Audio Engineering Society

Convention Paper 9775

Presented at the 142nd Convention
2017 May 20–23, Berlin, Germany

Sensory profiling of high-end loudspeakers using rapid methods - Part 2: Projective mapping with expert and naïve assessors

Davide Giacalone[1], Maciej Nitkiewicz[1], Samuel Moulin[2], Torstein Boðason[3], Jakob Lund Laugesen[4], and Søren Bech[2,3]

[1] *University of Southern Denmark, Odense, Denmark*
[2] *Bang & Olufsen A/S, Struer, Denmark*
[3] *Aalborg University, Aalborg, Denmark*
[4] *University of Copenhagen, Frederiksberg, Denmark*

Correspondence should be addressed to Davide Giacalone (dg@iti.sdu.dk)

ABSTRACT

This is the second of a series of papers evaluating the efficiency of rapid sensory profiling methodologies in the audio field [1]. The present paper introduces projective mapping [2] as a method for perceptual audio evaluation, and demonstrates its application for discrimination and description of a set of high-end loudspeakers. Additionally, the suitability of the method with both experts and naïve assessors was evaluated. The results showed a successful discrimination between the loudspeakers with the main differences primarily associated to bass strength and bass depth. A high degree of agreement was observed between perceptual configurations obtained separately by the expert and the naïve assessors, though the former outperformed the latter in the descriptive part of the method.

1 Introduction

1.1 Rapid methods for perceptual audio evaluation

Sensory analysis is a powerful tool aiming at providing a complete description of the sensory characteristics of consumer products. In the field of audio quality, listening tests are the main tools by which sensory data on audio products are acquired. Although this may be limited to a single subjective metric (using concepts such as the Basic Audio Quality (BAQ) or the Mean Opinion Score (MOS)) there are many instances where the experimenter may wish to characterize perceptually the stimuli under test on a number of individual sensory attributes [3, 4]. To this end, methods such as Descriptive Analysis (DA) are usually applied [5]. In conventional DA, a group of trained assessors (usually 8-10) develops a common vocabulary for describing a set of products, agree on their meaning (usually through the use of reference standards), and agree on the type of scales that will be used for the evaluation [3, 6]. The process involves several hours of listening sessions during which assessors' performance is monitored and feedback is provided to increase consistency between individual assessors. Once the panel is "calibrated", the actual evaluation takes place, usually with at least two replicate sessions to increase the statistical power and assess the panel reproducibility.

Conventional sensory methods, such as DA, with trained panels are very comprehensive and known to produce robust and repeatable results, as documented by numerous scientific publications (for examples of application in the audio field see e.g. [7] and [8]). Unfortunately, DA has major drawbacks that make it not well suited for industrial applications. It is a very time consuming method that requires several hours of listening sessions (mostly due to the extended training phase), and may represent a significant bottleneck if rapid feedbacks on a specific product or product set are needed. Secondly, it is a costly method. Training and maintaining a sensory panel can be a significant spending in the industry and the budget allocated to sensory assessment in product development projects is often insufficient for conducting a proper evaluation.

In order to address these drawbacks, there has been an increasing interest in recent years in developing alternative methodologies that may provide valid results without the need for extensive training. Such alternative rapid sensory methods may give a significant benefit for speeding up innovation processes and saving costs of expensive instrumental analysis by rapidly guiding R&D towards relevant consumer-relevant sensory parameters. Recent reviews on the topic [9, 10] classify rapid sensory methods into verbal-based, similarity-based, and reference-based methods. Verbal methods are based on monadic evaluations of a number of products on individual sensory descriptors. Examples of this class of methodology are Free Choice Profiling [11], Flash Profile [12], and check-all-that-apply (CATA) questionnaires [13]. Similarly to classic descriptive analysis, these methods produce a descriptive sensory profiling of the products, while bypassing the time-consuming steps of attribute and scaling alignment [9]. In the second class of methods, similarity-based, assessors are presented with all products simultaneously, and give a global evaluation expressed as perceived inter-product differences (unlike verbal-based methods which require assessors to decompose the stimulus into multiple attributes). A popular similarity-based method is projective mapping [2] – the method that will be the focus in this research – which introduced the idea of expressing product differences as Euclidean distances by means of projection onto a two dimensional space. The third class is referred as reference-based methods and include methods such as polarized sensory positioning [14], the pivot profile© [15], and polarized projective mapping [16]. The basic idea behind these methods is to compare products to one or more target products, so that the assessors' evaluation reflects the perceived distance from the chosen reference(s).

Many of these methods have gained huge popularity in recent years, and professionals in industries as varied as food and beverages, cosmetic, personal care, fabrics, and automotive are increasingly interested in rapid sensory approaches [9, 10]. Rapid sensory methods have recently been indicated as a possible alternative to DA in the context of audio evaluation [1]. Initial applications in the audio field have appeared for evaluation of sound environments [17, 18]. Even if rapid methods seem promising, several methodological questions need to be addressed before one can advocate their use with confidence. In particular, it is unclear if they can produce reliable data in the context of audio products, especially when small sensory differences need to be investigated.

1.2 Projective mapping

Projective mapping (PM) is a rapid sensory profiling method originally introduced to the field of sensory food evaluation by Risvik and colleagues [2]. The basic idea behind the method is that inter-perceived product differences can be expressed as Euclidean distances in a two-dimensional surface. The method consists in presenting a set of products simultaneously to each assessor, together with a large sheet of blank paper, usually of rectangular shape. Assessors are then instructed to evaluate the perceived similarities (or dissimilarities) between the products, by positioning them on the sheet in such a way that two products should be placed close if they are perceived as similar and far apart if they are perceived as different. Assessors are usually free to choose their own criteria, although in later applications some authors have found useful to restrict the task to a specific aspect [19].

The method produces data in the form of the position of each product on the sheet, which are then digitalized in data matrix with products as rows, and X-coordinate and Y-coordinate on the sheet as columns. It is customary to place the origin of the coordinate system in the bottom left corner (though it can be placed anywhere). 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, [20, 21]). The end result of PM and UFP is a perceptual map of the stimuli under test supplemented by sensory descriptors associated to each of them. An example of such map for an individual assessor is shown in Figure 1.

PM data are usually analyzed by Multiple Factor Analysis (MFA, [22]), a multivariate data analytical technique that seeks the common structure between several blocks of variables (i.e. the individual assessors in a PM task) describing the same observations (the products).

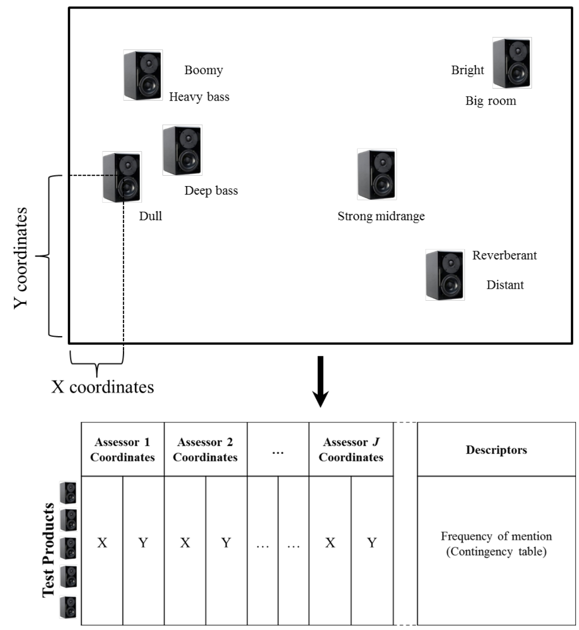

Fig. 1: Pictorial representation of a completed PM sheet for one assessor and resulting data structure over a panel.

Briefly, MFA works as follows: denoting X_j a matrix containing PM coordinates for the test products for $1,\ldots,j$ assessors, MFA starts by computing an initial PCA on each individual matrix X_j (containing sample coordinates for individual assessors), and subsequently transformed into a new matrix $X_{NEW\,j}$ such that:

$$X_{NEW\,j} = \frac{1}{\sqrt{\lambda_1^j}} \times X_j \tag{1}$$

where λ_1^j represents the first eigenvalue of the initial PCA of matrix X_j. The quantity $\sqrt{\lambda_1^j}$, called first singular value, is basically a matrix equivalent of the standard deviation. This procedure corresponds to a normalization (i.e. the first eigenvalues of the transformed $X_{NEW\,j}$ matrices are all equal to 1) that prevents the blocks with the largest variance from exerting an overwhelming influence: in the context of PM, this means accounting for individual assessors' differences in the use of the projective space. After this step, the data blocks are concatenated in a global data table on which a new PCA is run, i.e. by singular value decomposition of the matrix $X_{NEW} = [X_{NEW1}|X_{NEW2}|\ldots|X_{NEW\,j}]$. The descriptive data from UFP are usually treated as supplementary variables to the MFA on the PM coordinates. "Supplementary" means that these data are not used to construct the MFA model, but correlation coefficients of the UFP sensory descriptors are calculated and can be presented in the product space to aid the interpretation and reveal which underlying sensory attributes the assessors took into account when sorting the products on the PM sheet. It is important to remark that this solution provides a sensory configuration that is not necessarily driven by the sensory variables with the largest variance (unlike what would happen with DA data), but by those that are relatively more important for the assessor[1]. Accordingly, some authors observed that this method can be thought of as producing both quantitative and qualitative sensory information [24].

1.3 Aims of the present study

PM has gained popularity in recent years, with numerous publications demonstrating its validity as a sensory method and its suitability with trained assessors and consumers alike [19, 23, 25]. Although primarily associated with food sensory science, PM has been successfully employed in other product categories, e.g. for haptic evaluation of cosmetics [26] and different textured materials [27], and has been indicated as a potentially interesting method for evaluation of sounds as well [1].

[1] According to Pagès, this is the main advantage of PM over DA [23]. The latter produces (ultimately) a data matrix crossing products and descriptors. Such data, typically containing mean values over assessors, is then mean-centered column-wise and analyzed by Principal Component Analysis (PCA), either by giving identical weight to the same variables after a normalization procedure so that each descriptor gets the same variance (usually dividing the data by the sample standard deviation), or by keeping the weight of each descriptors proportional to its variance (unscaled PCA). Whichever solution one chooses, Pagès [23] observes, the weights given to the descriptors do not necessary correspond to the actual importance for the subject.

The present work introduces projective mapping as a method for perceptual audio evaluation. The first aim of this study is to evaluate the performance of this method for discrimination and description of a set of high-end loudspeakers. Additionally, we are interested in evaluating the method performance with different types of assessors. Current standards (e.g., ISO 8586-2 [28]) classify sensory assessors with regards to degree of sensory expertise, ranging from naïve assessors, essentially consumers or product users with no specific qualifications, to expert assessors, who are characterized by a high degree of sensory acuity and expertise. Usually, only the latter group is used for DA task in the audio field [3, 4, 5, 6], however, as already mentioned, PM is regarded as a consumer-friendly method due to its simplicity. Nevertheless, previous research has also indicated that sensory expertise can increase both discrimination and description in such tasks [29, 30]. It is of both scientific and practical interest to know whether these subjects can perform a valid assessment of audio products, where differences may be relatively subtle for inexperienced listeners. Therefore, a second aim of this research is to compare results obtained from experts and naïve assessors with regards to similarity of configurations obtained from projective mapping, as well as the semantic output obtained from UFP.

2 Methods

2.1 Stimuli

The test products consisted of 5 sets of high-end loudspeakers set up in a stereo configuration. The loudspeakers were selected from different brands and covered a wide of range of spatial impressions and spectral signatures. Three of them had built-in amplifiers, whereas two (Speakers 3 and 5) were passive and demanded external amplification. All loudspeakers were comparable in terms of frequency bandwidth, but had different timbral signatures. Additional information on the loudspeakers are given in [1]. The program material consisted of a set of 5 tracks selected by a tonmeister to cover a wide range of genres, dynamics, spatial, and timbral properties (Table 1). All tracks were level-aligned [2] to avoid any level-dependent effect on the sound perception and had an average duration of 25 seconds (minimum 19 s, maximum 34 s).

[2] The track levels were decreased to match the perceived loudness of the quietest track in order to avoid introducing additional noise to the original recordings.

In order to simulate the playback of the music clips trough the five sets of loudspeakers, a complete auralization system was used for the acquisition of loudspeakers' characteristics and for the reproduction of the clips over headphones. All tracks were convolved with binaural room impulse responses (BRIR) of the five pairs of loudspeakers set up in a stereo configuration (i.e. $\pm 30^{o}$). The impulse responses of the Left and Right loudspeakers were measured from -90^{o} to $+90^{o}$ with a resolution of 1^{o} (0^{o} being the frontal direction) using a custom Head and Torso Simulator (Brüel & Kjær HATS 4100). The measurements were performed in a listening room at Bang & Olufsen that follows the IEC 60268-13 recommendations [31] and the loudspeakers were level-aligned using a 5 s pink noise so that the level at the ears of the HATS was 74 dB(C). The complete auralization protocol used in the study is described in [1].

2.2 Assessors

In accordance with the aims of the study, two panels of assessors with different degree of sensory expertise were recruited. The first panel included 11 assessors defined as "experts" according to the ISO 8586-2 standard definition [28]. Their age varied from 27 to 61 (Mean=42). In order to increase the size and robustness in the dataset, nine of them performed a duplicate evaluation, resulting in a total of 20 PM evaluations for this panel. The second panel included 51 assessors (39 Male, 12 Female) with no previous experience of sensory evaluation, and therefore defined as "naïve" according to ISO 8586-2:2008. They were aged 18 to 61 (Mean= 34). The naïve assessors evaluated the stimuli in a single evaluation. Assessors in the expert panel were Bang & Olufsen's employees with extensive expertise in acoustics, and who attend listening training sessions on a weekly basis. Assessors in the consumer panel were recruited among students and staff at the University of Southern Denmark and had no prior experience with audio evaluation. All assessors self-reported normal hearing conditions except for four individuals (two in each panel). Diagnostic analysis on these assessors (not shown in the paper) did not indicate any observable deviations from the behavior of the other assessors. All assessors participated in the study on a voluntary basis without any compensation for their time.

Table 1: Music material used in the experiment.

Track	Genre	Title (Artist)
1	Rock	Bombtrack (Rage Against the Machine)
2	Classical /orchestra	Dance of the Tumblers (Minnesota Orchestra)
3	String quartet	Petite Symphonie à Cordes II. Assez vif (Scottish Ensemble)
4	Jazz	Sing Sang Sung (Gordon Goodwin's Big Phat Band)
5	Vocal	Tom's Diner (Suzanne Vega)

2.3 Experimental procedures

Assessors performed the PM task using a data acquisition software that simulated a projective mapping sheet and allowed assessors to drag and move each test stimulus on the surface, and also provided the possibility to add descriptors. An example of the user interface used in the experiment is shown in Figure 2. Upon arrival, assessors received a brief introduction to the task by an experimenter and were familiarized with the data acquisition software. According to the principles of PM, the assessors were instructed to listen to the audio samples and position them on the map close to each other if they sounded similar and further apart if they sounded different. It was emphasized to the assessors that they had to do so according to their own criteria and that there were no "right" or "wrong" ways to do this. Finally, assessors were asked to possibly provide a verbal description of the stimuli (UFP). Each assessor performed five consecutive PM evaluations on the loudspeakers (one for each track). The order of evaluation for the tracks was randomized between assessors. Within each track, the name of the loudspeakers was blinded to the assessors (with random 3-digit code for the consumer panel, and with sequential A-E letters for the expert panel) and changed from track to track to reduce reliance on memory and increase the focus on the percept. The order in which the audio samples appeared on the software was also randomized, however, assessors were free to listen each sample any number of time they want and go back and forth between them. Assessors were allowed to take short breaks between tracks and refreshments were also provided. The average time to complete the task (all 5 tracks) was 27 minutes for the consumer panel and 50 minutes for the expert panel. The reason behind the difference in time spent on the task is unclear, but it is ostensibly due to the different data acquisition softwares, and possibly to differences in the cognitive strategies adopted by the two panels [3].

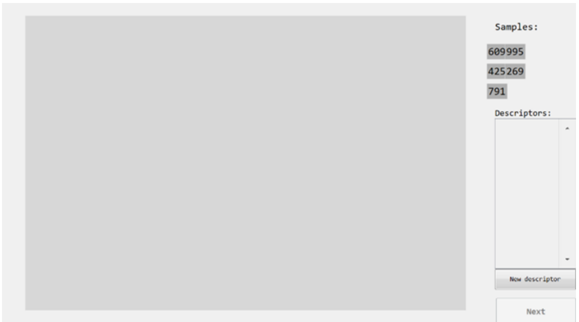

Fig. 2: Interface for the PM data collection in Acqui.

2.4 Experimental setup and reproduction system

The experiments took place in a quiet office environment for the panel of consumers, and in a listening room that follows the IEC 60268-13 recommendations for the expert panel. For the expert panel, the experiment was conducted on a Lenovo Thinkpad with an external monitor, an external sound card (M-Audio Fast Track Pro), and dynamic binaural rendering through a USB-connected head-tracker attached on the top of a set of Sennheiser HD 650 headphones that were playing the stimuli [1]. For the consumer panel, the experiment was conducted on a Lenovo Thinkpad with an external monitor, and an external sound card (Audio 4 by Native Instruments). The sound stimuli were played over a pair of Sennheiser HD650 headphones, though unlike the expert panel a static binaural rendering was used (no head-tracking). The data acquisition software was MaxMSP [32] for the expert panel, and Acqui [33] for the consumer panel.

[3] Since the experts are more used to listen to audio materials in a analytical way, they may potentially perceive subtler and more varied differences. Thus, the extra time to complete the tast could be due to them taking more aspects to take into account.

2.5 Data analysis

The experiment produced two types of data: (1) perceptual spaces generated from PM and (2) descriptive outputs from UFP.

The PM data consisted of the X and Y coordinates for each loudspeaker captured by the data acquisition software using a coordinate systems with the bottom left corner as the origin. These data were analyzed by Hierarchical Multiple Factor Analysis (HMFA, [34]) to assess similarities and dissimilarities between the loudspeakers, as well as to compare the perceptual spaces generated by the two panels. HMFA is a generalization of Multiple Factor Analysis (MFA, [22]) applicable to data that are hierarchically structured. The goal of the analysis is to integrate different groups of variables describing the same observations (in this case, the five loudspeakers), and to balance their influence by using a scaling procedure based on the variance associated with each group of variables. HMFA balances the role of the groups of variables by applying MFA in a stepwise fashion within each node of the hierarchy. Thus, it provides graphical displays that can be interpreted from a global perspective, as well as from the perspective of the groups of variables hierarchically defined [34]. In the present dataset, we defined four levels of hierarchy, resulting in the structure illustrated in Figure 3. HMFA applies a succession of MFA to each level of the hierarchy from the bottom to the top, in order to balance the groups of variables within every level. It starts by computing individual Principal Component Analysis (PCA) models on these mean-centered data and applies a scaling factor corresponding to the square root of the first eigenvalue obtained in the individual PCA models. HMFA proceeds sequentially to perform another MFA on the next four groups of variables, and then performs a global PCA on the merged dataset. The first level corresponds to the individual X- and Y- coordinates from each PM sheet, i.e. 405 groups of variables (20*5 for the experts and 51*5 for the consumers). The second level corresponds to the individual tracks and contains the PM coordinates (following the scaling procedure just described) and the contingency table with UFP data for each track. The third level of the hierarchy corresponds to the global configurations of the two panels taking all five tracks into account. The fourth and final level corresponds to a global PCA on the merged dataset (Fig. 3).

Graphical visualization was used to present and describe the perceptual space obtained by the PM procedure at the aggregated level. Partial configurations obtained from the experts and consumers were also retrieved from the HMFA model and compared both qualitatively (i.e. by visual inspection) and quantitatively by computing RV coefficients between configurations obtained by two panels. The RV coefficient [35] is a multivariate generalization of the simple correlation coefficient, and is widely used for estimating similarities between sensory spaces on the same products [25]. It takes values between 0 and 1, where 1 indicates identical configurations. RV coefficients were computed between partial configurations on a track by track basis considering either a two or a four MFA dimensions solution. Their significance was assessed using a permutation test as suggested in [36].

The second part of the analysis concerned the description of the stimuli from the UFP step. A contingency matrix containing the frequency of mention of each descriptor for each loudspeaker was cross-tabulated on track by track basis. Prior to that, semantic grouping of descriptors relating to the same percept was performed (e.g., "bass heavy", "heavy bass", and "bass sounds heavy"). Additionally, only descriptors mentioned at least five times in total were retained for further analysis. This decision was done to improve interpretability of the UFP data, and is based on common practices [20, 21, 25, 29]. In general, both procedures (semantic grouping and the use of a cut-off point) are customary and necessary to improve the interpretability of UFP data. The contingency tables obtained for each track were included in the respective HMFA models as supplementary variables, meaning that they were not used to compute the HMFA, but that correlation coefficients of the UFP descriptors with each model dimension were calculated and presented to aid the interpretation of the sensory spaces. All analyses were performed in `R` [37] using functions from the `FactoMineR` package [38].

3 Results

3.1 Perceptual distances between the loudspeakers (PM data)

The first part of the analysis pertains to the perceptual distances between the loudspeakers based on the PM configurations. Figure 4 shows the first two dimensions

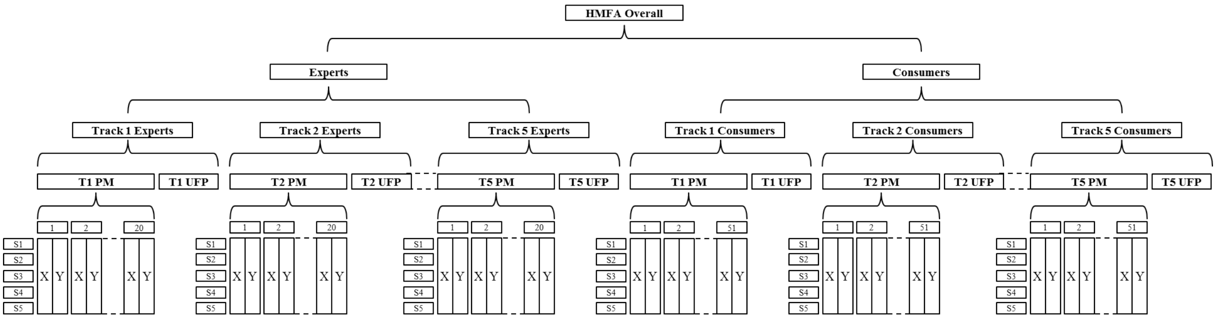

Fig. 3: Visual representation of the data structure for the HMFA analysis.

of the overall HMFA model with all the tracks included which account for 60.7% of the variance in the PM data. The first dimension shows that the largest difference is between Speaker 5 and Speaker 3, whereas the second dimension opposes Speaker 2 and 4 to the remaining three (Fig. 4). Inspecting the corresponding plots for the HMFA models on individual tracks (Figure 5) confirmed the indications of the overall analysis. All models had similar explained variance along the first two dimensions (Track 1: 66.1%; Track 2: 66.4%; Track 3: 66.4%; Track 4: 69.1%; Track 5: 61.4%) to the overall model. In particular, Track 1, Track 2, and Track 4 showed an almost identical ranking of the loudspeakers on the first two HMFA dimensions. The HMFA model for Track 3 shows an opposition between Speaker 5 and Speaker 1, 2, and 3 on the first dimension, with the latter being less differentiated than in the other models. The configuration for Track 5 (the vocal only track) was somewhat different than in other tracks. In this model the main opposition on the first dimension is between Speaker 3 and Speaker 4, whereas the second dimension singles out Speaker 5 from the rest. All models indicate that Speaker 5 sounded the most different from the other loudspeakers. There are several possible reasons behind this result, although this discussion is outside of the present paper.

The second part of the analysis of the PM data consisted in assessing the level of agreement between the expert and the consumer panel. Visual assessment of Figure 4B reveals that the partial points for the two separate configurations are very close to each other for all loudspeakers, indicating that the two panels provided highly congruent product spaces at an overall level. Looking at models for individual tracks, some minor differences are noticeable regarding the position of some loudspeakers in the second HMFA dimension (e.g., Speaker 1 in Track 1), although at an overall level there seem to be a good agreement as the coordinates on the first HMFA dimension are almost always identical for both panels. The RV coefficients between partial configurations fully confirmed these visual interpretations. As reported in Table 2, the RV coefficient between the expert and the consumer configuration was extremely high (0.97 considering two HMFA dimensions, meaning nearly identical configurations). The RV values considering the models for individual tracks ranged from 0.78 to 0.90, with an average of 0.85. Although slightly lower than in the overall model, these values are still considerably high (to put things into perspective, an RV coefficient of 0.85 would be considered a good performance when evaluating the replicability of DA results from the same panel (Lawless, 1984). On the other hand, when including the third and fourth dimensions, the RV values were lower and were generally not statistically significant. This means that configurations between experts and consumers were uncorrelated after the first two dimensions. This could indicate that the two panels focused on different things, or alternatively that higher order dimensions captured mostly random noise. Our interpretation is the latter because, when considered together with the UFP output (described in the next section), the HMFA models did not produce more than two interpretable dimensions.

3.2 Sensory descriptive profiling (UFP data)

The second part of the analysis pertained to the description of the loudspeakers provided by the assessors during the UFP step. Table 3 shows significant correlations of UFP terms with the first and second HMFA components for both panels, on a track by track basis. Recall that these descriptions are not actively used in the construction of the HMFA dimensions (which are based on

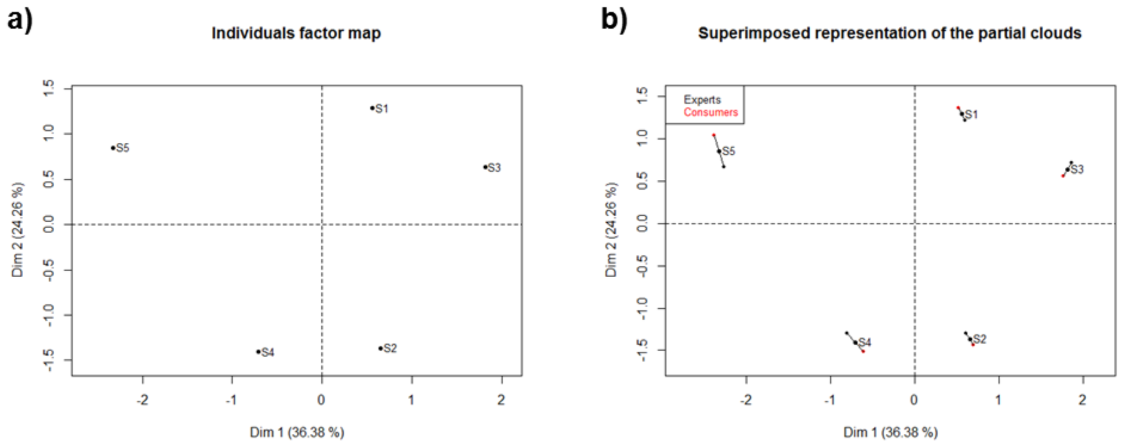

Fig. 4: Positions of the five loudspeakers on the first and second HMFA dimensions. Figure 4B shows the same plot with superimposed partial configurations obtained from the two panels.

the PM coordinates only), but are used for interpretation purpose, i.e. to understand the sensory properties underlying the perceptual configuration obtained from the PM data. The first thing to notice in Table 3 is that only the first HMFA dimension has a substantial number of significant correlations, whereas the second dimension hardly has any. Regarding the differences between the loudspeakers, they were clearly related to bass strength and bass depth for Track 1 and Track 4. In Track 1, the main HMFA dimensions opposes loudspeakers characterized as *boomy*, *heavy*, *deep bass* (primarily Speaker 5, according to Fig. 5) to those who sounded *thin* (primarily Speaker 3). The second dimension was less well described in the consumer panel, but for the experts was negatively correlated to the timbral attribute *warm* and the spatial attribute *wider*. Track 4 showed the exact same ranking of the loudspeakers on the first HMFA dimension and opposed *boomy*, *bass*, *heavy bass* versus *thin*. The second HMFA dimension was positively correlated with the attribute *muffled* (primarily associated with Speaker 4). The other three tracks were not so well characterized, with fewer significant UFP descriptors in the first dimension of their respective HMFA models and no significant correlations with the second dimension. Track 3 was poorly described by both panels. The only significant descriptor was *clear* which was positively correlated to the first dimension and thus negatively associated with Speaker 5 and positively with Speaker 3 (Table 3 and Fig. 5).

Track 5 (the vocal only) was somewhat different from the other tracks. For this track the expert panel did not produce any significant UFP correlations, whereas the results for the consumer panel suggest that the differences between the loudspeakers were primarily due to timbral and spatial aspects as the first dimension of this group opposed the attribute *clea* to *reverb*. Considering the differences between the two panels overall, Table 3 shows that the experts tended to have more significant UFP descriptors than the naïve assessors, indicating that it was easier for the expert assessors to verbalize their sensory impressions and to do so consistently. Moreover, the expert panel tended to use more precise descriptors, such as *boomy*, *kick*, *bright*, *thin*, and *warm*, which revealed their familiarity with known sensory lexicon in the audio field (e.g., [39]). The consumers, on the other hand, tended to used more integrated (i.e., that combine several product attributes into one) attributes, such as *bass*.

4 Discussion

The overall aim of this work was to introduce PM as a rapid sensory method for perceptual evaluation of audio products. The first sub-aim was to evaluate the performance of PM to characterize a set of loudspeakers from a sensory point of view. The method successfully discriminated between the loudspeakers produced interpretable configurations for all tracks. All HMFA models (with the partial exceptions of the model for

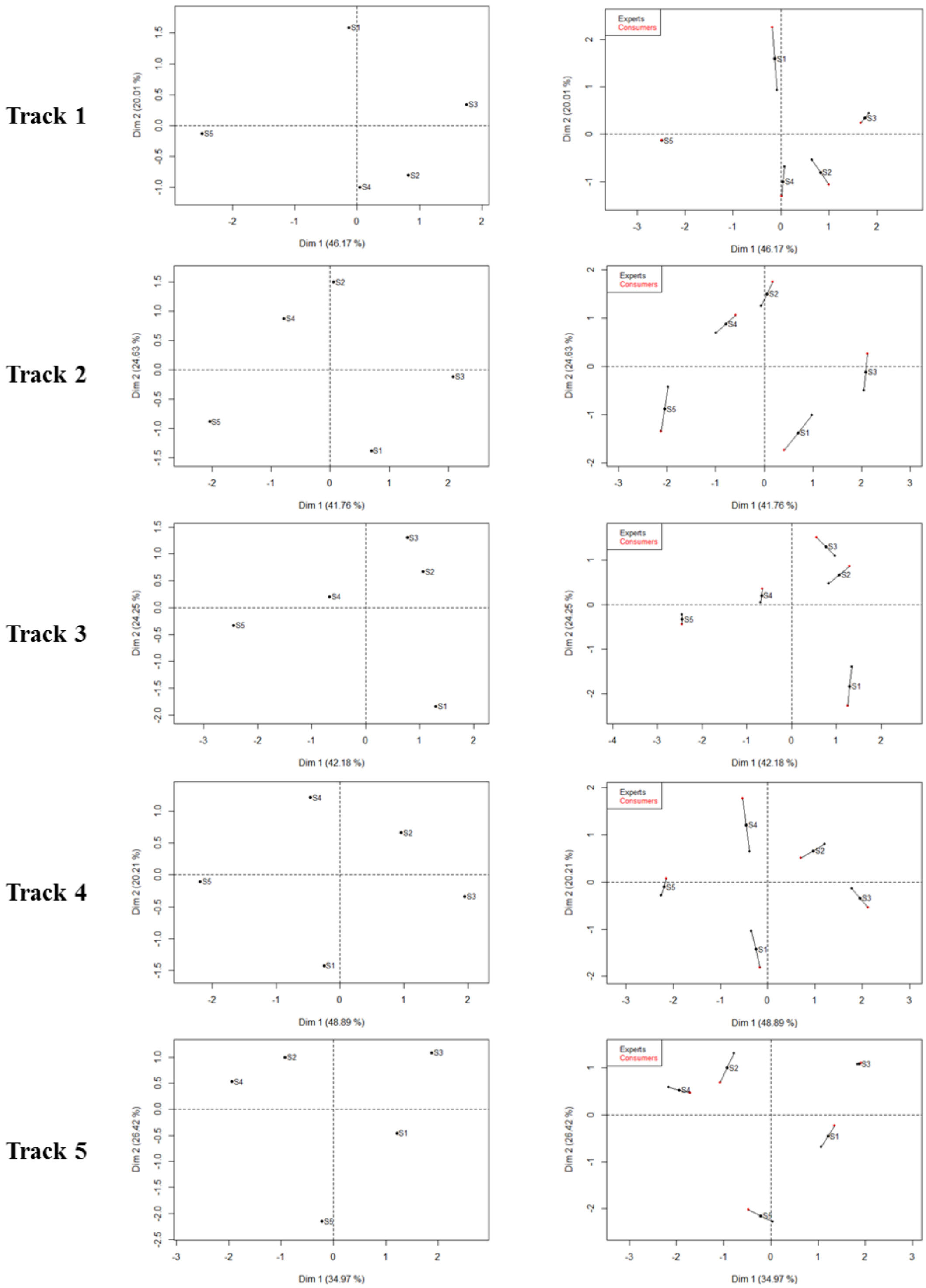

Fig. 5: Positions of the five loudspeakers on the first and second HMFA dimensions on a track by track basis (left), and superimposed partial configurations obtained from the two panels (right).

Table 2: RV coefficients between partial configurations obtained from experts and consumers considering 2 and 4 HMFA dimensions. The significance of RV coefficients is tested according to [37].*** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$; $^{n.s}$ $p > 0.5$.

Experts vs Consumers	RV (2 HMFA dimensions)	RV (4 HMFA dimensions)
All tracks	0.97 ***	0.93 $^{n.s}$
Track 1	0.85 ***	0.78 *
Track 2	0.78 *	0.75 $^{n.s}$
Track 3	0.90 ***	0.82 $^{n.s}$
Track 4	0.84 *	0.73 $^{n.s}$
Track 5	0.90 ***	0.81 $^{n.s}$

Track 5 – the vocal only) showed very similar indications regarding the perceptual differences between the loudspeakers, confirming the validity of PM as a sensory method.

According to the UFP descriptors obtained during the experiment, the main differences between the products pertained to bass depth and bass strength for Tracks 1-4, and to clarity and reverb in Track 5 (which only included vocals). Nevertheless, all models produced no more than two HMFA interpretable dimensions. This suggests that PM can provide a coarse description of the main perceptual differences, but possibly not as detailed as a method like DA would. This conclusion may also be related of the specific product set used in the study, so more research is needed to conclude with certainty that this is the case. However, the key thing to consider is that PM is a very fast method compared to classical DA, therefore the possible loss of precision must be evaluated in the light of the significant time and cost savings in the data collection process enabled by this technique. The data analysis process is also made significantly faster by the use of dedicated software such as that used in this work, so that results can be plotted within minutes from the data collection process. Handling the UFP data is the only relatively laborious step for the experimenter, however, unless the results need to be formally reported (e.g. in a publication), the cleaning and semantic grouping steps may not be necessary. Moreover, in some cases the protocol can be modified to reduce the number the UFP terms, e.g. by setting a maximum number of descriptors allowed per each loudspeaker. Framing the task towards a particular sensory aspect (“partial” PM) is also known to reduce the number and improve the quality of UFP terms.

An important thing to notice is that PM should also be seen as a supplement, not necessarily as an alternative to DA. For example, in the context of a classical DA process [4], PM could be used as a basis for selecting the stimuli (e.g., to reduce the number of products from a larger pool). The UFP step in PM is also very useful as a word elicitation technique for DA and other consensus vocabulary techniques like CATA and Flash Profiling [30, 25].

The second aim of the work was to evaluate the performance of assessors with different degrees of sensory expertise in a PM task. In the case study presented, the two panels produced very similar indications regarding the differences between the loudspeakers for all tracks, and nearly identical configurations when considering all tracks in combined dataset. This suggests that untrained assessors can be used validly for sensory profiling of audio with the PM task, which is in good agreement with previous research on different product categories [25, 26, 29, 30]. This result is important as consumer panels can be an inexpensive alternative to trained panels. Moreover, as software support for PM continues to grow, it might be interesting in some cases to conduct PM tests remotely using consumers in their own homes, making it possible to significantly reduce the amount of time needed to prepare, conduct, and evaluate PM experiments [4]. It is important to be cautious about the expert vs consumer aspect. The two panels had very different sizes, with the consumer panel having 2.55 times more data points than the experts. The size of the consumer panel was chosen according to extant research in food science that indicate a minimum of 50 assessors when performing PM with consumers [40]. However, it would be interesting to know whether the same results would

[4] An example of such application is "NappingPlayer", a software for conducting PM experiments on video quality in Android tablets (publicly available at https://github.com/slhck/napping-player).

Table 3: Negative and positive correlations coefficients between UFP descriptors and HMFA dimensions for the two panels (Experts vs. Consumers). The significance of the correlation coefficient is estimated using a t-distribution with 3 degrees of freedom. Only significant correlations ($r \neq 0$ with $p<0.05$) are shown.

Track	Dim 1 (Exp.)	Dim 1 (Cons.)	Dim 2 (Exp.)	Dim 2 (Cons.)
1	Boomy (–0.95) Deep bass (–0.94) Heavy (–0.94) Kick (–0.94) Bass Heavy (–0.93) Narrow (–0.90)	Bass (–0.91) Deep (-0.89)	Warm (–0.90) Wider (–0.90)	-
	Thin (0.92)	Flat (0.90)	-	-
2	Dull (–0.95) Narrow (–0.94)	-	-	-
	Bright (0.87) Close (0.97)	-	-	-
3	-	-	-	-
	-	Clear (0.93)	-	-
4	Boomy (–0.92)	Bass (–0.91) High bass (–0.91) Heavy bass (-0.87)	-	-
	Thin (0.91) Narrow (0.90)	-	-	Muffled (0.91)
5	-	Clear (-0.93)	-	-
	-	Reverb (0.87)	-	-

be obtained if the two panels had approximately the same size. Moreover, due to practial reasons the experimental conditions were slightly different for the two panels. In particular, different sound cards and rendering methods (static vs. dynamic binaural) may have provided different timbral and spatial impressions. Since the task in each experiment was to compare the different loudspeakers, we can hypothesize that these different experimental setups had a minor impact on the results (at least, this is what the results suggest) but this would require to be further investigated. Additionally, it should be noted that 9 out 20 PM evaluations from the experts were in fact duplicate. A detailed analysis of this aspect is beyond the scope of this paper but it requires further attention as some learning may have occurred between the first and the second evaluation, thus possibly reducing the variance in the experts' assessor data. We are currently working on a separate publication where the issue of consistency across evaluations and inter-assessor reliability are explicitly addressed.

Another point to consider is that, although perceptually the two panels gave identical indications, the descriptive output (UFP) of the experts produced more and better (i.e. more precise and interpretable) descriptors, indicating that experts outperformed consumers as soon as language was involved. Taken collectively, the results obtained appear consistent with research on the topic indicating that sensory and product expertise are relatively more important for verbalization of sensory perceptions than for perceptual ability [29, 41, 42, 43]. In practical applications, an expert panel seems therefore a better option to conduct PM with, if a description of the perceptual differences is needed. Whether this is necessary depends on several factors that the experimenter should consider in order to make a choice on the most appropriate type of assessors, such as the goal of the test, previous knowledge of the product set, existence of baseline data, whether or not PM is used as a stand-alone method, etc.

Overall, this research suggests that PM can be a useful tool for perceptual evaluation of audio products, especially valuable when time and/or budget are limited. However, further research is advised to better understand the method and provide best practises for its application in the audio field. Future work should include a proper comparison with a method based on consensus vocabulary, which would be needed to further validate the method but especially to understand the level of detail that PM can provide for product sets that present perceptual differences in more than two dimensions. The appropriate benchmark for such comparison is DA, but it would also be interesting to compare PM vis-à-vis other rapid descriptor-based methods such as CATA and Flash Profiling. A possible method improvement in future applications could consist of framing the task towards a pre-defined sensory aspect, as this strategy has been reported to ease the task for the assessors and improve the signal-to-noise ratio in PM tests [19]. Moreover, the speed of the method is such that separate 'partial' PM evaluations (e.g., running separate evaluation on timbral and spatial aspects) can be performed in the same session. These datasets could then be examined separately to look at individual aspects, or combined in a HMFA model (in the same way as the different tracks were combined in this paper) to obtain a global characterization of the products under test. Other issues to consider in future work include the possibility to run an initial training step prior to evaluation which has been reported to increase the assessors' performance in PM task [30], and considerations about the choice and number of audio samples that is reasonable to include.

5 Summary

This work has presented an application of the projective mapping method for perceptual evaluation of five high-end loudspeakers evaluated by both naïve and expert assessors. The results showed that PM was effective in providing an overview of the main perceptual differences between the products under test, and UFP data supplemented the analysis by revealing the underlying sensory attributes responsible for said differences. A high agreement between configurations obtained by the two panels was observed for all evaluated tracks, suggesting that PM as a method might be applicable also with subjects without formal training in sensory evaluation. Nevertheless, expert assessors provided a higher number of significant descriptors in the UFP part, and these were generally more precise than those provided by the naïve assessors. Taken collectively, these results indicate that sensory expertise is related to better descriptive – but not necessarily perceptual – ability in a PM task. Several methodological questions stemming from this research are identified. Future research is advised to understand the correct range of application of PM and provide best practices for the use of PM in perceptual audio evaluation.

References

[1] Moulin, S., Bech, S., and Stegenborg-Andersen, T., "Sensory Profiling of High-End Loudspeakers Using Rapid Methods-Part 1: Baseline Experiment Using Headphone Reproduction," in *Audio Engineering Society Conference: 2016 AES International Conference on Headphone Technology*, Audio Engineering Society, 2016.

[2] Risvik, E., McEwan, J. A., Colwill, J. S., Rogers, R., and Lyon, D. H., "Projective mapping: A tool for sensory analysis and consumer research," *Food Quality and Preference*, 5(4), pp. 263–269, 1994.

[3] Bech, S. and Zacharov, N., *Perceptual audio evaluation-Theory, method and application*, John Wiley & Sons, 2006.

[4] Pedersen, T. H. and Zacharov, N., "How many psycho-acoustic attributes are needed," *Journal of the Acoustical Society of America*, 123(5), pp. 3163–3163, 2008.

[5] Stone, H. and Sidel, J., *Sensory Evaluation Practises (3rd Ed.)*, Tragon Corporation, 2003.

[6] Lawless, H. T. and Heymann, H., *Sensory evaluation of food: principles and practices*, Springer Science & Business Media, 2010.

[7] Zacharov, N. and Koivuniemi, K., "Perceptual audio profiling and mapping of spatial sound displays," in *Proceedings of the Int. Conf. on Auditory Display*, pp. 95–104, 2001.

[8] Zacharov, N. and Lorho, G., "Sensory analysis of sound (in telecommunications)," in *Proc. of European Sensory Network Conference*, volume 5, 2005.

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

[10] Varela, P. and Ares, G., "Sensory profiling, the blurred line between sensory and consumer science. A review of novel methods for product characterization," *Food Research International*, 48(2), pp. 893–908, 2012.

[11] Williams, A. A. and Langron, S. P., "The use of free-choice profiling for the evaluation of commercial ports," *Journal of the Science of Food and Agriculture*, 35(5), pp. 558–568, 1984.

[12] Dairou, V. and Sieffermann, J.-M., "A comparison of 14 jams characterized by conventional profile and a quick original method, the flash profile," *Journal of Food Science*, 67(2), pp. 826–834, 2002.

[13] Adams, J., Williams, A., Lancaster, B., and Foley, M., "Advantages and uses of check-all-that-apply response compared to traditional scaling of attributes for salty snacks," in *7th Pangborn sensory science symposium*, 2007.

[14] Teillet, E., Schlich, P., Urbano, C., Cordelle, S., and Guichard, E., "Sensory methodologies and the taste of water," *Food Quality and Preference*, 21(8), pp. 967–976, 2010.

[15] Thuillier, B., Valentin, D., Marchal, R., and Dacremont, C., "Pivot© profile: A new descriptive method based on free description," *Food Quality and Preference*, 42, pp. 66–77, 2015.

[16] Ares, G., de Saldamando, L., Vidal, L., Antúnez, L., Giménez, A., and Varela, P., "Polarized projective mapping: Comparison with polarized sensory positioning approaches," *Food Quality and Preference*, 28(2), pp. 510–518, 2013.

[17] Blumenthal, D. and Herbeth, N., "Use of rapid sensory methods in the automotive industry," *Rapid Sensory Profiling Techniques: Applications in New Product Development and Consumer Research*, p. 427, 2014.

[18] Kaplanis, N., Bech, S., Tervo, S., Pätynen, J., Lokki, T., van Waterschoot, T., and Jensen, S. H., "A method for perceptual assessment of automotive audio systems and cabin acoustics," in *Audio Engineering Society Conference: 60th International Conference: DREAMS (Dereverberation and Reverberation of Audio, Music, and Speech)*, Audio Engineering Society, 2016.

[19] Dehlholm, C., Brockhoff, P. B., Meinert, L., Aaslyng, M. D., and Bredie, W. L., "Rapid descriptive sensory methods–comparison of free multiple sorting, partial napping, napping, flash profiling and conventional profiling," *Food Quality and Preference*, 26(2), pp. 267–277, 2012.

[20] Perrin, L., Symoneaux, R., Maître, I., Asselin, C., Jourjon, F., and Pagès, J., "Comparison of three sensory methods for use with the Napping® procedure: Case of ten wines from Loire valley," *Food Quality and Preference*, 19(1), pp. 1–11, 2008.

[21] Perrin, L. and Pagès, J., "Construction of a product space from the Ultra-Flash Profiling method: Applications to 10 red wines from the Loire Valley," *Journal of Sensory Studies*, 24(3), pp. 372–395, 2009.

[22] Abdi, H., Williams, L. J., and Valentin, D., "Multiple factor analysis: principal component analysis for multitable and multiblock data sets," *Wiley Interdisciplinary Reviews: Computational Statistics*, 5(2), pp. 149–179, 2013.

[23] Pagès, J., "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(7), pp. 642–649, 2005.

[24] Chollet, S., Lelièvre, M., Abdi, H., and Valentin, D., "Sort and beer: Everything you wanted to know about the sorting task but did not dare to ask," *Food Quality and Preference*, 22(6), pp. 507–520, 2011.

[25] Reinbach, H. C., Giacalone, D., Ribeiro, L. M., Bredie, W. L., and Frøst, M. B., "Comparison

of three sensory profiling methods based on consumer perception: CATA, CATA with intensity and Napping®," *Food Quality and Preference*, 32, pp. 160–166, 2014.

[26] Parente, M., Ochoa Andrade, A., Ares, G., Russo, F., and Jiménez-Kairuz, Á., "Bioadhesive hydrogels for cosmetic applications," *International Journal of Cosmetic Science*, 37(5), pp. 511–518, 2015.

[27] Dacleu-Ndengue, J., J, F., J, B., Massi, F., Zahouani, H., and Delafosse, D., "Perception of textured materials: visual and tactile evaluation through Napping® procedure," in *11th Pangborn sensory science symposium*, volume 16, p. 16, 2015.

[28] ISO, "ISO 8586-2:2008. Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Expert sensory assessors," 2008.

[29] Giacalone, D., Ribeiro, L., and Frøst, M., "Perception and Description of Premium Beers by Panels with Different Degrees of Product Expertise," *Beverages*, 2(5), 2016.

[30] Liu, J., Grønbeck, M. S., Di Monaco, R., Giacalone, D., and Bredie, W. L., "Performance of Flash Profile and Napping with and without training for describing small sensory differences in a model wine," *Food Quality and Preference*, 48, pp. 41–49, 2016.

[31] IEC, *IEC 60268–13 Sound system equipment – Part 13: Listening tests on loudspeakers*, International Electrotechnical Commission, 1998.

[32] Cycling'74, "Max/MSP: A graphical programming environment for music, audio, and multimedia," 2006.

[33] Laugesen, J. L., "Acqui - Acquisition of Sensory and Cosumer Profiles," 2016.

[34] Le Dien, S. and Pagès, J., "Hierarchical multiple factor analysis: Application to the comparison of sensory profiles," *Food Quality and Preference*, 14(5), pp. 397–403, 2003.

[35] Robert, P. and Escoufier, Y., "A unifying tool for linear multivariate statistical methods: the RV-coefficient," *Applied Statistics*, pp. 257–265, 1976.

[36] Josse, J., Pagès, J., and Husson, F., "Testing the significance of the RV coefficient," *Computational Statistics & Data Analysis*, 53(1), pp. 82–91, 2008.

[37] Team, R. C., "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2013," 2014.

[38] Lê, S., Josse, J., Husson, F., et al., "FactoMineR: an R package for multivariate analysis," *Journal of Statistical Software*, 25(1), pp. 1–18, 2008.

[39] Pedersen, T. H. and Zacharov, N., "The development of a sound wheel for reproduced sound," in *Audio Engineering Society Convention 138*, Audio Engineering Society, 2015.

[40] Vidal, L., Cadena, R. S., Antúnez, L., Giménez, A., Varela, P., and Ares, G., "Stability of sample configurations from projective mapping: How many consumers are necessary?" *Food Quality and Preference*, 34, pp. 79–87, 2014.

[41] Chollet, S. and Valentin, D., "Impact of training on beer flavor perception and description: are trained and untrained subjects really different?" *Journal of Sensory studies*, 16(6), pp. 601–618, 2001.

[42] Guerrero, L., Gou, P., and Arnau, J., "Descriptive Analysis Of Toasted Almonds: A Comparison Between Expert And Semi-Trained Assessors," *Journal of Sensory Studies*, 12(1), pp. 39–54, 1997.

[43] Lawless, H. T., "Flavor description of white wine by "expert" and nonexpert wine consumers," *Journal of Food Science*, 49(1), pp. 120–123, 1984.