Received: 29 June 2020 | Revised: 24 September 2020 | Accepted: 24 September 2020
DOI: 10.1111/joss.12620

ORIGINAL ARTICLE

Journal of Sensory Studies WILEY

Sound quality perception of loudspeakers evaluated by different sensory descriptive methods and preference mapping

Ingunn Berget[1] | Søren Bech[2,3] | Davide Giacalone[4] | Samuel Moulin[2] | Mads Erling Pedersen[1] | Paula Varela[1] | Tormod Næs[1]

[1]Nofima, Ås, Norway
[2]Bang og Olufsen a/s, Struer, Denmark
[3]Department of Electronic Systems, The Technical Faculty of IT and Design, Signal and Information Processing, Aalborg University, Aalborg, Denmark
[4]Department of Technology and Innovation, Faculty of Engineering, University of Southern Denmark, Odense, Denmark

Correspondence
Ingunn Berget, Nofima, Ås, Norway.
Email: ingunn.berget@nofima.no

Funding information
Norges Forskningsråd, Grant/Award Number: NFR 262308

Abstract

Quantitative Descriptive Analysis (QDA) was compared with novel sensory methodology on five high-end loudspeakers using five tracks. The focus of the study was on the ability of different sensory methods to predict quality assessments (liking) made by consumers. Preference mapping indicates that projective mapping (PM) and Check-All-That-Apply (CATA) when performed with experts can provide results comparable to QDA, although slightly less accurate. PM done by consumers is less comparable to QDA. Combinations of the sensory tests were also attempted without any clear improvement of results.

Practical applications: The results indicate that it may be possible to replace QDA with more rapid sensory methodology for assessing sensory quality of audio products. This may speed up and simplify product development process. However, further studies will be needed to consolidate the results.

1 | INTRODUCTION

Sensory analysis of perceived sound quality and the relationship to the physical variables defining the sound field has always been an important issue in both the professional and audiophile communities. Scientific analysis and understanding of how to quantify either general sound quality as perceived by consumers, or specific attributes perceived by trained listeners have, however, been lagging behind for many years. The physical variables with an assumed influence on sound quality, for example frequency response or distortion characteristics, of the audio system and international standards for how to quantify these variables have, on the other hand, been known for years. The development and commercial introduction of low-bit-rate systems like mp3 raised an increased awareness and need for listening tests or sensory analysis of such systems.

An important goal has been to produce scientifically valid results that could be used to develop physical measures correlating with the perceptual aspects. Furthermore, there was also a need for new paradigms for listening tests that were sensitive enough to quantify the—often very small—perceived differences between different commercial implementations of the standardized codecs. This resulted in an increased effort to quantify and understand perceived sound quality. Methods like Quantitative Descriptive sensory Analysis (QDA, Stone, Sidel, Woolsey, & Singleton, 2008) and preference mapping originally developed for sensory analysis of food quality were highly inspirational in that process. A detailed summary and references are given in Bech and Zacharov (2006).

Quantitative descriptive sensory analysis (QDA, Stone et al., 2008) is the most widely used descriptive sensory method. While QDA is known to produce reliable and precise quantification of sensory differences between products, it is also characterized by some limitations, primarily due to the extensive training required and associated time and cost requirements. To address these limitations new rapid sensory methods have been proposed (Varela & Ares, 2012). Unlike QDA, rapid methods require little to no training and thus are much faster to apply, which makes them advantageous in a product development context where time is of the essence and rapid feedbacks, albeit less accurate, are preferable.

Some of these methods, such as Check-All-That-Apply (CATA, Adams, Williams, Lancaster, & Foley, 2007)—where assessors are requested to check attributes from a predefined list and projective mapping (Risvik, McEwan, Colwill, Rogers, & Lyon, 1994)—where assessors generate two-dimensional product maps based on perceived similarity and differences, have been gaining popularity in recent years. PM is also referred to as Napping (Pagès, 2003). There are numerous publications supporting their validity for sensory assessment for both food and non-food products. Besides speed, another aspect is that these methods are supposedly more holistic and intuitive, since they are not limited to a sensory profile defined by experts. This may make data more relevant to consumer preferences than sensory spaces generated by trained assessors. For instance, multivariate configurations from PM data reflect the relative importance given by the assessors to specific sensory aspects, not necessarily the attribute with the most variance, as in a principal component analysis (PCA) based on QDA data (Pagès, 2005). This might be more relevant to hedonics, as it has been suggested that PM data collected from untrained consumers may conceptually be closer to an internal preference map than to a QDA product space (Torri et al., 2013). Projective mapping has even been suggested, based on consumers preferences, as a good tool to reveal drivers of liking and choice from a holistic perspective (Varela et al., 2017).

Preference mapping uses multivariate tools to create bi-dimensional plots (maps) that link product characteristics and consumer preferences. Preference mapping is one of the most important methodologies for identifying the sensory attributes that are the main drivers of liking and the most preferred products(s) for different consumer groups. It is used with success in a large number of publications (see e.g., Meullenet, Xiong, & Findlay, 2007; Næs, Varela, & Berget, 2018) and has become a standard tool in sensory studies of food, but also other product categories, including audio (Mattila, 2001; Zacharov & Koivuniemi, 2001). In order to do preference mapping, liking/preference data obtained from consumers and a descriptive sensory profile for the same products are needed. Typically, the two data sets are first analyzed separately using ANOVA to study product differences with respect to liking and individual sensory attributes, and then combined using some type of regression analysis.

There are conceptually two different ways of performing preference mapping. In *internal preference mapping* the sensory data is modeled from the consumer data while in *external preference mapping* the order is switched, so that consumer preferences are predicted from sensory data. Both approaches have their pros and cons as discussed in for instance (Næs et al., 2018). In most cases, principal component analysis (PCA) is used first on the independent data block, then the responses are regressed onto the components (principal component regression or PCR), but other dimension reduction techniques can be applied depending on the data. Jaeger, Wakeling, and MacFie (2000) highlighted the potential issue in preference mapping based on QDA that trained panelists and consumers may not perceive products in the same way, and that consumers, when forming their preferences, synthesize sensory perception to form a simplified overview of product characteristics.

In the context of preference mapping, an open question regarding rapid sensory methods is, thus, whether they are equally suitable or can add additional insight to the classical approach with QDA. While several papers have compared rapid sensory methods with respect to their descriptive aspects (e.g., Dehlholm, Brockhoff, Meinert, Aaslyng, & Bredie, 2012; Reinbach, Giacalone, Ribeiro, Bredie, & Frøst, 2014), little research has been published on comparing their performance in a preference mapping context. A few examples in the food domain can be found in Ares, Varela, Rado, and Giménez (2011), Dooley, Lee, and Meullenet (2010), Giacalone, Bredie, and Frøst (2013), which used CATA and/or projective mapping in a preference mapping context. When comparing to the classical approach some authors concluded that these methods give similar results to external preference mapping. Other applications using rapid methods can be found in Varela, Beltrán, and Fiszman (2014), using preference ranking and open comments in Withers et al. (2014)), and using taxonomic free sorting and liking in Faye et al. (2006) with focus on liking of textiles.

In this paper, the use of CATA and projective mapping are investigated for the purpose of supplementing or potentially replacing standard QDA for preference mapping in assessing loudspeaker quality. For simplicity PM will be used as an acronym of projective mapping. Attempts will also be made to combine different descriptive methodologies using multi-block regression methodology (Næs, Tomic, Afseth, Segtnan, & Måge, 2013). Both internal and external preference mapping will be considered. The experiments used for the purpose are based on assessment of sound quality of loudspeakers using different music tracks, using data from previous studies where the loudspeakers were assessed by QDA (Moulin, Bech, & Stegenborg-Andersen, 2016), PM (Giacalone et al., 2017), and CATA (Hicks, Moulin, & Bech, 2018).

2 | DATA SETS

2.1 | Experimental strategy

A standard QDA experiment (Moulin et al., 2016) is used as baseline to be compared with the rapid methods CATA (Hicks et al., 2018) and PM (Giacalone et al., 2017). A hedonic liking test provided data for the preference mapping analyses. The experiments were conducted at three different sites by different experimenters. The common stimuli set included five different audio tracks (Table 1) and five different

TABLE 1 Description of tracks

Track	Genre	Title—Artist—Album
R	Rock	*Bombtrack*—Rage Against the Machine—RATM
C	Classical orchestra	*Dance of the Tumblers*—Eiji O./Minnesota Orchestra—Tutti! Orchestral sampler
S	String quartet	*Petite Symphonie à Cordes II. Assez vif*—Scottish Ensemble—Linn 40th Anniversary Collection
J	Jazz	*Sing Sang Sung*—Gordon Goodwin's Big Phat band—Swingin' for the Fences
V	Vocal	*Tom's Diner*—Suzanne Vega—Solitude Standing

Journal of Sensory Studies

TABLE 2 Specifications of the 5 selected loudspeakers

Loudspeaker	Specifications
Loudspeaker 1	3-Way active loudspeaker, sealed box
Loudspeaker 2	4-Way active loudspeaker, sealed box
Loudspeaker 3	3-Way passive loudspeaker, ported box
Loudspeaker 4	3-Way active loudspeaker, ported box
Loudspeaker 5	Full-range electrostatic loudspeaker, bipolar directivity

standardized listening room (IEC TR 60268-13, 1998). The recorded signals were then reproduced over the same headphone system at the different sites. The auralization system has been tested and the results validated for listening test on loudspeaker systems (Bech, Ellermeier, Ghani, Gulbol, & Martin, 2005; Christensen et al., 2005; Hegarty, Choisel, & Bech, 2007). Specific details of the auralization system used for the present series of test are described in Moulin et al. (2016). Please note that both dynamic and static binaural versions of the auralization system were used in the experiments as shown in Figure 1. The static version was used due to practical reasons and because the results of the first experiments indicated that the influ-

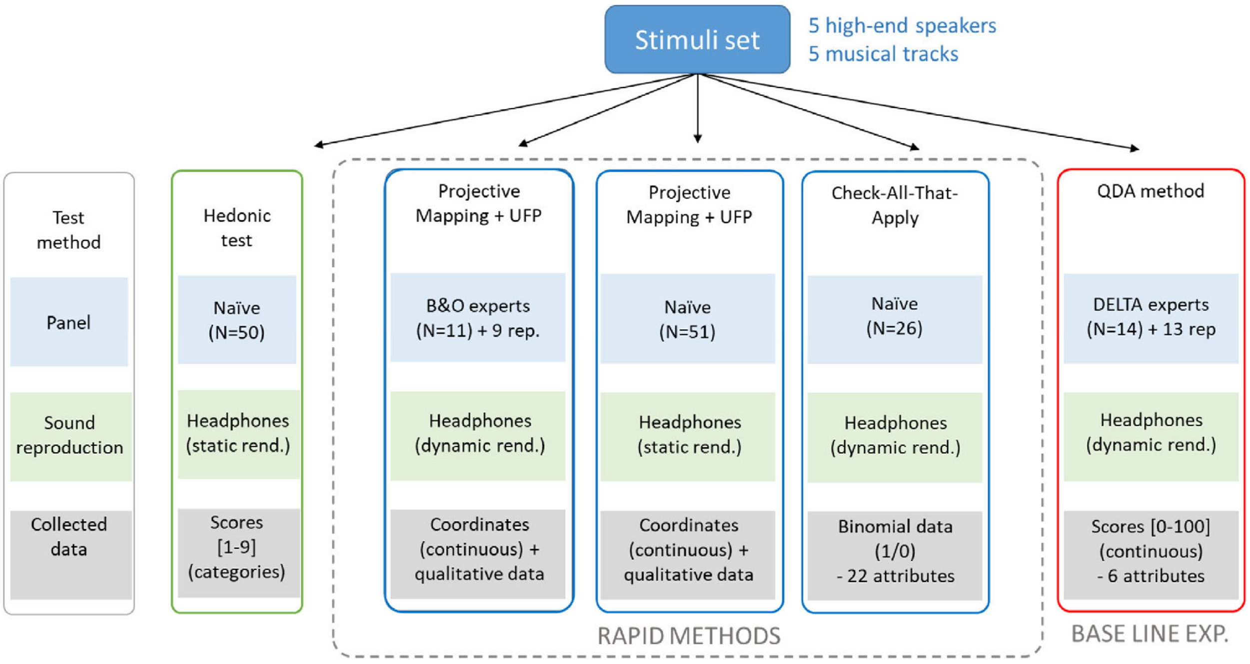

FIGURE 1 Overview of the data sets used in the study

high-end loudspeakers (Table 2). The loudspeakers were selected from Bang & Olufsen's library of loudspeakers that have all been characterized both in term of physical performance (e.g., frequency response and power response) and perceived characteristics (e.g., performance on selected sensory attributes). The loudspeakers were selected to represent both small and large perceived between-loudspeakers differences for both spatial and timbral characteristics. The tracks were selected from Bang & Olufsen's music library to ensure that the differences between the loudspeakers in the spatial and timbral domains would be audible during the tests. Further information on the loudspeakers' specifications and the tracks is given in Moulin et al. (2016). An important factor for the perceived sound quality is the acoustics characteristics of the listening room, and as the test was conducted at different sites this could be a confounding factor that would be difficult to account for in the statistical analysis. To eliminate the problem, an auralization method was used including a dummy-head recording of loudspeakers setup in a stereo configuration, and placed in a ence would have a minor influence on the results. This is confirmed by the analysis conducted in this paper and further discussed in Section 4.6. The experimental strategy of the project is illustrated in Figure 1.

2.2 | Descriptive and hedonic measurements collected

2.2.1 | Quantitative descriptive analysis (QDA)

The base line QDA experiment, was performed by FORCE Technology, SenseLab at their facilities using a trained panel of 14 assessors (Moulin et al., 2016). The panel was trained on a regular basis to perform QDA listening tests. The attributes were selected using the Sound Wheel developed by FORCE Technology (ITU-R, 2017; Pedersen & Zacharov, 2015) and a 2-phase process including:

TABLE 3 CATA attributes. Attributes also included in the QDA are marked with bold letters

Bass depth	Depth (deep)
Bass precision (imprecise)	Depth (flat)
Bass precision (precise)	Detailed
Bass strength (loud)	Enveloping
Bass strength (soft)	Full
Boomy	Midrange strength (loud)
Bright	Midrange strength (soft)
Brilliance (a little)	Powerful
Brilliance (a lot)	Precise
Clean	Treble strength (loud)
Dark	Treble strength (soft)

1. an individual word elicitation/selection based on the Sound Wheel
2. final selection of attributes based on step 1.

This was followed by an individual rating phase where the assessors rated the headphone-based reproduction of the 25 combinations of loudspeaker and track for each attribute using a continuous scale with no labels except the verbal end-points. The test was repeated for all subjects. The selected attributes were (a) Attack (Danish: *attack*), (b) Bass Depth (*Basdybde*), (3) Brilliance (*Brillians*), (4) Dark/Bright (*mørk/lys*), (5) Presence (*naervaerende*), and (6) Width (*Bredde*). The verbal description of each attribute is given in Moulin et al. (2016).

2.2.2 | CATA

The CATA test was performed at Bang & Olufsen's facilities using 26 naïve assessors and a list of 22 attributes (Hicks et al., 2018). The selection of attributes for the CATA list was based on FORCE Technology's Sound Wheel (see Section 2.2.1) where 28 attributes formed the first group of attributes. To identify the relevant attributes for the stimuli to be tested a pilot experiment was conducted using nine naïve assessors. The test included five sessions, one for each track, and the assessors could switch at will between the five loudspeakers reproductions. The assessors were asked to evaluate each of the 28 attributes along with its verbal definition and to rate yes/no based on its ability to describe the differences between the loudspeakers. The assessors were also asked to rate how well they understood the verbal description of the attribute on a continuous scale. The frequency of selection for each attribute and track was calculated and based on a scoring system (Hicks et al., 2018), a list including 15 attributes were selected. This included attributes that had a "double" meaning like Dark—Bright or Depth (flat)—Depth (deep) so they were split into two separate attributes each. The final list thus included the 22 attributes shown in Table 3. The attributes Bass Depth (Basdybde), Brilliance (Brillians), and Dark/Bright (mørk/lys) were common with the QDA. The test procedure first included a familiarization part and then the assessor, for one track at a time, could switch at will between the loudspeaker reproductions and was asked to indicate which attributes from the list should be included for each loudspeaker.

2.2.3 | Projective mapping—PM

Projective mapping (PM) was carried out with two panels, one trained assessor panel (N = 11 and no overlap with the QDA trained panel) and one consumer panel (N = 51), respectively, at Bang & Olufsen and the University of Southern Denmark. Hearing performance was not measured, but it was assumed that students are in the age group which is termed "otologically normal and thus represent the most sensitive listener" (ISO-7029, 2017). In order to increase the size and robustness in the dataset for the trained assessors, nine of them performed a duplicate evaluation, resulting in a total of 20 p.m. evaluations for this panel. Assessors in both panels performed five consecutive PM evaluations of the loudspeakers (one for each track). The order of evaluation for the tracks was randomized between assessors. In both cases, assessors performed the PM task using a data acquisition software that simulated a projective mapping sheet on a computer screen. Assessors were instructed to listen to the audio samples and position them on the map as typically done in PM experiments, that is, close to each other if they sounded similar and further apart if they sounded different. Since PM in itself is essentially a sorting task, it has become customary to instruct the assessors, once they have reached a final configuration, to add a list of sensory descriptors that they find appropriate to describe the samples. This procedure is known as Ultra-Flash Profiling (UFP, Perrin and Pagès (2009) and all assessors (both trained and consumer panel) were asked to provide a verbal description (free elicitation) of the stimuli. Experts on sound quality simplified comments and organized the descriptions into words/phrases as described in Varela and Ares (2012). Additional information on the PM experiments are given in Giacalone et al. (2017).

The PM data consist of the X (horizontal axis) and Y (vertical axis) coordinates on the PM for each loudspeaker captured by the data acquisition software using the bottom left corner as the origin of the coordinate system. The UFP data consisted of the frequency of mention of each descriptor for each loudspeaker and are used as supplementary data in the analyses (MFA, multiple factor analysis) and guides interpretation of the components identified. Only words used more than five times are included in the analyses. The two tests for consumers and experts are here called PMC (C for consumers) and PME (E for experts), respectively and are analyzed separately.

2.2.4 | Hedonic test

The hedonic liking test was conducted using 99 naïve listeners representing the average consumer (42.7% women, mean age was 23.8 and the SD of the groups was 4.8) at the University of Southern Denmark. All consumers participated in the experiment voluntarily with

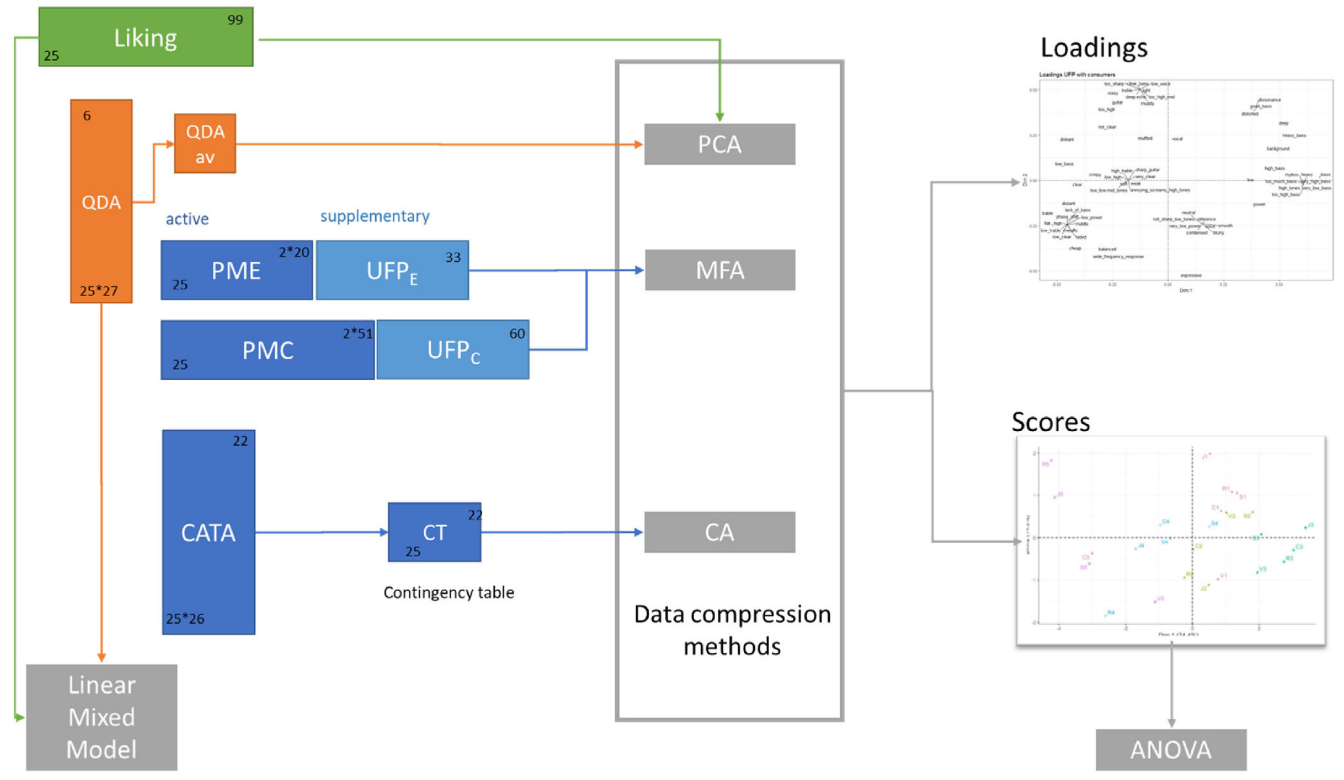

FIGURE 2 Overview of the statistical methods used in the study, from data set to results (plots). Number of rows and columns for each data set is indicated in lower left and upper right corner, respectively. For experts 209 words/phrases were identified after pre-processing, and 33 words were used five or more times. For consumers 60 words/phrases were identified for 51 consumers

no compensation for their time, recruitment was based on a convenience sample based on availability and the only reason for exclusion was self-reported hearing impairment. Assessors were asked to rate all loudspeakers using the 9-pt hedonic scale (ranging from 1 = Dislike extremely to 9 = Like extremely) for one track at a time (answering the question "How much do you like this loudspeaker?"). Consumers were allowed to go back and forth between loudspeakers as often as they wanted within a single track. When they had rated all five loudspeakers for one track, they moved on to the next track and the process was repeated for every track. Both the order in which consumers rated the tracks and the order of the loudspeaker within each track were fully randomized. Different sets of consumers participated in CATA, PMC, and the hedonic tests.

3 | STATISTICAL METHODS

An overview of the statistical methods used can be found in Figure 2.

3.1 | Preference mapping

In this paper we applied both internal and external preference mapping using each of the descriptive techniques as input, but we also investigated the option of using multi-block methods combining more than one descriptive method in order to improve results (see e.g., Måge, Menichelli, & Næs, 2012). In both cases the linear vector model, focusing on direction of preference, was used.

For all sensory methods, the goal is to obtain a set of scores for each sample in a low dimensional space that can be linked with the liking values. The data compression is important both for handling collinearity and for interpretation purposes. Methods for data compression were chosen dependent on the nature of the descriptive sensory data. The different sensory methods applied here (QDA, CATA, and PM) provide researchers with data of different characteristics and structure. This influences the choice of method for analyzing the data for the purpose of dimension reduction. It is well established to do dimension compression for QDA and liking with principal component analysis (PCA). For QDA, the panel averages were used for preference mapping, however PC-ANOVA (Luciano & Næs, 2009, see below) was conducted on both the panel average and the complete data set.

PM data were analyzed using multiple factor analysis (MFA, Escofier & Pagès, 1994), which is the most used method for analyzing PM data, although other methods have also been applied (Næs, Berget, Liland, Ares, & Varela, 2017; Tomic, Berget, & Næs, 2015) providing comparable results. In order to compare all 25 samples (five speakers and five tracks) MFA was conducted treating the different assessors as the different data blocks. Each data block then consisted of *X* and *Y* coordinates for all speakers and tracks. In a previous work on the same data (Pedersen et al., 2018), the PM data were analyzed using hierarchical MFA (hMFA, Le Dien & Pagès, 2003) because the loudspeakers were compared for one track at a time. In the present study, we have chosen to analyze all data in one MFA analysis with track*speaker as the rows in order to get similar sample spaces for all methods. Data from consumers (PMC) and experts (PME) were analyzed in the same manner.

Raw data from the CATA are binary in nature, and the common way of analyzing such data at the overall level is to do correspondence analysis (CA, Greenacre, 1984) on the contingency table. CA on contingency tables for binary or quantitative data is analoges to PCA on numerical data. Here, the contingency table was obtained for the 25 track*speaker combinations.

3.2 | Mixed model ANOVA

ANOVA was used for analyzing both descriptive and liking data. For the QDA data, loudspeaker and track were included as fixed effects and the assessors as random effects. All two-way interactions between fixed factors were included. The ANOVA model was fitted to single attributes, and for the two first PCs after PCA on the complete QDA data. This approach is sometimes referred to as PC-ANOVA (Luciano & Næs, 2009). The advantage of the approach is that it gives a direct assessment of importance of the design factors relative to what is seen along the PC axes.

A similar model was fitted to the first two components extracted from the data compression of QDA, CATA, PME, PMC, and the liking data, but without interaction between the fixed factors and the consumer/assessor effect since this is not possible without replication of each sample.

A model including consumer segments was also fitted to the liking data in order to study individual differences in more details. Consumers were split in two segments according to their loadings on the second component in the PCA of the liking data (internal mapping), segmentation based on PCA plots has been discussed in for instance Endrizzi, Gasperi, Rodbotten, and Naes (2014).

3.3 | Methods for comparing score spaces

The scores for the different descriptive sensory methods will be compared through visualization of the scores, the SMI (similarity index, Indahl, Næs, & Liland, 2018) and the RV coefficient (Robert & Escoufier, 1976). The SMI is a measure between 0 and 1 (or sometimes expressed as a percentage) where 1 (or 100) means perfect fit, and is simply defined as the explained variance of the standardized scores for one of the blocks predicted from the other. The SMI and RV are different in the sense that the RV depends on the relative size of the eigenvalues, while the SMI does not. The SMI is therefore sometimes more relevant than the RV coefficient for comparing matrices when the interest also lies in components beyond the first, which is where the RV places the most emphasis (Tomic et al., 2015). For even more detailed insight, the Pearson correlation coefficient was computed between each pair of scores from QDA and the other methods.

3.4 | Multi-block preference mapping based on SO-PLS

Multi-block regression refers to combining two or several data sets as independent variables in the regression. This can be done by concatenating the input blocks and using standard partial least squares (PLS), but other methods have also been developed (Westerhuis, Kourti, & MacGregor, 1998). A newer development is the SO-PLS regression (Næs et al., 2013; Romano, Tomic, Liland, Smilde, & Næs, 2019), which is a sequential procedure that focuses on the additional explained variance as new blocks are incorporated.

In the case of two blocks of explanatory variables (X_1 and X_2), one of the blocks X_1 is first fitted to Y by PLS regression. Then, the X_2 is orthogonalized with respect to X_1 and a new PLS regression is run. The scores from the two models are combined in a regression equation with Y as the dependent block (here consumer liking data). For the SO-PLS procedure we will here confine ourselves to using QDA as X_1 and then see if it is possible to increase the predictive ability of the consumer liking data by incorporating descriptive data obtained with consumers by adding either from CATA or PM data as X_2. An application of the method is presented by Niimi et al. (2018).

3.5 | Software

The preference mapping was done using R (R Core Team, 2019). The ANOVA models were fitted using lmer from lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmerTest (Kuznetsova, Brockhoff, & Christensen, 2017). The multivariate analyses (PCA, MFA, hMFA, and CA) were done using FactoMineR (Lê, Josse, & Husson, 2008). The RV and SMI were computed using MatrixCorrelations (Indahl et al., 2018). Visualizations of results were done using *emmeans* (Lenth, 2020) and *ggplot2* (Wickham, 2016). SO-PLS was performed in Matlab (Matlab, R2018b), using the code downloaded from www.nofimamodelling.org.

4 | RESULTS AND DISCUSSION

4.1 | Sensory description of loudspeakers obtained with QDA

The scores and loadings plot for QDA (using two components) are presented in Figure 3. The order of the speakers along PC1 (principal component 1) are 5–4–(2–1) and 3 when going from left to right (Figure 3a). The first component is quite dominating (accounting for

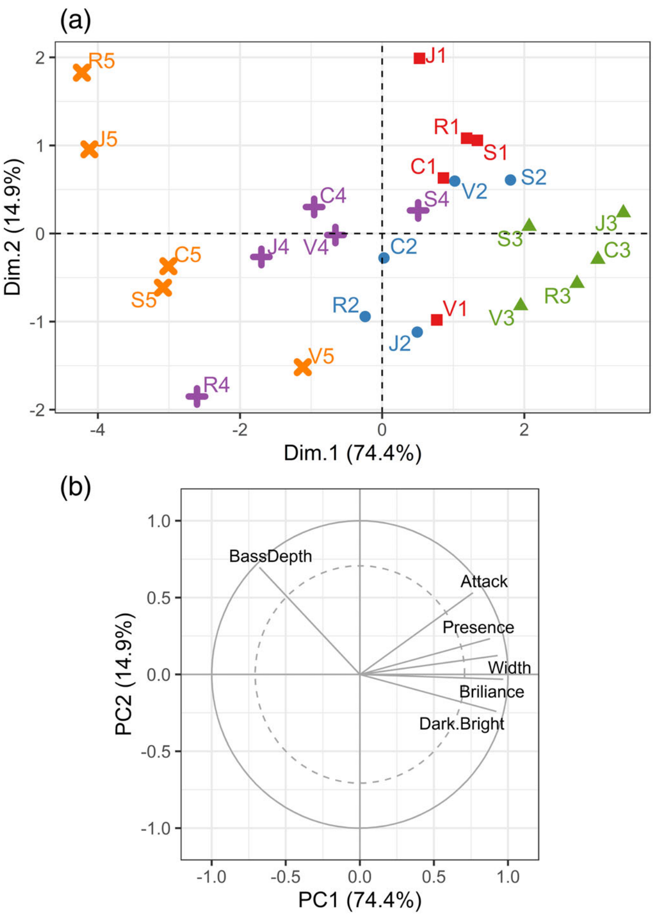

FIGURE 3 PCA of QDA data, scores (a) and correlation loadings (b). In the score plot labels indicate track (R = Rock, C = Classical Orchestra, S = String Quarted, J = Jazz, V = Vocal) and loudspeaker (numbers). Speakers are also marked by different symbols and colors

74% of the variance) and represents a contrast between bass depth and the other attributes (Figure 3b), showing that Speaker 3 scores low on bass depth and high on the other attributes. Speaker 1 is separated from Speaker 2 only in the PC2 (principal component 2) direction. This component may contrast the attribute Dark/Bright, the only attribute with a negative loading on PC2, to the rest. Speaker 5 has the highest variability between tracks with respect to PC2.

Separate ANOVA models for the two first principal components (based on the complete QDA data) show that speaker was clearly the dominating effect for the first component, but a significant effect was also detected for the track and the interaction ($p < .001$ all terms, Table 4). For Component 2, all fixed effects were found to be significant, but less than for the speaker along the first component. Interaction plots for both components are shown in Figure 4. The speaker effect is totally dominating along Component 1. Moreover, track S (string quartet) is the track mainly responsible for the interaction. There is also a quite clear effect of speaker along Component 2, but in this case the interaction is more pronounced. The interaction is mainly due to track V (vocal) and to a certain extent S (string quartet) which is represented by a quite flat line in the plot.

4.2 | Comparison of sensory methods

Score plots (two components) for PME, PMC, and CATA are shown in Figure 5a–c. Explained variances for the two first components of each method are summarized in Table 4, together with p-values from the described ANOVA models. As can be seen, the explained variances are lower than for QDA (Figure 3). This is quite natural given the nature of the data acquisition and also the fact that QDA is based on replicated data.

When comparing the scores for the different methods, the effect along the first component is very similar to that of QDA, with Speakers 3 and 5 spanning the first component and the rest in the middle. The QDA (Figure 3) gives the best separation of the samples. PC-ANOVA indicated a clear significant effect of speaker on the first dimensions for PME, PMC, and CATA (p-values summarized in Table 4). Moreover, for all methods the speaker effect was comparable to the estimated effect with QDA, although discrimination was better with QDA. Speaker 3 scored highest on this component for all methods. With CATA and PMC the track also had a significant effect on the first dimension, but when going to pairwise comparisons the methods would provide some different results. The order of tracks is **V**-**S**-C-**J**-**R** for CATA, and **V**-S-R-J-**C** for PMC (groups that are significantly different are highlighted).

Relations along the second component are generally less clear visually. According to PC-ANOVA, effect of track was significant for the second dimension with PMC, but only nearly significant effect with PME. However, when comparing the effects, the ranking of tracks was different. With PMC, the tracks were divided into two groups V-J (high scores) versus S-C-R (low scores), whereas the pattern was less clear for PME.

Further comparisons on methods can be supported by RV, SMI and Pearson correlations between components (Table 5). Based on RV and SMI it is clear that PME has the highest similarity with QDA along the first axis, followed by CATA. The PMC is quite different from QDA. When looking into two components, RV coefficients indicate best correlation between PME and QDA, whereas SMI indicates higher agreement between QDA and CATA. Since RV is dominated by the component with largest eigenvalues (see Indahl et al., 2018), SMI can be considered a better tool for assessing similarity in this case. The pairwise correlations in Table 5 supports that CATA and QDA are most alike, whereas similarity between PM and QDA is lower. Similarity between QDA and the other methods was even more pronounced when comparing the configurations from the scores averaged over the tracks. In this case the RV (SMI) values for the two component solutions were 0.98 (0.85) for PME, 0.91 (0.80) for PMC, and 0.90 (0.72) for CATA.

When comparing scores between the rapid sensory methods (CATA, PME, PMC), neither RV nor SMI are particularly high. For the comparison with two components, the values were (SMI = 0.29, RV = 0.42) for PME/PMC (SMI = 0.34, RV = 0.40) for PMC/CATA and (SMI = 0.39, RV = 0.53) for PME/CATA. RV coefficients are noticeably larger than the corresponding SMI, since with RV it is mainly the first component that is emphasized.

TABLE 4 Overview of ANOVA results for different data sets

Data	Data reduction method	Comp1% variance	*p*-value speaker	*p*-value track	*p*-value interaction (s)	Comp2% variance	*p*-value speaker	*p*-value track	*p*-value interaction
QDA complete[a]	PCA	74.4	<.001	<.001	<.001	14.9	<.001	<.001	<.001
QDA average[b]	PCA	74.0	<.001	.34		15.0	.43	.76	
PME[b]	MFA	27.9	<.001	.85	–	8.9	.17	.06	–
PMC[b]	MFA	12.0	<.0001	.002	–	9.9	.003	<.001	–
CATA[b]	CA	50.9	<.001	.006	–	13.7	.44	.56	–
Liking[b]	PCA	13.6	.004	.04	–	9.2	.83	.05	–
Liking[c]	Values	–	<.001	<.001	<.001	–	–	–	–
Liking[d]	Values	–	<.001	.70	.04				

Note: Columns show the data set, method used for data reduction, explained variance for component 1, *p*-values for speaker, track and interaction(s) (when relevant), then the same for the second component. Note that there are three different models for liking data, defined by superscripts b–d. For the model with consumer segments (model 4), segment was included as a fixed factor and consumer(segment) as a random factor. The *p*-value for the interaction is for track*speaker. The other interactions had *p*-values <.001. The *p*-value for the segment effect was .55.

[a]ANOVA model: Score = Track + Speaker + Track*Speaker + Assessor + Speaker*Assessor + Track*Speaker + Track*Speaker*Assessor.

[b]ANOVA model: Score = Track + Speaker.

[c]ANOVA model: Liking = Track + Speaker + Track*Speaker + Consumer.

[d]ANOVA model: Liking = Track + Speaker + Track*Speaker + Segment + Track*Segment + Speaker*Segment + Track*Speaker*Segment + Consumer (segment).

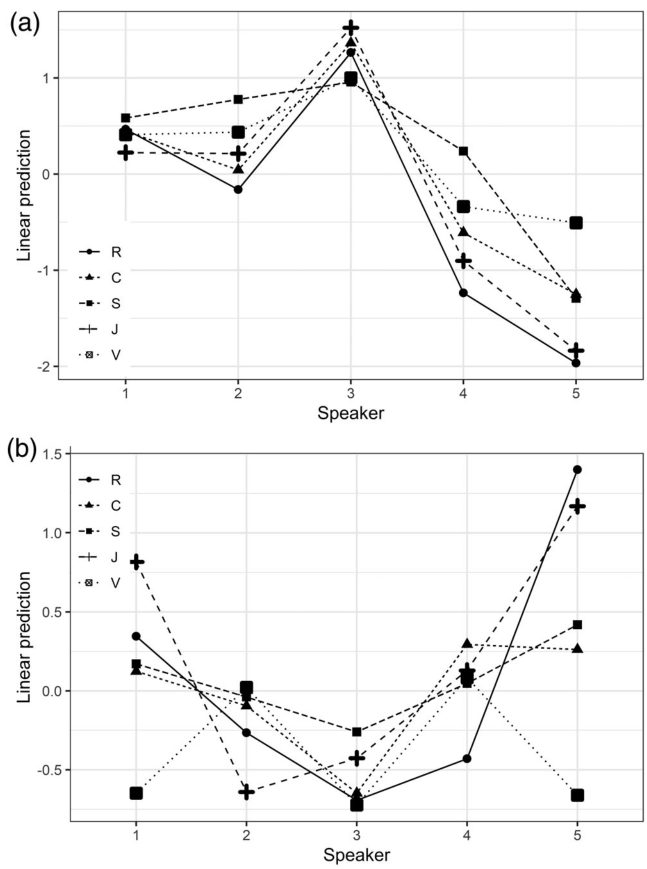

FIGURE 4 QDA. Interaction plots for PC1 and PC2. *SE* = 0.30 and for PC1, and 0.25 for PC2. The different speakers (1–5) are indicated along the *x*-axis whereas the tracks are indicated by different lines and symbols

For interpretation of the components from the rapid methods, words from the UFP data are projected to the MFA components for PME and PMC (Figure 6a,b), whereas for CATA we can consider the loadings from CA on the contingency tables (Figure 6c). Comparing these plots with the loading plot from QDA (Figure 3b) it is evident verbal descriptors in sensory space are in general consistent across methods. Additionally, more diverse or comprehensive descriptions of the loudspeakers can be obtained with these methods although the data become less structured as the variability explained by dimensions 1 and 2 is lower than for the QDA.

The free elicitation of words by experts in particular was very comprehensive, and resulted in 209 words after the simplification by sound experts showing a large degree of individual variation in the descriptions. However, only 33 of the words were used five or more times and only these are included in Figure 6a. For the UFP with consumers, 60 words were generated in total, all were used five or more times and were included in the plot (Figure 6b). As the number of assessors in PME was much lower than for PMC, the percentages kept is not the same for the two data sets, however, when relaxing the filtering for UFP with experts, the plot in Figure 6b became overloaded and difficult to read. Therefore, it was decided to use the five occurrences for both data sets, although this means that different criteria were used for experts and consumers in terms of percentages.

4.3 | Internal preference mapping

In internal preference mapping, the consumer preferences (liking) are used as input data, and the data are compressed by PCA, before the sensory attributes are linked to the scores through. The PCA for the liking data is presented in Figure 7.

The scores (Figure 7a) show a slightly different pattern than for QDA (Figure 3a). Compared to QDA the samples are less separated as most of the samples are crowded around the center, although some

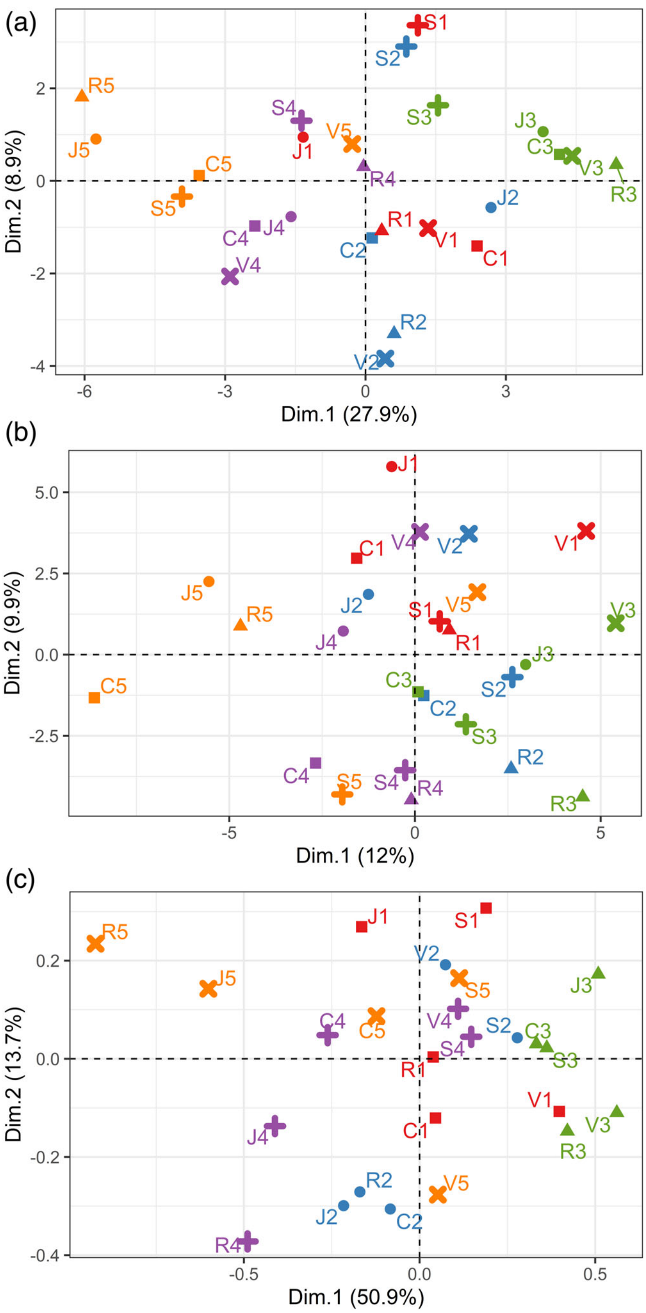

FIGURE 5 Score plot for rapid sensory methods (a) MFA of PME data, (b) MFA of PMC data, (c) scores from correspondence analysis (CA) of the contingency table from CATA. In the score plot labels indicate track (Rock, C = Classical Orchestra, S = String Quarted, J = Jazz, V = Vocal) and loudspeaker (numbers). Speakers are also marked by different symbols and colors

track and speaker combinations stand out. Most evident is two tracks (J and R) for Speaker 3 with high positive scores on PC1. In addition, three tracks (J, V, and R) for Speaker 5 have high positive scores on PC2. The order of the loudspeakers along PC1 (with the exception of Speaker 5) resembles, however, that from QDA (Figure 3a).

The consumer loadings in Figure 7b, are mainly on the left-hand side, where bass depth is located. This shows that they have rated the liking of samples (speaker and track combinations) on the left side in the corresponding score plot (Figure 7a) higher than samples on the right side. The consumer loadings therefore tell which area of the sensory space that are most preferred. All tracks with Speaker 4 are on this side, so Speaker 4 is the most liked, whereas Speaker 3 is the least liked. Since the first axis distinguishes primarily between bass depth and the rest (see Figure 3), this attribute is considered the most favorable property and is likely to be a driver of liking for loudspeakers. For Component 2, there are consumers in both the positive and negative direction of the score (Figure 7b). In order to investigate these individual differences further, consumers were separated into two segments based on the second component as indicated in Figure 7b. These segments were later on analyzed by ANOVA and external preference mapping as will be shown below. Segment 1 is located on the positive side of PC2, whereas Segment 2 comprises consumers on the negative side. The separation line is set at the median of the loadings on PC2 to obtained balanced (equally sized) groups. The second component reflects differences related to Dark. Bright (positive side) and BassDepth, Presence and Width (negative side). When looking into other components (PC3 and PC4) which explain less of the variation, no interesting patterns could be seen.

Sensory loadings for internal preference mapping (Figure 7b) and QDA (Figure 3b) show similar patterns, an exception is the attribute attack which has moved in the plot when going from internal to external mapping. Note that the second axis is turned upside down in the internal mapping plot, but this is arbitrary and has no influence on the conclusion.

The average liking for the different speakers were 4.8 (3), 5.2 (5), 5.3 (1), 5.3 (2), and 5.5 (4), the overall *SE* was 0.09. The overall average for tracks varied between 4.9 (Rock) and 5.5 (String Quarted). The differences in liking between the best liked and the least liked were relatively small taken into account that a 9-point scale was used.

The ANOVA for the liking data confirmed that Speaker 3 was significantly less liked than the other speakers ($p < .001$), whereas pairwise comparisons did not prove significant differences between

TABLE 5 Comparison between QDA and rapid sensory methods for A components

	RV			SMI			Correlation per component		
A	PME	PMC	CATA	PME	PMC	CATA	PME	PMC	CATA
1	0.79	0.55	0.67	0.79	0.55	0.67	0.89	0.74	0.82
2	0.76	0.45	0.68	0.49	0.40	0.66	0.26	0.40	0.78
3	0.74	0.40	0.67	0.45	0.28	0.51	0.18	−0.03	0.34
4	0.73	0.44	0.67	0.43	0.29	0.48	0.24	−0.14	−0.02

Note: PME, projective mapping with experts; PMC, projective mapping with consumers. For each data set similarity to product map is measured by RV and SMI with A components, and the correlation between components from QDA and the other sensory methods.

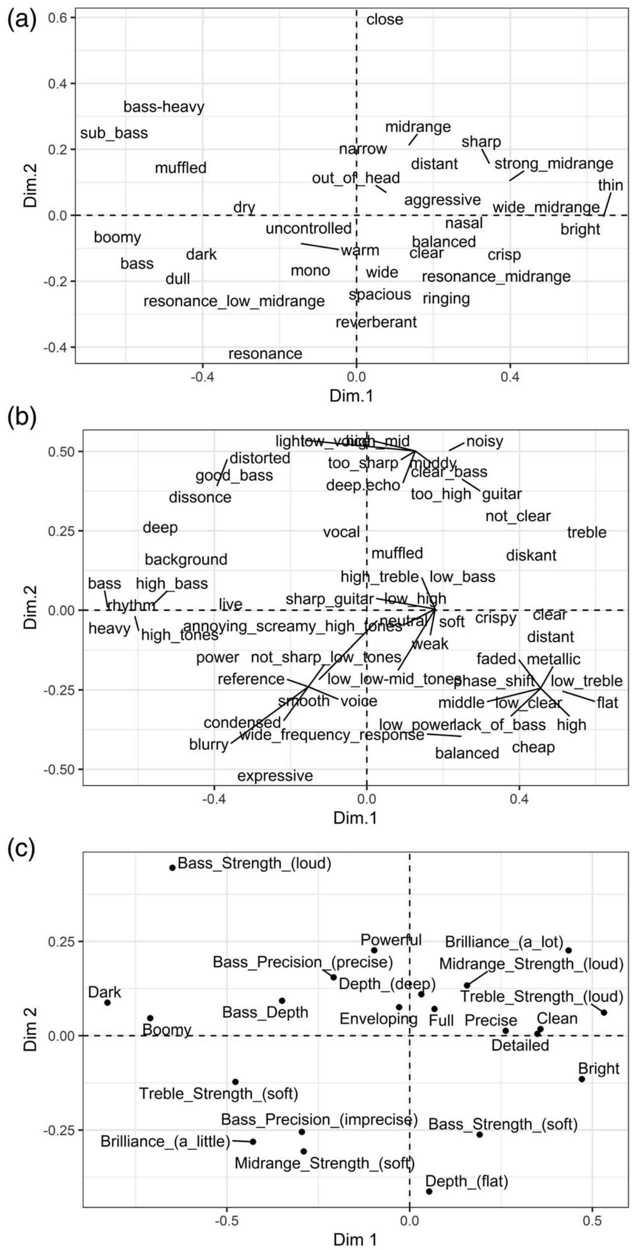

FIGURE 6 (a) Words from UFP projected to the MFA components for PME. (b) Words from UFP projected to the MFA components for PMC. (c) Loading plot from correspondence analysis (CA) of the contingency table from CATA

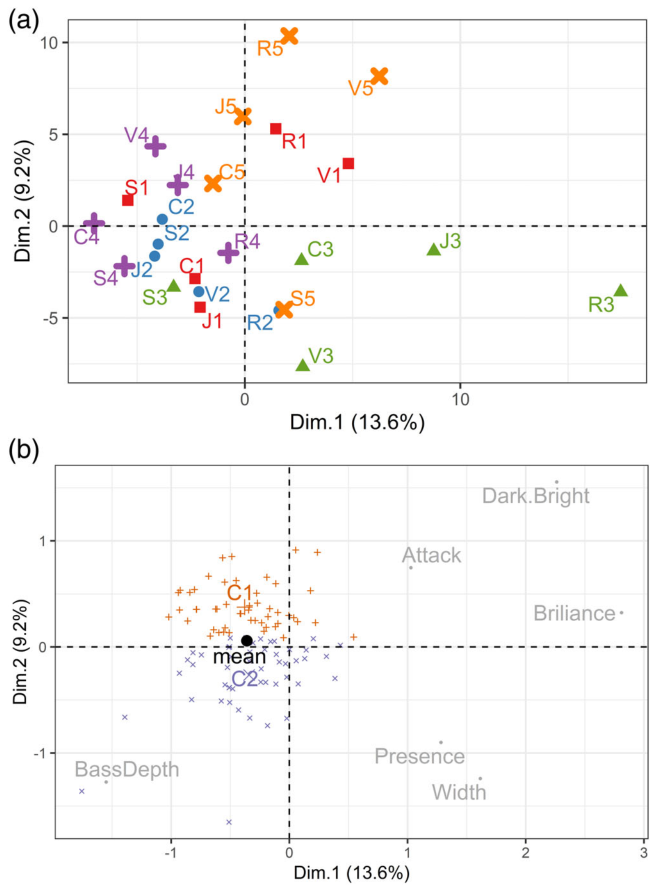

FIGURE 7 Internal preference mapping. (a) Scores from the PCA on the liking data. The first letter reflects the track (R = Rock, C=Classical Orchestra, S = String Quartet, J = Jazz, V = Vocal) whereas speakers are noted L1–L5 and are also differentiated by colors. (b) Consumer loadings and sensory loadings after regressing QDA data onto the principal components of the liking. Consumers are segmented according to loadings on PC2 (above/below median). Projections of mean values for the complete data set and the two segments are shown

the other speakers (p-values for the comparison larger than .05). When modeling the average there was a significant effect of track, as well as a significant interaction between track and speaker (Table 4). However, when including the segments defined in Figure 7b in the ANOVA, the main effect of track was not significant, whereas all interactions (both two- and three-ways) were significant. Figure 8 shows the interaction plot for track and speaker for each segment. For Segment 1 (which have high loadings on PC2, Figure 7b) there is a strong interactions between speaker and track, and in particular for the rock track they rate the speakers very differently. When evaluating Speaker 3, the rock track obtains the poorest average rating, whereas the same track gets the best rating for Speaker 5 which is the speaker with the strongest bass effect. The other segment shows large variations between track for Speakers 3, 5, and 1, but only small differences for Speakers 2 and 4, and have less variation which track is liked best. Overall Speaker 5 scored relatively more on track R (Rock) compared to other speakers. The best-liked speaker (4) performed well on all tracks.

4.4 | External preference mapping—Comparison across sensory methods

Consumer loading plots for external preference mapping based on each of the sensory methods described, are shown in Figure 9. For all methods, the average liking goes in the same direction, indicating that there was a tendency towards preference for loudspeakers with the strongest bass depth (speaker numbers 4 and 5). Bass depth is a driver

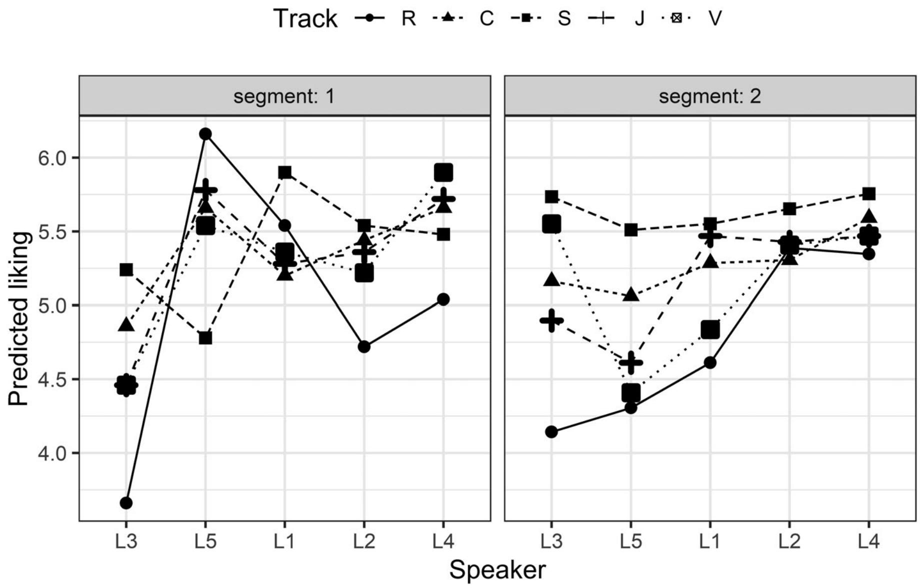

FIGURE 8 Interaction between speaker and track for liking for each segment. Loudspeakers on the horizontal axis are ordered from least (left side) to best liked (right side). The tracks are indicated by different lines and symbols (R = Rock, C=Classical Orchestra, S = String Quartet, J = Jazz, V = Vocal)

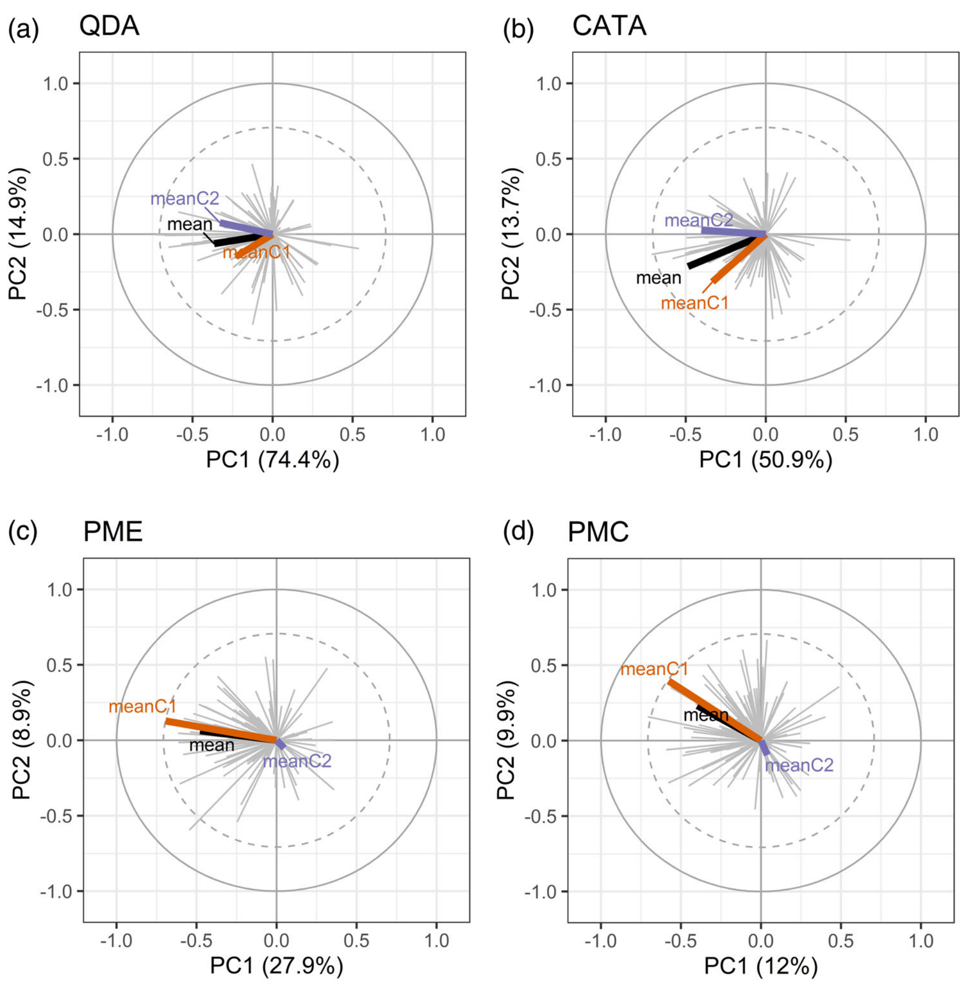

FIGURE 9 Consumer loadings from external preference mapping with different sensory methods (a) QDA, (b) CATA, (c) PME, and (d) PMC. In each plot the mean for all consumers and each of the two segments are shown

of liking when comparing with loadings from QDA (Figure 3b). This corresponds well with words on the left side from PME, PMC, and CATA attributes (Figure 6).

Averages of the two segments identified from the internal preference mapping (Figure 7b), have arrows in the same direction as the overall mean for the external mapping with QDA and CATA, and only

TABLE 6 Summary statistics for R^2 for the relation between the different descriptive methods and liking (external preference mapping)

	QDA	CATA	PME	PMC
Min	<0.001	0.002	0.003	0.003
First quartile	0.022	0.030	0.033	0.037
Median	0.072	0.085	0.070	0.081
Mean	0.099	0.103	0.098	0.109
Third quartile	0.127	0.140	0.138	0.151
Max	0.537	0.521	0.446	0.470

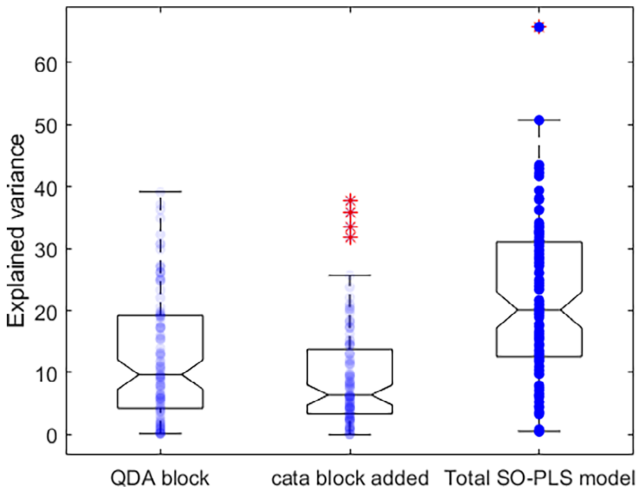

FIGURE 10 Boxplot for explained variance in SO-PLS model with one block (QDA), the additional explained variance from CATA (CATA block) and the total explained variance in liking from the complete model

a small angle between them. For the preference mapping with PM (both experts and consumers) on the other hand, the arrow corresponding to the average for the second segment, points opposite to that of Segment 1 (and the overall mean), but is very short. Hence, external mapping using PM does not reveal direction of liking for Segment 2. This indicates that consumers in Segment 2 have a weaker preference pattern than the rest. Segment 1 has similar directions as the overall mean.

In external preference mapping one regression model for each consumer is obtained. The regression coefficients are often referred to as consumer loadings. In order to investigate whether any of the sensory methods provides a better relation between reported liking and sensory properties of the products, the descriptive statistics for the R^2 of the regression models were investigated. The descriptive statistics are shown in Table 6. The average is around 0.1 for all methods. This range of R^2 is quite common when looking at individual consumers in preference mapping, nevertheless the analyses usually provide information at the population level. The third quantile is slightly higher for PMC than for the others, including QDA (0.151 vs. 0.127). Based on numbers in Table 6, there is no clear evidence that any of the sensory methods give better or poorer relation between sensory descriptions and consumer liking. This is consistent with results obtained in the food-related sensory literature (Adams et al., 2007; Ares et al., 2011; Dehlholm et al., 2012). In the present study the main liking tendency was mostly related to the first component, for other situations where the relation between sensory and liking is more complex, the differences between methods may possibly be more expressed.

4.5 | Combining sensory methods for external preference mapping

As was seen above (Table 6) it is hard to find good relations between consumer liking based and sensory data. This is partially because consumers tend to give noisy answers in surveys, but may also reflect that the sensory data may not capture the details that consumers respond to when reporting liking. In particular, trained panels could provide descriptions which may not necessarily reflect consumers perception; they do not reflect individual differences in sensory perception in a wide consumer sample, and do not represent different points of view of what are the attributes more important for hedonic perception to different consumers (Ares & Varela, 2017). It is therefore of interest to see if different sensory methods can enhance relations between descriptive and liking data when combined. In order to check this, SO-PLS was applied with scores from the two first components from QDA as the first X-block and then scores from the two first factors of either CATA or PMC as the second block. The liking data represent the Y-block. The explained variances for all consumers are plotted in Figure 10 using box-plots. As can be seen, the QDA and CATA are comparable while the combined plot indicates slightly better predictions. This result is, however, not supported by cross-validation and is probably mainly due to the fact that the number of independent variables is increased. The conclusion is that in the present study combining the two does not improve relations between descriptive and hedonic data in any substantial way. Similar results were obtained for PME (not shown).

4.6 | Limitations and future research

Possible limitations to acknowledge, include the number of assessors used in the CATA and the two PM experiments, which were lower than what the literature would suggest. Specifically, Ares and collaborators, using bootstrap resampling, suggested that 60–80 assessors are necessary to obtain stable product profiles using CATA (Ares, Tárrega, Izquierdo, & Jaeger, 2014), and at least 50 (naïve) assessors using PM (Vidal et al., 2014). While the direct relevance of such recommendations to the present work is uncertain—as they are based on sensory evaluation of food—it is possible that configurational similarities between the different methods would increase with the inclusion of additional subjects and/or by using replicate evaluations with the same subjects. In general, determining the optimal number of assessors for evaluation of perceived sound quality using rapid sensory methods seems a useful continuation of the present research.

Another factor that could influence the inferences resulting from comparing the different experiments is the application of static or dynamic binaural rendering as reproduction method (Moulin et al., 2016). Dynamic rendering was applied for the QDA, CATA, and PM using trained assessors' experiments and static rendering for the other experiments. The dynamic rendering was applied in the initial experiments to ensure perceptual preservation of the spatial properties of the two-channel (stereo) reproduction. However, only two (Presence and Width) of the six attributes elicited in the QDA experiment could be argued to be related to spatial properties while the other four were representing timbral aspects. This is in accordance with results (Rumsey, Zieliński, Kassier, & Bech, 2005), showing that approximately 70% of an overall sound quality rating is due to timbral aspects and the remaining 30% is accounted for by spatial properties. Furthermore, *Width* is related to the perceived width of the sound scape (stereo image) and this should be preserved also in a static binaural rendering situation, and it is unclear if perception of Presence is directly related to spatial aspects according to Sound Wheel categorization (ITU-R, 2017). The loading plots (see Figures 1, 3b, and 6 confirm that the majority of the attributes are related to timbral aspects of the reproduction. The influence of static or dynamic rendering is thus believed to have a negligible influence on the comparisons between the tested sensory methods.

The results indicate that the effort and time consuming QDA method could be substituted by either CATA using naïve subjects or PM using trained listeners. The implications of this would be important for the large number of sensory tests, for example, bench marks and in development projects that are performed on a daily basis in the audio industry. However, a QDA profile with few attributes was used in this study. Hence, further experiment employing a larger number of subjects needs to be performed to consolidate the conclusions of the present set of experiments.

5 | CONCLUSIONS

In the present study, different methods were compared to Quantitative Descriptive Analysis (QDA) to obtain descriptive sensory profiles of a set of high-end loudspeakers. The ability of the different methods to predict consumer liking was investigated. The results indicate that the different sensory methods provided comparable results, although the direction of the most preferred product was less evident with projective mapping with consumers compared to the other methods. With QDA and CATA, the drivers of liking can be identified through the attributes and selected CATA terms, whereas projective mapping should be accompanied with UFP to obtain an interpretation of the components and find potential drivers of liking. All methods indicated that bass depth opposed is important for perception and reported liking of the loudspeakers. The main conclusion is that it seems that little is lost if QDA is substituted by CATA or PM at least when using experts for the latter.

ACKNOWLEDGMENTS

We would like to thank for the financial support received the Norwegian Foundation for Research Levy on Agricultural Products FFL, through the research program "FoodSMaCK", Spectroscopy, Modeling and Consumer Knowledgeort (grant number NFR 262308). Students at the University of Southern Denmark who helped carrying out the PMC experiment and the hedonic test are also thankfully acknowledged, so as the different institutions, experimenters and participants involved in the data collection.

ORCID

Ingunn Berget https://orcid.org/0000-0003-1027-1472
Davide Giacalone https://orcid.org/0000-0003-2498-0632

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, Minneapolis, USA.

Ares, G., Tárrega, A., Izquierdo, L., & Jaeger, S. R. (2014). Investigation of the number of consumers necessary to obtain stable sample and descriptor configurations from check-all-that-apply (CATA) questions. *Food Quality and Preference*, *31*, 135–141. https://doi.org/10.1016/j.foodqual.2013.08.012

Ares, G., & Varela, P. (2017). Trained vs. consumer panels for analytical testing: Fueling a long lasting debate in the field. *Food Quality and Preference*, *61*, 79–86. https://doi.org/10.1016/j.foodqual.2016.10.006

Ares, G., Varela, P., Rado, G., & Giménez, A. (2011). Identifying ideal products using three different consumer profiling methodologies. Comparison with external preference mapping. *Food Quality and Preference*, *22*, 581–591. https://doi.org/10.1016/j.foodqual.2011.04.004

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *1*(1), 1–48.

Bech, S., Ellermeier, W., Ghani, J., Gulbol, M. -A., & Martin, G. (2005). *A listening test system for automotive audio - Part 2: Initial verification*. Audio Engineering Society Convention 118, Barcelona, Spain.

Bech, S., & Zacharov, N. (2006). *Perceptual audio evaluation–theory, method and application*. Chichester, England: John Wiley and Sons.

Christensen, F., Lydolf, M., Martin, G., Minnaar, P., Pedersen, B., & Song, W. -K. (2005). *A listening test system for automotive audio - Part 1: System description*. Audio Engineering Society Convention 118, Barcelona, Spain.

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 and Preference*, *26*, 267–277. https://doi.org/10.1016/j.foodqual.2012.02.012

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 and Preference*, *21*, 394–401. https://doi.org/10.1016/j.foodqual.2009.10.002

Endrizzi, I., Gasperi, F., Rodbotten, M., & Naes, T. (2014). Interpretation, validation and segmentation of preference mapping models. *Food Quality and Preference*, *32*, 198–209. https://doi.org/10.1016/j.foodqual.2013.10.002

Escofier, B., & Pagès, J. (1994). Multiple factor analysis (AFMULT package). *Computational Statistics & Data Analysis*, *18*, 121–140. https://doi.org/10.1016/0167-9473(94)90135-X

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 and Preference*, *17*, 604–614. https://doi.org/10.1016/j.foodqual.2006.05.006

Giacalone, D., Bredie, W. L. P., & Frøst, M. B. (2013). "All-in-one test" (AI1): A rapid and easily applicable approach to consumer product testing. *Food Quality and Preference*, *27*, 108–119. https://doi.org/10.1016/j.foodqual.2012.09.011

Giacalone, D., Nitkiewicz, M., Moulin, S., Boðason, T., Lund Laugesen, J., & Bech, S. (2017). *Sensory Profiling of High-End Loudspeakers Using Rapid Methods—Part 2: Projective Mapping with Expert and Naïve Assessors*, In Proceedings of the 142nd Audio Engineering Society Convention, Paper 9975, Berlin, Germany, 20–23 May 2017.

Greenacre, M. (1984). *Theory and applications of correspondence analysis*. London, UK: Academic Press.

Hegarty, P., Choisel, S., & Bech, S. (2007). *A listening test system for automotive audio - Part 3: Comparison of attribute ratings made in a vehicle with those made using an auralization system*. Audio Engineering Society Convention 123, New York, USA.

Hicks, L., Moulin, S., & Bech, S. (2018). Sensory profiling of high-end loudspeakers using rapid methods - Part 3. Check-all-that-apply with naïve assessors. *The Journal of the Audio Engineering Society*, *66*, 329–342. https://doi.org/10.17743/jaes.2018.0015

IEC TR 60268–13. (1998). *Sound system equipment - Part 13: Listening tests on loudspeaker*. https://webstore.iec.ch/publication/1213.

Indahl, U. G., Næs, T., & Liland, K. H. (2018). A similarity index for comparing coupled matrices. *Journal of Chemometrics*, *32*, e3049. https://doi.org/10.1002/cem.3049

ISO-7029. (2017). *Acoustics — Statistical distribution of hearing thresholds related to age and gender*. Geneva, Switzerland: International Organization for Standardization.

ITU-R (2017). BS 2399: Methods for selecting and describing attributes and terms, in the preparation of subjective tests. In *Report, BS Series*. Geneva, Switzerland: International Telecommunication Union Radiocommunication Assembly (ITU-R).

Jaeger, S. R., Wakeling, I. N., & MacFie, H. J. H. (2000). Behavioural extensions to preference mapping: The role of synthesis. *Food Quality and Preference*, *11*, 349–359. https://doi.org/10.1016/S0950-3293(00)00009-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*, 26. https://doi.org/10.18637/jss.v082.i13

Le Dien, S., & Pagès, J. (2003). Hierarchical multiple factor analysis: Application to the comparison of sensory profiles. *Food Quality and Preference*, *14*, 397–403. https://doi.org/10.1016/S0950-3293(03)00027-2

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*, 18. https://doi.org/10.18637/jss.v025.i01

Lenth, R. (2020). emmeans: Estimated marginal means, aka least-squares means. R package version 1.4.6.

Luciano, G., & Næs, T. (2009). Interpreting sensory data by combining principal component analysis and analysis of variance. *Food Quality and Preference*, *20*, 167–175. https://doi.org/10.1016/j.foodqual.2008.08.003

Matlab. (R 2018b). Matlab. Mathworks, Inc.

Mattila, V. V. (2001). *Descriptive analysis of speech quality in mobile communications: Descriptive language development and external preference mapping*. Audio Engineering Society Convention 111. Audio Engineering Society

Meullenet, J.-F., Xiong, R., & Findlay, C. J. (2007). *Multivariate and probabilistic analyses of sensory science problems*. Ames: Blackwell Publishing.

Moulin, S., Bech, S., & Stegenborg-Andersen, T. (2016). *Sensory profiling of high-end loudspeakers using rapid methods-Part 1: Baseline experiment using headphone reproduction*. AES International Conference on Headphone Technology, Ålborg, Denmark, 24-26 August.

Måge, I., Menichelli, E., & Næs, T. (2012). Preference mapping by PO-PLS: Separating common and unique information in several data blocks. *Food Quality and Preference*, *24*, 8–16. https://doi.org/10.1016/j.foodqual.2011.08.003

Niimi, J., Tomic, O., Næs, T., Jeffery, D. W., Bastian, S. E. P., & Boss, P. K. (2018). Application of sequential and orthogonalised-partial least squares (SO-PLS) regression to predict sensory properties of Cabernet Sauvignon wines from grape chemical composition. *Food Chemistry*, *256*, 195–202. https://doi.org/10.1016/j.foodchem.2018.02.120

Næs, T., Berget, I., Liland, K. H., Ares, G., & Varela, P. (2017). Estimating and interpreting more than two consensus components in projective mapping: INDSCAL vs. multiple factor analysis (MFA). *Food Quality and Preference*, *58*, 45–60. https://doi.org/10.1016/j.foodqual.2016.11.012

Næs, T., Tomic, O., Afseth, N. K., Segtnan, V., & Måge, I. (2013). Multiblock regression based on combinations of orthogonalisation, PLS-regression and canonical correlation analysis. *Chemometrics and Intelligent Laboratory Systems*, *124*, 32–42. https://doi.org/10.1016/j.chemolab.2013.03.006

Næs, T., Varela, P., & Berget, I. (2018). *Individual differences in sensory and consumer science: Experimentation, analysis and interpretation*, Chippenham, England: Woodhead Publishing.

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 and Preference*, *16*, 642–649. https://doi.org/10.1016/j.foodqual.2005.01.006

Pagès, J. (2003). Recueil direct de distances sensorielles: application à l'évaluation de dix vins blancs du Val-de-Loire. *Sciences des Aliments*, *23*, 679–688. https://doi.org/10.3166/sda.23.679-688

Pedersen, M. E., Moulin, S., Berget, I., Varela, P., Næs, T., & Beck, S. (2018). *Sensory profiling of loudspeakers using rapid methods compared to traditional QDA*, Eurosense, Eight European Conference on Sensory and Consumer Research Verona, Italy.

Pedersen, T. H., & Zacharov, N. (2015). *The development of a sound wheel for reproduced sound*. Audio Engineering Society Convention 138, Warsaw, Poland, 6 May.

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*, 372–395. https://doi.org/10.1111/j.1745-459X.2009.00216.x

R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from https://www.R-project.org/

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 and Preference*, *32*, 160–166. https://doi.org/10.1016/j.foodqual.2013.02.004

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 and Preference*, *5*, 263–269. https://doi.org/10.1016/0950-3293(94)90051-5

Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: The RV-coefficient. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, *25*, 257–265. https://doi.org/10.2307/2347233

Romano, R., Tomic, O., Liland, K. H., Smilde, A., & Næs, T. (2019). A comparison of two PLS-based approaches to structural equation modeling. *Journal of Chemometrics*, *33*, e3105. https://doi.org/10.1002/cem.3105

Rumsey, F., Zieliński, S., Kassier, R., & Bech, S. (2005). On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality. *The Journal of the Acoustical Society of America*, *118*, 968–976. https://doi.org/10.1121/1.1945368

Stone, H., Sidel, J., Woolsey, A., & Singleton, R. C. (2008). Sensory evaluation by quantitative descriptive analysis. In *Descriptive sensory analysis in practice* (Vol. 28, pp. 23–34). Trumbull, Connecticut: Wiley.

Tomic, O., Berget, I., & Næs, T. (2015). A comparison of generalised procrustes analysis and multiple factor analysis for projective mapping data. *Food Quality and Preference*, *43*, 34–46. https://doi.org/10.1016/j.foodqual.2015.02.004

Torri, L., Dinnella, C., Recchia, A., Naes, T., Tuorila, H., & Monteleone, E. (2013). Projective mapping for interpreting wine aroma differences as perceived by naïve and experienced assessors. *Food Quality and Preference*, *29*, 6–15. https://doi.org/10.1016/j.foodqual.2013.01.006

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. https://doi.org/10.1016/j.foodres.2012.06.037

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. https://doi.org/10.1016/j.foodqual.2013.03.004

Varela, P., Berget, I., Hersleth, M., Carlehög, M., Asioli, D., & Næs, T. (2017). Projective mapping based on choice or preference: An affective approach to projective mapping. *Food Research International*, *100*, 241–251. https://doi.org/10.1016/j.foodres.2017.08.049

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 and Preference*, *34*, 79–87. https://doi.org/10.1016/j.foodqual.2013.12.006

Westerhuis, J. A., Kourti, T., & MacGregor, J. F. (1998). Analysis of multiblock and hierarchical PCA and PLS models. *Journal of Chemometrics*, *12*, 301–321. https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5%3C301::AID-CEM515%3E3.0.CO;2-S

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag Retrieved from https://ggplot2.tidyverse.org

Withers, C., Methven, L., Qannari, E. M., Allen, V. J., Gosney, M. A., & MacFie, H. J. H. (2014). Taxonomic free sorting: A successful method with older consumers and a novel approach to preference mapping. *Journal of Sensory Studies*, *29*, 182–189. https://doi.org/10.1111/joss.12093

Zacharov, N., & Koivuniemi, K. (2001). *Unravelling the perception of spatial sound reproduction: Analysis & external preference mapping*. Audio Engineering Society Convention 111. Audio Engineering Society.

How to cite this article: Berget I, Bech S, Giacalone D, et al. Sound quality perception of loudspeakers evaluated by different sensory descriptive methods and preference mapping. *J Sens Stud*. 2020;e12620. https://doi.org/10.1111/joss.12620