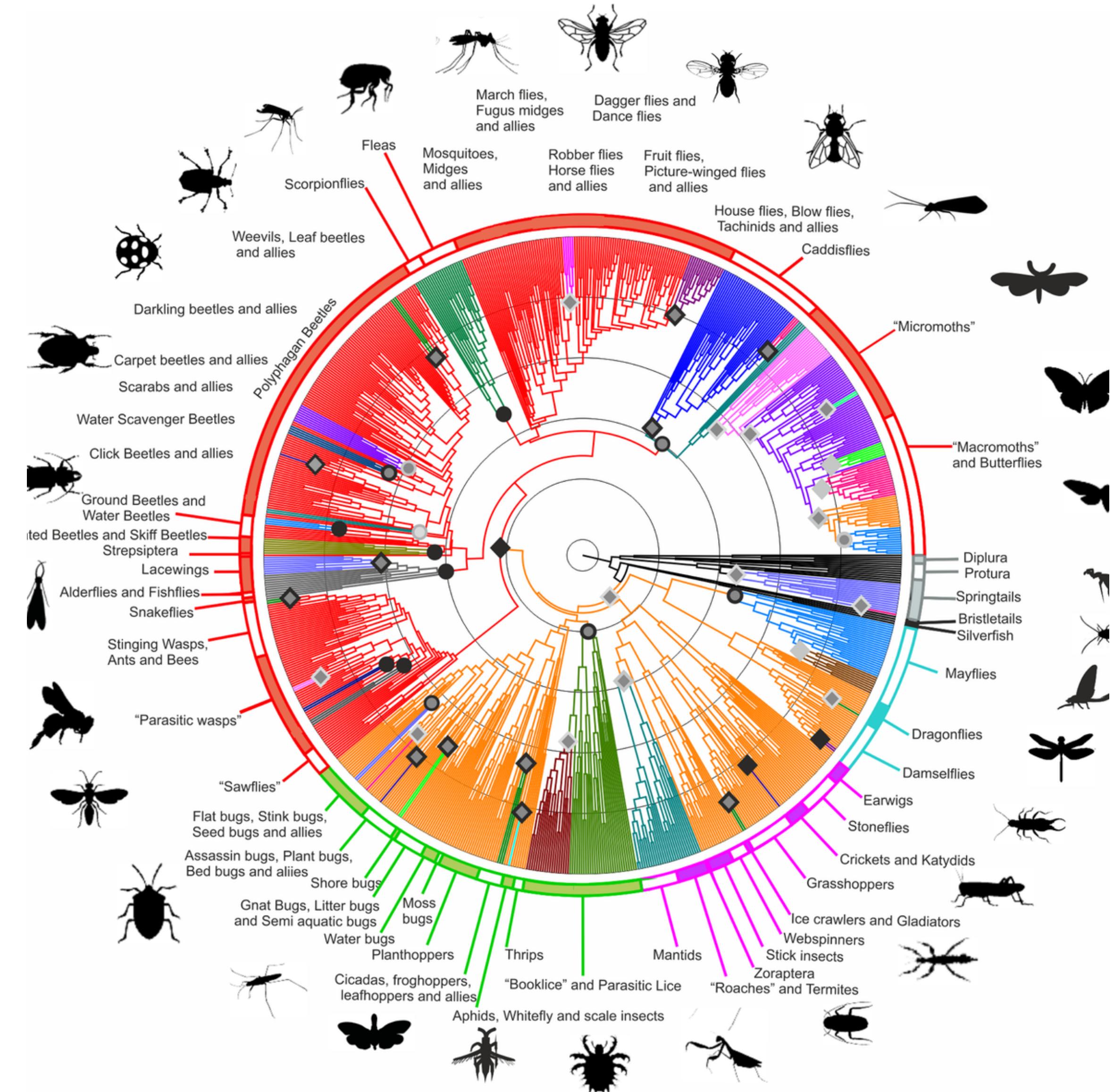


An introduction to HMFA

S. Lê



Rainford JL, Hofreiter M, Nicholson DB, Mayhew PJ (2014)
Phylogenetic Distribution of Extant Richness Suggests Metamorphosis
Is a Key Innovation Driving Diversification in Insects.
PLoS ONE 9(10): e109085.

Case study 1

MTMM data



A quick introduction to **MTMM** data

<http://davidakenny.net/cm/mtmm.htm>

- A set of t traits are each measured by m methods.
- The resulting data are tm measures, and the correlation matrix is called a multitrait-multimethod matrix.
- The matrix was originally proposed by Donald T. Campbell and Donald Fiske (1959). The matrix is commonly abbreviated as MTMM.

A quick introduction to **MTMM** data

<http://davidakenny.net/cm/mtmm.htm>

- **Convergent validity:** measures of the same trait should be strong (Same-trait, different-method correlations are in bold and called the validity diagonal.).
- **Discriminant validity:** a measurement method should discriminate between different traits. Different-trait, different-method correlations should not be too high, especially relative to same-trait, different-method correlations.
- **Method variance:** variance due to method can be detected by seeing if the different-trait, same-method correlations are stronger than the different-trait, different-method correlations.

Example of MTMM matrix

- Construct validity of assessment centers: Appropriate use of confirmatory factor analysis and suitable construction principles
- Journal of social behavior and personality 12(5):65-84
- Martin Kleinmann & Olaf Köller

	Group Discussion			Role Playing			Presentation		
	OC	A/J	IS	OC	A/J	IS	OC	A/J	IS
Group Discussion									
Oral Communication	1.00								
Analysis/Judgment	.58	1.00							
Interpersonal Skills	.64	.45	1.00						
Role Playing									
Oral Communication	.71	.46	.45	1.00					
Analysis/Judgment	.52	.47	.43	.69	1.00				
Interpersonal Skills	.45	.36	.35	.61	.82	1.00			
Presentation									
Oral Communication	.54	.35	.42	.65	.52	.37	1.00		
Analysis/Judgment	.34	.29	.24	.40	.42	.34	.76	1.00	
Interpersonal Skills	.24	.22	.22	.19	.31	.28	.51	.64	1.00

Note: GD: Group Discussion; RP: Role Playing; PR: Presentation; OC: Oral Communication; A/J: Analysis/Judgment; IS: Interpersonal Skills. All correlations are significant: $p < .05$ (one-tailed). $N = 70$.

NICHD Study of Early Child Care and Youth Development

<https://www.nichd.nih.gov/research/supported/seccyd>

- SECCYD began as the Study of Early Child Care in 1991.
- Working with more than 1,300 children and their families, the researchers analyzed how different child care arrangements related to measurements of the children's health, behavior, school performance, and other indicators of development in infancy, early childhood, middle childhood, and middle adolescence.

Longitudinal Multitrait-Multimethod Models for Developmental Research

Kevin J. Grimm, Robert C. Pianta & Timothy Konold

- In this study, children are assessed by their mother, father, and teacher in first, third, fourth, and fifth grade according to three criteria:
 - their social skills,
 - the concept of external behavior problem,
 - the concept of internal behavior problem.

Longitudinal Multitrait-Multimethod Models for Developmental Research

Kevin J. Grimm, Robert C. Pianta & Timothy Konold

- We also have at our disposal the following variables:
 - the ethnic group (Ethnic),
 - the gender (Gender),
 - whether the family was considered poor (Poor; 0 = no, 1 = yes),
 - the number of caregiver hours per week working outside the home (Caregiver),
 - the income-to-needs ratio (Incm_nds; 0 = no income, 1 = income just meets needs),

Longitudinal Multitrait-Multimethod Models for Developmental Research

Kevin J. Grimm, Robert C. Pianta & Timothy Konold

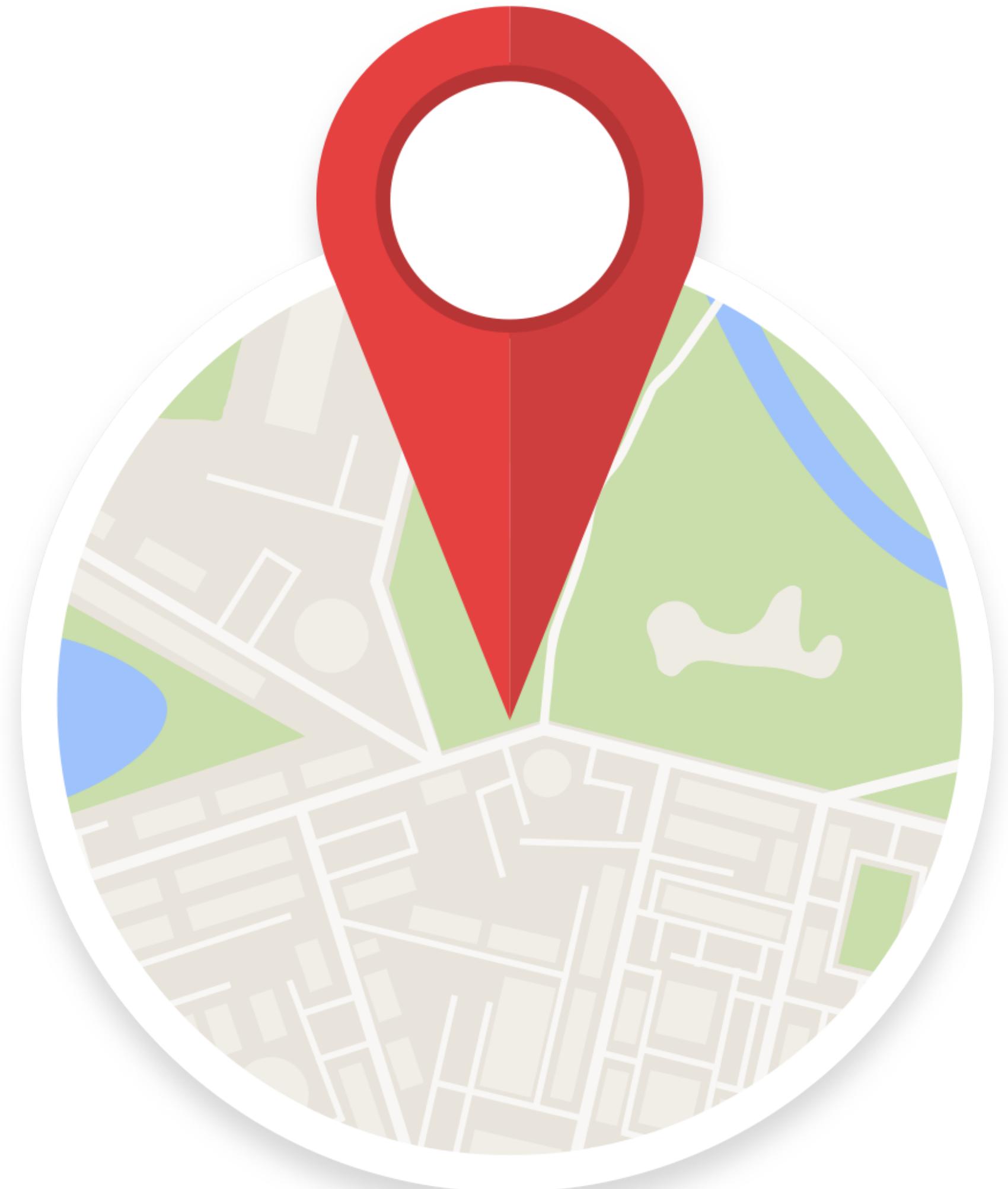
- We also have at our disposal the following variables:
 - the mother years of education (Mom_ed; centered at 12),
 - the father years of education (Dad_ed; centered at 12),
 - the mother depression score (Mom_dep),
 - the father depression score (Dad_dep),
 - the maternal sensitivity score (Mom_sen),
 - the father sensitivity score (Dad_sen).

Analysing multitrait multimethod data with exploratory multivariate analysis... the French way: a multiple factor analysis perspective

Lê, S.

- The dataset comprises 247 rows (children), and 47 variables, our main variables of interest being the 36 child behavior measures:
 - for each grade, each child is measured according to 3 variables (social skills, external and internal behaviour),
 - and 3 assessors (mother, father, and teacher);
 - each child is measured during 4 years

**Which question do
you want to
address?
How would you structure
your data?**



Case study 2

Multiway data



Multiway data analysis

From Wikipedia, the free encyclopedia

Multiway data analysis is a method of analyzing large data sets by representing the data as a [multidimensional array](#). The proper choice of array dimensions and analysis techniques can reveal patterns in the underlying data undetected by other methods.^[1]

Contents [hide]

- 1 History
- 2 Composition of multiway data analysis
 - 2.1 Multiway data
 - 2.2 Multiway model
 - 2.3 Multiway application
 - 2.4 Multiway processing
- 3 See also
- 4 References

History [edit]

The study of multiway data analysis was first formalized as the result of a conference held in 1988. The result of this conference was the first text specifically addressed to this field, Coppi and Bolasco's *Multiway Data Analysis*.^[2] At that time, the application areas for multiway analysis included [statistics](#), [econometrics](#) and [psychometrics](#). In recent years, applications have expanded to include [chemometrics](#), [agriculture](#), [social network analysis](#) and the [food industry](#).^[3]

Composition of multiway data analysis [edit]

Multiway data [edit]

Multiway data analysts use the term *way* to refer to a dimension of the data while reserving the word *mode* for the methods or models used to analyze the data.^{[2]:xviii}

In this sense, we can define the various *ways* of data to analyze:

- *One-way data* is a [vector](#), with a single data value for each discrete or continuous value of the single dimension.
- *Two-way data* is a [matrix](#), with a single data value for each discrete or continuous value of two separate dimensions; a [spreadsheet](#) can be used to visualize such data in the case of discrete dimensions.
- *Three-way data* can be viewed as a stack of matrices (or similarly, as a workbook of multiple [spreadsheets](#)), adding a third dimension. Such data might represent the temperature at different locations (two-way data) sampled over different times (the third dimension, leading to three-way data)

An introduction to multiway data and their analysis

Computational Statistics & Data Analysis 18 (1994) 3-13

Renato Coppi

Multiway data: definitions and notation

In traditional multivariate analysis the “data matrix” is usually defined as a set of “elementary data” x_{ij} ($i = 1, I$; $j = 1, J$) concerning the values taken by J variables X_j on I statistical units (individuals, areas, etc.). In this case, we are dealing with “two-way data”, since each element of the data set is characterised by a pair of indices: i for the mode “units” and j for the mode “variables”.

When the “elementary datum” is referred to three or more indices (or ways, or criteria of classification), we get “multiway data”. A common example of this type of data is provided by the “three-way array”

$$X = \{x_{ijk}\}, \quad i = 1, I; j = 1, J; k = 1, K,$$

where i denotes the “units”, j denotes the “variables” and k denotes the “occasions” (different times of observation, different locations, different experimental conditions, and so on).

Correspondence to: Renato Coppi, Dipartimento di Statistica, Probabilità e Statistiche Applicate, Università “La Sapienza”, P.le A. Moro 5, 00185, Roma, Italy.

0167-9473/94/\$07.00 © 1994 – Elsevier Science B.V. All rights reserved
SSDI 0167-9473(94)00003-2

mental conditions, and so on). In this case, the three ways are fully crossed, namely on each occasion the same I units are observed with respect to the same J variables. Moreover, each “way” corresponds to a different “mode” (respectively: units, variables, occasions). We get a “three-way, three-mode array”.

The Three-Mode Company

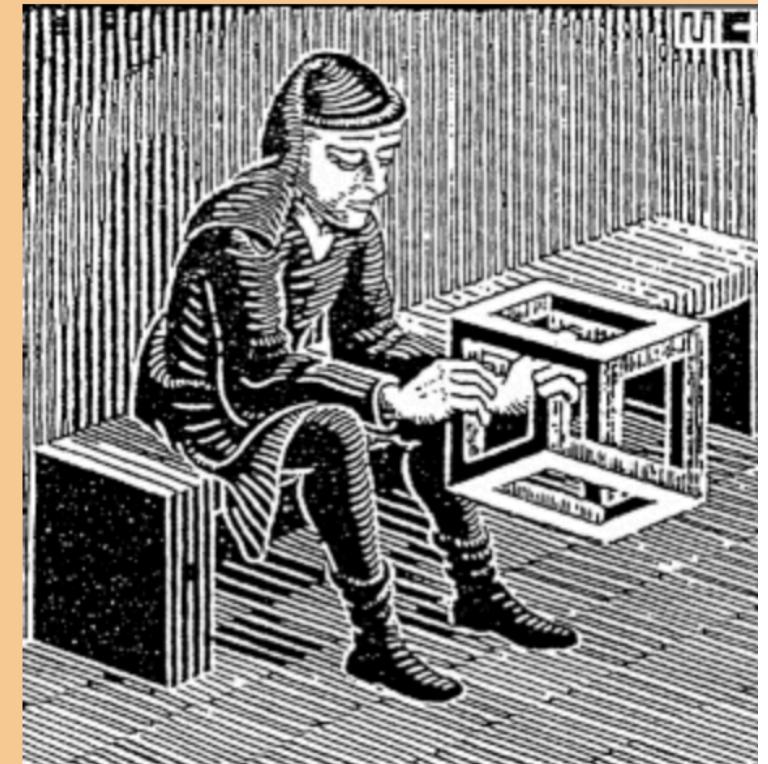
<http://three-mode.leidenuniv.nl>



The screenshot shows a browser window with the URL "three-mode.leidenuniv.nl" in the address bar. The page title is "The Three-Mode Company" in large green letters, with the subtitle "any devoted to creating three-mode software and promoting three-mode data analysis" below it. A navigation bar at the top includes links for "I. Kroonenberg", "University", "Leiden, The Netherlands", "Faculty of Social and Behavioural Sciences", "Centre for Child & Family Studies", and "HOME".

The Three-Mode Company is dedicated to promoting all of the analysis of three-way data. To this end it carries out the following activities.

- Developing and distributing of [3WayPack](#) a user-friendly software package for three-mode analysis.
- Maintaining a [bibliography](#) of theoretical and applied papers on three-mode analysis.
- Maintaining a [repository](#) of full-text three-mode books (such as [Kroonenberg, 1983](#)).
- Maintaining an [address list](#) (including home page links) of active researchers in the field.
- Building a library of publicly available three-mode [data sets](#).
- Maintaining a list of [available software](#) for three-mode analysis.
- Collecting photographs of all researchers connected one way or another with three-mode analysis.



The Three-Mode Data Analyst

This Three-Mode Data Analyst is officially called *Man with Cuboid* and is displayed here by courtesy of the artist. Escher's work is protected by copyright. Any unauthorized reproduction is prohibited.

Applied Multiway Methods

Applied Multiway Data Analysis. Hoboken, NJ: Wiley
Peter M. Kroonenberg
BN: 978-0-470-16497-6 (Hardcover), 608 pages, January 2008
S \$125 [Wiley US](#) / Euro 104 [Wiley UK](#) / UK Stirling 84;

Analyzing four-way sensory data resulting from individual vocabulary profiling of audio: A comparison between HMFA and PARAFAC2

**9th Sensometrics Meeting
2008 July 22nd, Ontario, Canada**



**Gaëtan Lorho
Nokia Corporation**

Sensory evaluation in the field of multimedia

- Any system/application capturing, processing or reproducing audio/visual material
 - We want to evaluate a system
 - Our stimulus is a media clip ‘interacting’ with this system
 - Number of stimuli = number of systems × number of media clips

Loudspeaker systems



Camera optics systems



Video encoding algorithms

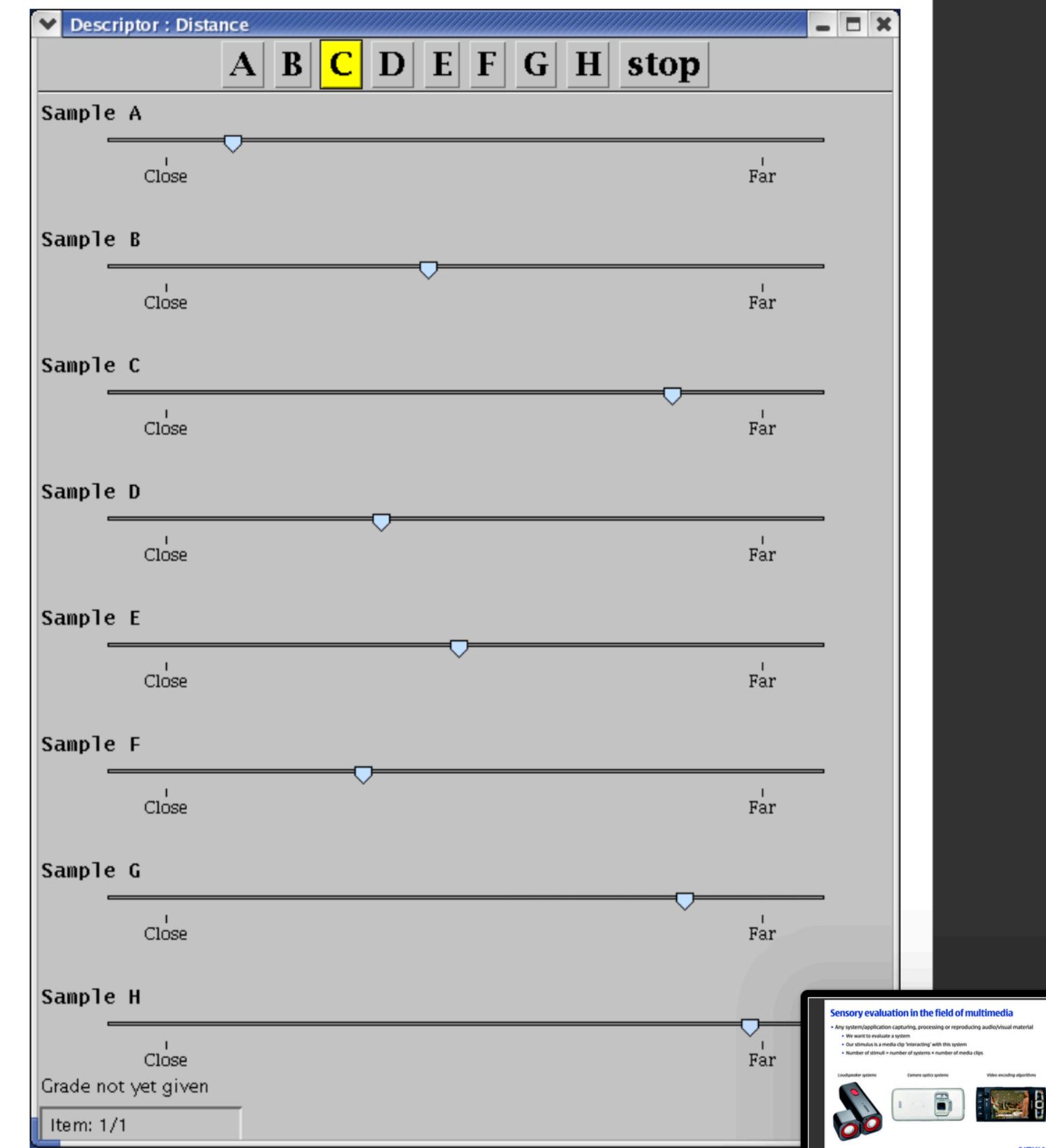


NOKIA

An example of audio sensory profiling

- A sensory evaluation of techniques to reproduce music over headphones
 - 8 'systems' evaluated
 - 3 'music clips' considered
- Sensory profiling inspired by *Flash Profile*
 - Comparative evaluation approach
 - 8 systems for a given music clip
- Outcome for each assessor:
 - ⇒ One set of attributes
 - ⇒ Three sensory profiles

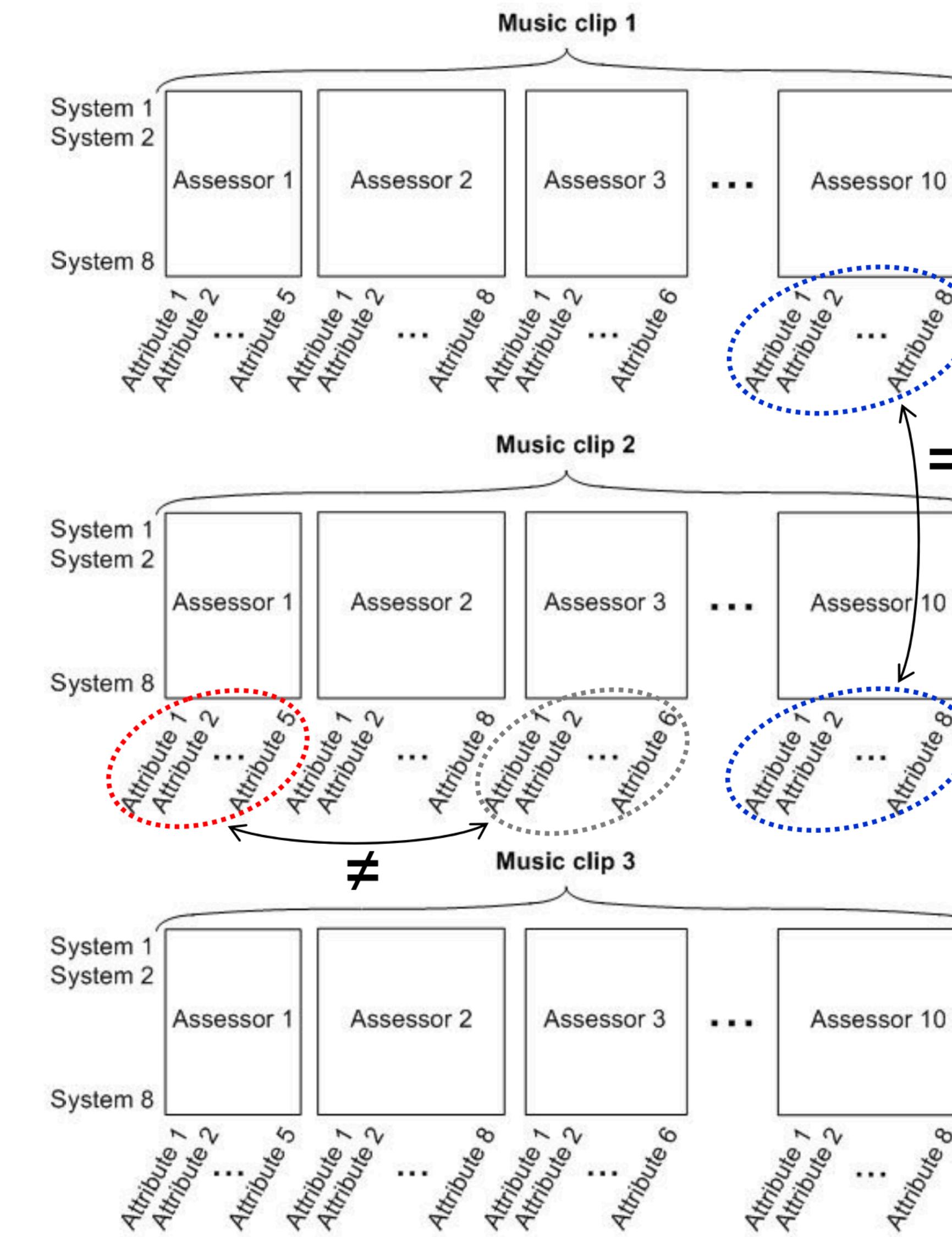
*Final attribute scaling UI
One Attribute, One music clip, 8 systems*



NOKIA

The resulting four-way sensory dataset

- 8 systems
- 3 music clips
- 10 assessors
- 4 to 8 attributes per assessor



**Which question do
you want to
address?
How would you structure
your data?**



Case study 3

A bit of **lightness** in this
world of **heavies**



Our starting point

- APA Lokki, T., Pätynen, J., Kuusinen, A., Vertanen, H., & Tervo, S. (2011). ~~Concert hall acoustics assessment with individually elicited attributes. The Journal of the Acoustical Society of America, 130(2), 835-849.~~
- ISO 690 LOKKI, Tapio, PÄTYNEN, Jukka, KUUSINEN, Antti, et al. ~~Concert hall acoustics assessment with individually elicited attributes. The Journal of the Acoustical Society of America, 2011, vol. 130, no 2, p. 835-849.~~
- MLA Lokki, Tapio, et al. "Concert hall acoustics assessment with individually elicited attributes." ~~The Journal of the Acoustical Society of America~~ 130.2 (2011): 835-840.

The Helsinki Conservatory of Music



Sello Hall in Espoo Finland

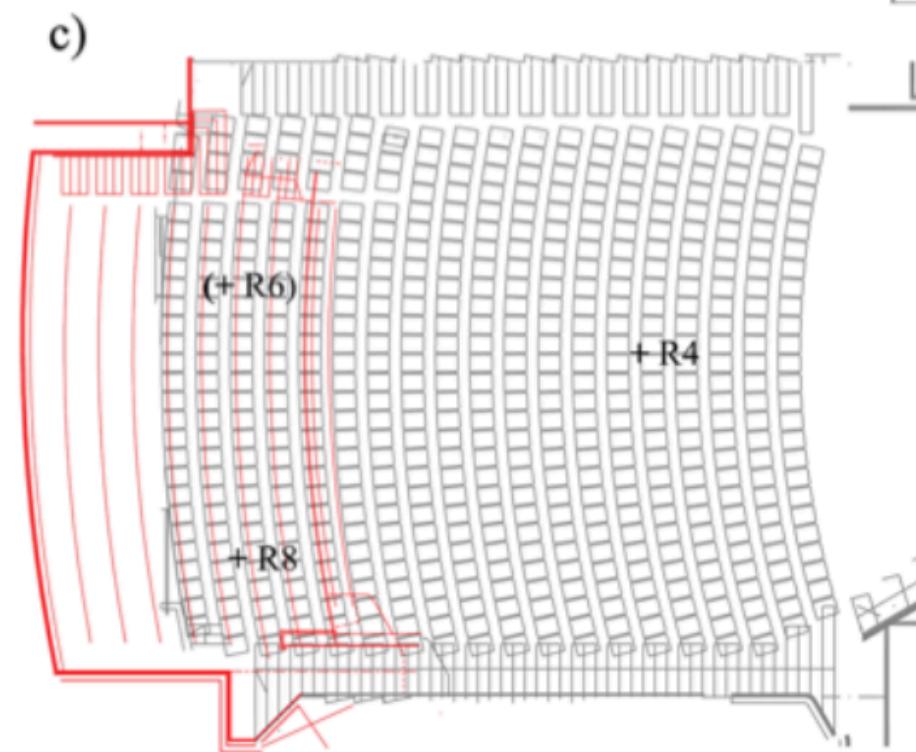
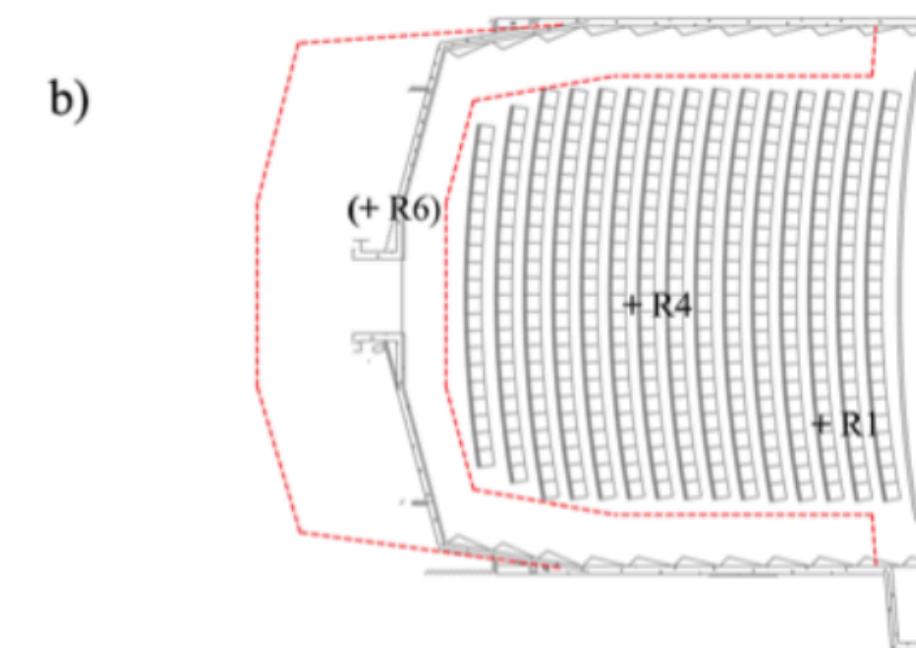
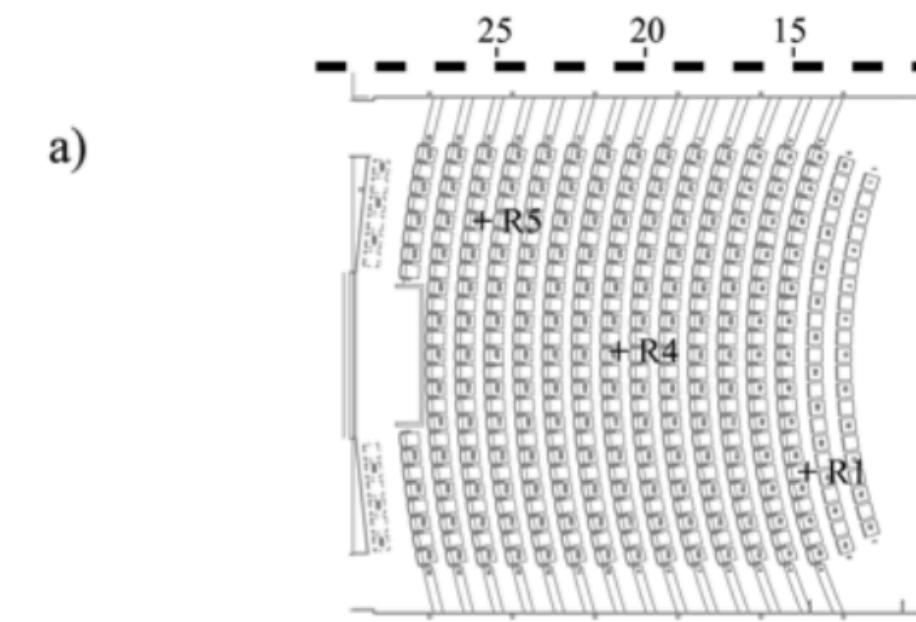


Tapiola Hall in Espoo Finland



Concert hall acoustics assessment with individually elicited attributes

- As we said, 3 concert halls
- 3 positions in each concert halls



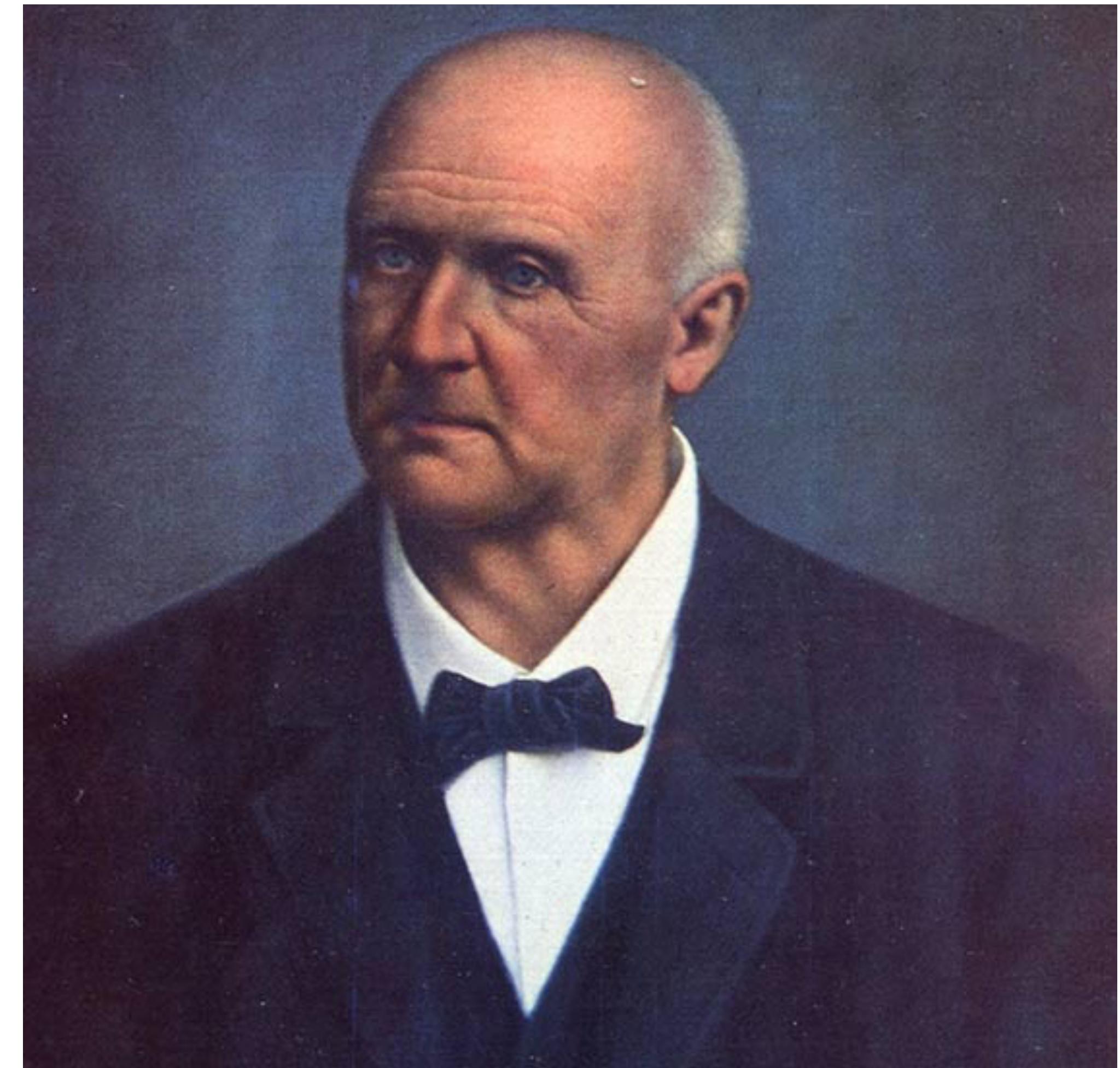
Concert hall acoustics **assessment** with individually elicited attributes

- Mozart, An aria of Donna Elvira from the opera Don Giovanni, Act II, Scene III, bars 1– 15, 26.0 s



Concert hall acoustics **assessment** with individually elicited attributes

- Bruckner, Symphony no. 8, movement II,
bars 41–61, 29.0 s



Concert hall acoustics **assessment** with individually elicited attributes

- Mahler, Symphony no. 1, movement IV, bars 71–86, 17.5 s



Concert hall acoustics **assessment** with individually elicited attributes

- Beethoven, Symphony no. 7, movement I,
bars 12–18, 23.5 s



Concert hall acoustics **assessment** with individually elicited attributes

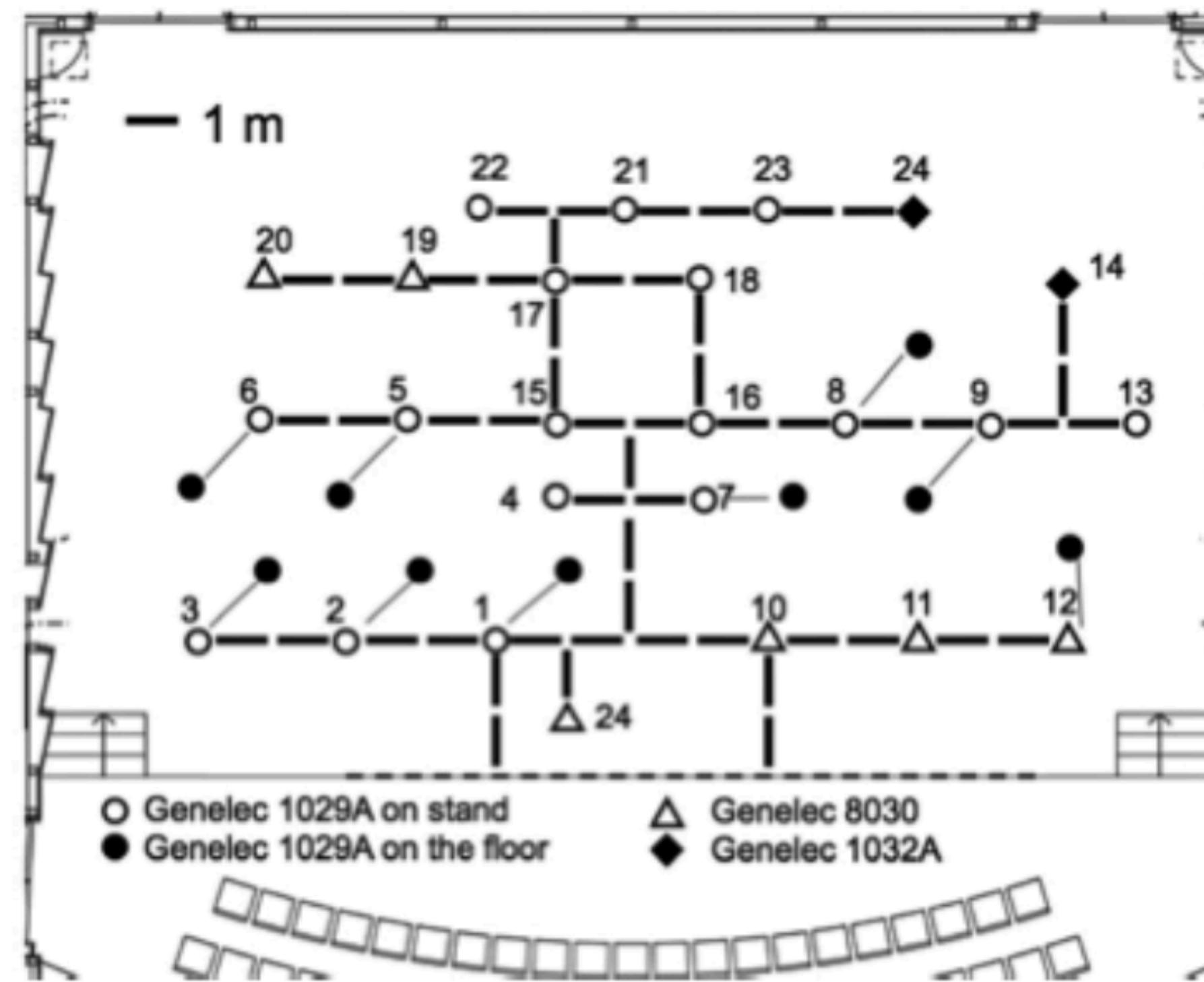


FIG. 3. Layout of the loudspeaker orchestra on the stage of a concert hall.

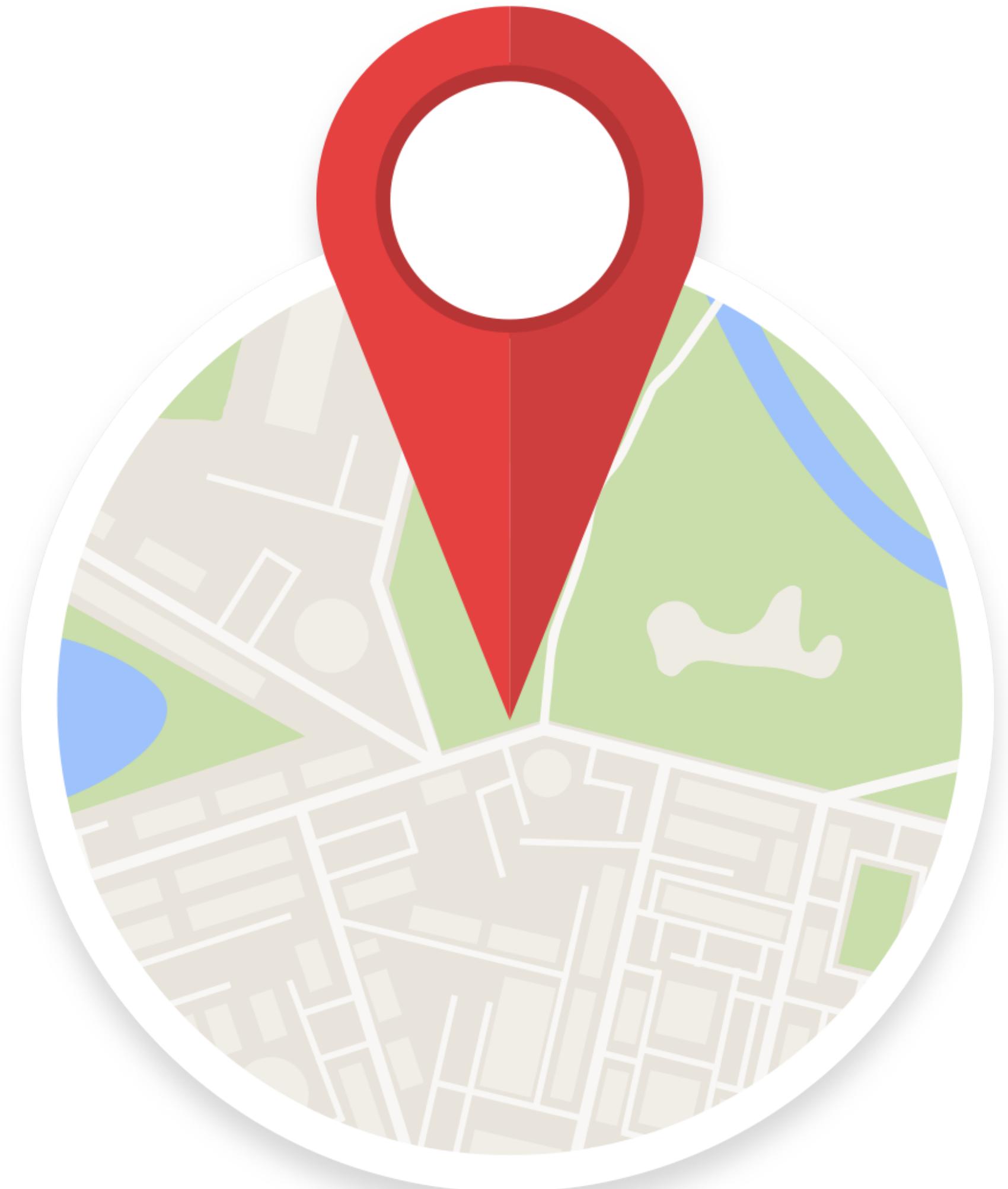
Concert hall acoustics assessment with individually elicited attributes

- Clarity
- Amount of reverb
- Deepness
- Distance
- Shape of space
- Reverberance
- Sharpness
- Intimacy
- Reverberation
- ...

What about the data?

- 20 assessors
- 9 stimuli per assessor (3 halls, 3 positions)
- 4 to 6 descriptors per assessor
- 4 music samples

**Which question do
you want to
address?
How would you structure
your data?**



Case study 4

Classifying genotypes based on **phenotypic** and **genetic** data



Crop Breeding & Genetic

Hierarchical Multiple-Factor Analysis for Classifying Genotypes Based on Phenotypic and Genetic Data

Jorge Franco, José Crossa, Santosh Deshpande

First published: 01 January 2010 | <https://doi.org/10.2135/cropsci2009.01.0053>

All rights reserved. No part of this periodical may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Permission for printing and for reprinting the material contained herein has been obtained by the publisher.

ABSTRACT

A numerical classification problem encountered by breeders and gene-bank curators is how to partition the original heterogeneous population of genotypes into non-overlapping homogeneous subpopulations. The measure of distance that may be defined depends on the type of variables measured (i.e., continuous and/or discrete). The key points are whether and how a distance may be defined using all types of variables to achieve effective classification. The objective of this research was to propose an approach that combines the use of hierarchical multiple-factor analysis (HMFA) and the two-stage Ward Modified Location Model (Ward-MLM) classification strategy that allows (i) combining different types of phenotypic and genetic data simultaneously; (ii) balancing out the effects of the different phenotypic, genetic, continuous, and discrete variables; and (iii) measuring the contribution of each original variable to the new principal axes (PAs). Of the two strategies applied for developing PA scores to be used for clustering genotypes, the strategy that used the first few PA scores to which phenotypic and genetic variables each contributed 50% (i.e., a balanced contribution) formed better groups than those formed by the strategy that used a large number of PA scores explaining 95% of total variability. Phenotypic variables account for much variability in the initial PA; then their contributions decrease. The importance of genetic variables increases in later PAs. Results showed that various phenotypic and genetic variables made important contributions to the new PA. The HMFA uses all phenotypic and genetic variables simultaneously and, in conjunction with the Ward-MLM method, it offers an effective unifying approach for the classification of breeding genotypes into homogeneous groups and for the formation of core subsets for genetic resource conservation.

Case study 5

On 28/Oct/2019, I sent you the abstract below, which was submitted to Food Quality and Preference.

I would be grateful if you could take the time to read this paper and comment on its suitability for publication.



Exercise

Run an HMFA with a PCA/
MCA program...in R

