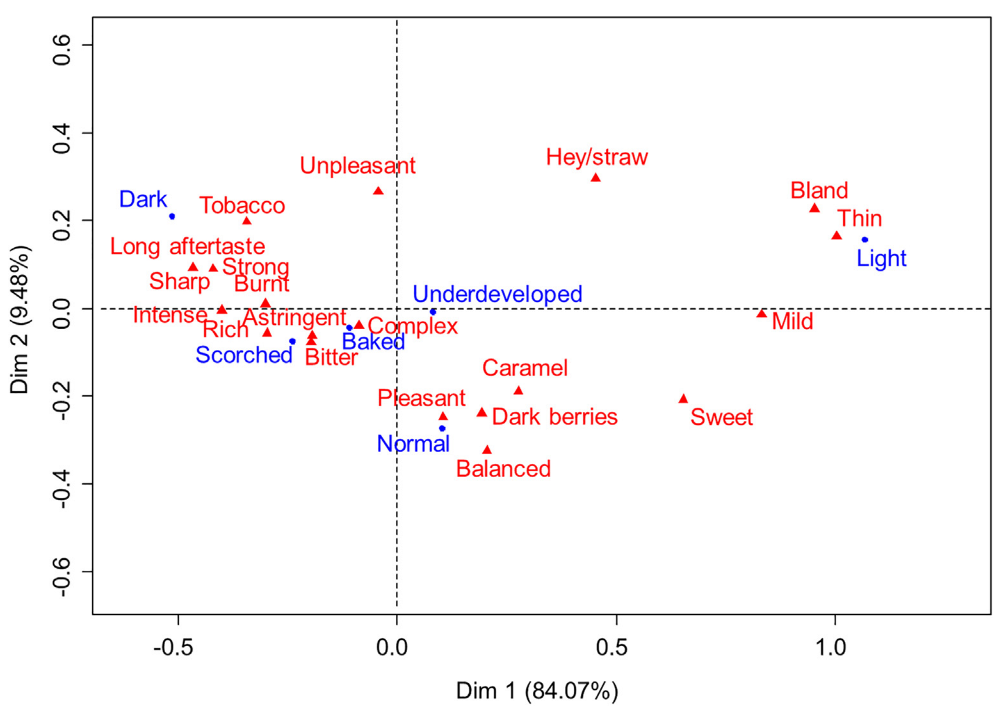

Fig. 6. Bi-plot showing associations of samples and CATA attributes on the first two CA dimensions.

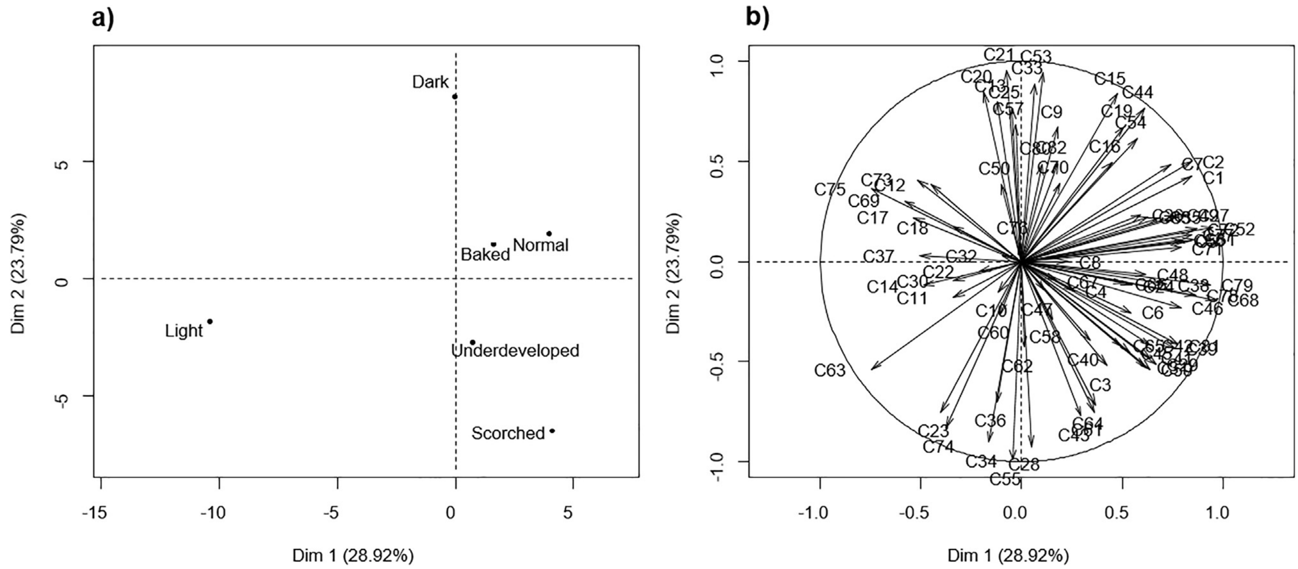

Fig. 7. Internal preference map showing the position of the samples based on their hedonic scores (a) and the direction of individual preference for each individual consumer (b).

complexity stands in an inverse U-shaped relationship with overall sensory intensity. A quadratic model with intensity as predictor and complexity as response confirmed this intuition as the model obtained[4] revealed a significant downward slope associated with the quadratic term ($F(2,3) = 7.86, p = 0.06; R^2 = 0.83$). The underlying phenomenon here seems to be that for very low level of intensity complexity is low as there a few recognizable sensory qualities in the stimulus. In our dataset this is the case of the Light sample where the short roasting time prevented the formation of many aroma compounds. For high intensity levels complexity is also low as the present of strong sensory inputs may dominate the percept, such as the sample Dark. The Normal roast, characterized by moderate intensity sensory, can be understood as having an optimal complexity level where many flavors are recognizable but no one flavor is dominating or off-putting. The association between the Normal roast and the CATA attribute *Balanced* observed in the consumer data also supports this interpretation. Overall, the results seem consistent with the expectation that the Normal roast (characterized by absence of defects and a high 'clean cup' score) would correspond to a coffee brew with a fully developed aroma profile but lacking dominating off-flavors.

Importantly, the consumer test results showed that the Normal roast was the most liked, significantly preferred over all other samples, except Baked (although this lack of difference may be due to insufficient statistical power). Even though the observed differences in average liking ratings were not very large, these results do suggest that absence of defects is relevant to consumer liking of coffee. The main implication for the coffee industry is that roasters may be able to pinpoint at specific markers (chemical and sensory) that may be used to set up internal quality control scoring systems (for aspects pertaining to coffee roasting) in both quality control and product development. For instance, in a product development context roasters would first identify an optimal roast degree, based on their own subjective and/or on a consumer test. Then, 'clean cup'-*like* evaluations can be used internally to further optimize the roasting profiles, e.g. with respect to timing aspects. Because we expect practical applications in the coffee branch to involve a smaller sensory range than the one used in this paper (particularly with respect to visual variation), we strongly recommend that roasters validate their internal evaluations against consumer test results obtained from their target population of interest.

Taken overall, the results of this study provide a comprehensive characterization of chemical and perceptual markers associated with common coffee defects, and demonstrate that a 'clean cup' (a coffee without defects) is associated with higher consumer preference. From a sensory scientific point of view, this research indicates that the attribute 'clean cup' describes a coffee of average sensory intensity, high in acidity, and having with many recognizable flavor attributes.

4.1. Limitations and future research

We acknowledge several limitations in this research that should be kept in mind in order to correctly qualify the results. First, the study only used a single origin *Arabica* bean, and thus the conclusions may not readily generalize to other coffee varieties and origin. For example, the specific finding that the Normal roast was high in acidity is almost certainly related to the choice of coffee (Kenyan coffee is supposed to be high in acidity) whereas it may be considered a defect in different varieties (e.g., Sumatran coffee).

Furthermore, the research only considered defects germane to the roasting process, thus not including other important sources of off-flavours in coffee – e.g., problems related to production, processing, and storage of the green beans (Agresti et al., 2008; Wintgens, 2009). The impact of these defects on the sensory quality of coffee could be a relevant venue for future research.

One additional aspect that deserve attention in future research is the heterogeneity in consumer preferences for coffee. Previous studies have shown that coffee is a product for which different consumer segments (in terms of preferences for specific sensory attributes) can be identified (Varela, Beltrán, & Fiszman, 2014). The data collected here also suggest that this is the case. Interestingly, we note that consumer liking ratings (not shown here) were approximately normally distributed for the two best liked sample (Normal and Baked), but rather bi-modal for the two worst liked sample (Dark and Light), with a significant proportion of the consumers giving high liking ratings for these two roasting profiles. Accordingly, Fig. 7 shows scores and loadings from an internal preference map obtained by performing a PCA on a matrix containing the hedonic scores for the six samples, from which it can be seen that, although the majority of the consumers' preference vectors are located in the direction of the Normal roast, several consumers are also located in other areas of the plots, including a sizeable minority expressing high preference for the Light and Dark roasts. We refrained from discussing this aspect in the paper because our sample size is insufficient to attempt a robust segmentation. However, it is clear that understanding this heterogeneity in relation common coffee defects may be a useful direction for future research. It would especially be interesting to link different preferences to consumers' background. Previous research have

[4] $Complexity = -1.74 + 2.29*Intensity - 0.13*Intensity^2$

pointed at several factors that may contribute to defining coffee preference segments, including gender (Cristovam, Russell, Paterson, & Reid, 2000), product usage (Masi et al., 2013) and, more recently, physiological differences in terms of taste sensitivity (Hayes et al., 2010; Masi, Dinnella, Monteleone, & Prescott, 2015) and coffee metabolism rate (Masi, Dinnella, Pirastu, Prescott, & Monteleone, 2016).

5. Conclusion

This work has investigated common roasting defects in coffee considering compositional (GC–MS), perceptual (sensory descriptive analysis with a trained panel and consumer-based CATA) and affective (consumer liking) aspects. The sensory and GC–MS analyses revealed identical information regarding the overall inter-sample differences, and pointed at a large influence of the roasting process in the aroma and sensory profiles of the roasts.

The results indicated a significant increase in aroma compound concentration associated with prolonged roasting time and temperature, resulting in an increase in sensory attributes typically associated with the roasting process - such as *Bitter*, *Astringent*, *Burnt*, *Licorice*, and *Tobacco* - as well as to an overall increase in flavor intensity. The Normal roast generally obtained values close to average with respect to sensory attribute intensity and aroma compounds concentration, consistent with the idea that a coffee without defects corresponds to a brew with a fully developed aroma profile is related but lacking dominating off-flavors. Supporting this interpretations, consumers described the Normal sample as the most *Balanced*. Most importantly, the Normal coffee obtained the highest consumer liking ratings.

Taken overall, these results provide a solid basis for understanding chemical and sensory markers associated with common roasting defects, which coffee professionals may use to set up internal protocols in the context of quality control and product development applications.

Acknowledgements

We thank Kontra Coffee (Dag Hammarskjolds Alle 36, 2100 Copenhagen, Denmark) for supplying the coffee used in the project and for the use of their roasting equipment. We also thankfully acknowledge the help of Belinda Nielsen, Julie Geertsen and Sarah Østergaard Brandt from University of Copenhagen with preparing the coffee samples for the sensory evaluation.

Appendix A. Contributions of aroma and sensory attributes to the MFA dimensions

Table A.1
Correlation coefficients size and significance between aroma compounds and sensory attributes with the first and second dimension of the MFA model.

Dimension 1	*r*	*p*	Dimension 2	*r*	*p*
1-Furfurylpyrrole	0.99	0.0001	Pyrrole	0.59	0.2150
2-Acetylfuran	0.98	0.0006	Pyridine	0.56	0.2433
Furfuryl alcohol	0.98	0.0006	4-Ethyl-2-methoxyphenol	0.56	0.2462
2-Methylbutanal	0.97	0.0009	2,5-Dimethylfuran	0.54	0.2664
2,3-Hexandione	0.97	0.0011	Phenol	0.53	0.2835
Intensity	0.97	0.0015	Difurfuryl ether	0.53	0.2837
Bitter	0.97	0.0016	Maltol	0.46	0.3589
Burnt	0.96	0.0021	Furfuryl methyl ether	0.46	0.3622
Dimethyl Trisulfide	0.96	0.0022	Indole	0.39	0.4404
Bitter Aftertaste	0.96	0.0024	3-Methylthiophene	0.38	0.4522
Trimethylpyrazine	0.96	0.0027	1-(1-H-pyrrol-2-yl) ethanone	0.37	0.4639
Burnt Aftertast.	0.96	0.0029	Roasted ryebread	0.36	0.4770
3-Methylbutanal	0.95	0.0035	2-Furfuryl methyl disulfide	0.36	0.4779
Tobacco	0.95	0.0035	Dimethyl disulfide	0.35	0.4966
2,3-Dimethylpyrazine	0.95	0.0036	Citrus	0.31	0.5443
Astringent	0.95	0.0038	Acidic aftertaste	0.30	0.5570
2-Vinylfuran	0.94	0.0050	Acidic	0.29	0.5829
2-Ethylpyrazine	0.94	0.0052	Octanoic acid	0.28	0.5954
Licorice	0.93	0.0068	Dark berries	0.24	0.6401
2-Furfuryl methyl disulfide	0.93	0.0080	3-Methylbutanal	0.23	0.6637
1-(1-H-pyrrol-2-yl) ethanone	0.93	0.0081	2-Methylbutanal	0.21	0.6947
Dimethyl disulfide	0.92	0.0087	Hexanal	0.20	0.6978
3-Methylthiophene	0.92	0.0094	Furfuryl alcohol	0.16	0.7674
Octanoic acid	0.92	0.0102	Dimethyl trisulfide	0.15	0.7778
2,6-Dimethylpyrazine	0.89	0.0180	Sweet	0.13	0.8107
Maltol	0.87	0.0235	2,3-dimethyl Pyrazine	0.12	0.8217
2,3-Butanedione	0.87	0.0259	1-Furfurylpyrrole	0.12	0.8272
Chocolate	0.85	0.0317	Trimethyl pyrazine	−0.01	0.9839
Difurfuryl ether	0.85	0.0321	2-Acetylfuran	−0.07	0.8988
2-Methylpyrazine	0.85	0.0328	Chocolate	−0.12	0.8176
Phenol	0.85	0.0331	2,3-Hexandione	−0.13	0.7989
Furfuryl methyl ether	0.83	0.0401	2,6-Dimethylpyrazine	−0.14	0.7878
2,5-Dimethylpyrazine	0.83	0.0416	2-Ethyl-6-methylpyrazine	−0.16	0.7657

(*continued on next page*)

Table A.1 (*continued*)

Dimension 1	*r*	*p*	Dimension 2	*r*	*p*
2-Ethyl-6-methylpyrazine	0.83	0.0419	2-Ethylpyrazine	−0.18	0.7334
4-Ethyl-2-methoxyphenol	0.83	0.0431	2-vinylfuran	−0.18	0.7320
Pyridine	0.82	0.0440	2-Methylpyrazine	−0.19	0.7126
2,5-Dimethylfuran	0.81	0.0490	Intensity	−0.22	0.6740
Pyrrole	0.78	0.0648	2,5-Dimethylpyrazine	−0.23	0.6674
Indole	0.71	0.1120	Bitter	−0.24	0.6522
Dihydro-2-methyl-3-furanone	0.67	0.1490	Tobacco	−0.25	0.6395
Complexity	0.34	0.5102	Burnt	−0.26	0.6214
Butanoic acid	0.27	0.5979	Bitter aftertaste	−0.27	0.6002
Hexanoic acid	0.27	0.6041	Burnt aftertaste	−0.28	0.5967
2,3-Pentadione	0.13	0.8116	Complexity	−0.29	0.5833
Acetic acid	0.04	0.9392	2,3-Butanedione	−0.29	0.5716
2-Furfural	−0.04	0.9465	Astringent	−0.30	0.5677
Dark berries	−0.44	0.3874	Licorice	−0.30	0.5572
Acidic Aftertaste	−0.45	0.3728	Dihydro-2-methyl-3-furanone	−0.42	0.4063
Acidic	−0.58	0.2257	Butanoic acid	−0.89	0.0178
Hexanal	−0.67	0.1465	Acetic acid	−0.89	0.0174
Roastet ryebread	−0.88	0.0198	Hexanoic acid	−0.89	0.0168
Citrus	−0.94	0.0058	2,3-Pentadione	−0.94	0.0056
Sweet	−0.98	0.0007	2-Furfural	−0.96	0.0021

References

Agresti, P. D. M., Franca, A. S., Oliveira, L. S., & Augusti, R. (2008). Discrimination between defective and non-defective brazilian coffee beans by their volatile profile. *Food Chemistry, 106*, 787–796.

Ares, G., & Jaeger, S. R. (2013). Check-all-that-apply questions: Influence of attribute order on sensory product characterization. *Food Quality and Preference, 28*, 141–153.

Baggenstoss, J., Poisson, L., Kaegi, R., Perren, R., & Escher, F. (2008). Coffee roasting and aroma formation: Application of different time- temperature conditions. *Journal of Agricultural and Food Chemistry, 56*, 5836–5846.

Bhumiratana, N., Adhikari, K., & Chambers, E. (2011). Evolution of sensory aroma attributes from coffee beans to brewed coffee. *LWT-Food Science and Technology, 44*, 2185–2192.

Borchgrevink, C. P., Susskind, A. M., & Tarras, J. M. (1999). Consumer preferred hot beverage temperatures. *Food Quality and Preference, 10*, 117–121.

Brown, F., & Diller, K. R. (2008). Calculating the optimum temperature for serving hot beverages. *Burns, 34*, 648–654.

Cristovam, E., Russell, C., Paterson, A., & Reid, E. (2000). Gender preference in hedonic ratings for espresso and espresso-milk coffees. *Food Quality and Preference, 11*, 437–444.

Di Donfrancesco, B., Gutierrez Guzman, N., & Chambers, E. (2014). Comparison of results from cupping and descriptive sensory analysis of colombian brewed coffee. *Journal of Sensory Studies, 29*, 301–311.

Feria-Morales, A. M. (2002). Examining the case of green coffee to illustrate the limitations of grading systems/expert tasters in sensory evaluation for quality control. *Food Quality and Preference, 13*, 355–367.

Fisk, I. D., Kettle, A., Hofmeister, S., Virdie, A., & Kenny, J. S. (2012). Discrimination of roast and ground coffee aroma. *Flavour, 1*, 14.

Giacalone, D., Duerlund, M., Begh-Petersen, J., Bredie, W. L., & Frst, M. B. (2014). Stimulus collative properties and consumers' flavor preferences. *Appetite, 77*, 20–30.

Giacalone, D., Fosgaard, T. R., Steen, I., & Münchow, M. (2016). Quality does not sell itself: Divergence between objective product quality and preference for coffee in naïve consumers. *British Food Journal, 118*, 2462–2474.

Hayes, J. E., Wallace, M. R., Knopik, V. S., Herbstman, D. M., Bartoshuk, L. M., & Duffy, V. B. (2010). Allelic variation in tas2r bitter receptor genes associates with variation in sensations from and ingestive behaviors toward common bitter beverages in adults. *Chemical Senses, 36*, 311–319.

Hervé, M. (2015). Diverse basic statistical and graphical functions (rvaide memoire). r package.

Husson, F., et al. (2005). Multiple factor analysis with confidence ellipses: A methodology to study the relationships between sensory and instrumental data. *Journal of Chemometrics, 19*, 138–144.

ISO (2008). Iso 6668:2008. green coffee -preparation of sample for use in sensory analysis.

Lawless, H. T., & Heymann, H. (2010). *Sensory evaluation of food: Principles and practices.* Springer Science & Business Media.

Lê, S., Josse, J., Husson, F., et al. (2008). Factominer: An R package for multivariate analysis. *Journal of Statistical Software, 25*, 1–18.

Lee, H., & O'Mahony, M. (2002). At what temperatures do consumers like to drink coffee?: Mixing methods. *Journal of Food Science, 67*, 2774–2777.

Masi, C., Dinnella, C., Barnabà, M., Navarini, L., & Monteleone, E. (2013). Sensory properties of under-roasted coffee beverages. *Journal of Food Science, 78*, 1290–1300.

Masi, C., Dinnella, C., Monteleone, E., & Prescott, J. (2015). The impact of individual variations in taste sensitivity on coffee perceptions and preferences. *Physiology & Behavior, 138*, 219–226.

Masi, C., Dinnella, C., Pirastu, N., Prescott, J., & Monteleone, E. (2016). Caffeine metabolism rate influences coffee perception, preferences and intake. *Food Quality and Preference, 53*, 97–104.

Meyners, M., Castura, J. C., & Carr, B. T. (2013). Existing and new approaches for the analysis of CATA data. *Food Quality and Preference, 30*, 309–319.

Münchow, M. (2016). *SCAE roasting foundation.* CoffeeMind Press.

Nair, K. P. (2010). *The agronomy and economy of important tree crops of the developing world.* Elsevier.

Ponte, S. (2002). The 'latte revolution'? regulation, markets and consumption in the global coffee chain. *World Development, 30*, 1099–1122.

R Core Team. (2014). R: A language and environment for statistical computing. r foundation for statistical computing, Vienna, Austria. 2014.

Ribeiro, J., Augusto, F., Salva, T., & Ferreira, M. (2012). Prediction models for arabica coffee beverage quality based on aroma analyses and chemometrics. *Talanta, 101*, 253–260.

Ribeiro, J., Augusto, F., Salva, T., Thomaziello, R., & Ferreira, M. (2009). Prediction of sensory properties of brazilian arabica roasted coffees by headspace solid phase microextraction-gas chromatography and partial least squares. *Analytica Chimica Acta, 634*, 172–179.

Ribeiro, J., Ferreira, M., & Salva, T. (2011). Chemometric models for the quantitative descriptive sensory analysis of arabica coffee beverages using near infrared spectroscopy. *Talanta, 83*, 1352–1358.

SCAA (2009). SCAA Protocols – Cupping for specialty coffee. Specialty Coffee Association of America.

Schenker, S., Handschin, S., Frey, B., Perren, R., Escher, F. (1999). Structural properties of coffee beans as influenced by roasting conditions. In ASIC, 18e colloque, Helsinki.

Schenker, S., Handschin, S., Frey, B., Perren, R., & Escher, F. (2000). Pore structure of coffee beans affected by roasting conditions. *Journal of Food Science, 65*(3), 452–457.

Steen, I., Waehrens, S. S., Petersen, M. A., Münchow, M., & Bredie, W. L. (2017). Influence of serving temperature on flavour perception and release of *Bourbon caturra* coffee. *Food Chemistry, 219*, 61–68.

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 and Preference, 32*, 152–159.

Wintgens, J. (2009). Green coffee defects. Coffee: Growing, Processing, Sustainable Production: A Guidebook for Growers, Processors, Traders, and Researchers, pages 758–788.

Yang, N., Liu, C., Liu, X., Degn, T. K., Munchow, M., & Fisk, I. (2016). Determination of volatile marker compounds of common coffee roast defects. *Food Chemistry, 211*, 206–214.