

COVER PAGE

Document downloaded by @DAEL

Sat Nov 14 03:28:33 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Objective Characterization of Perceptual Dimensions Underlying the Sound Reproduction of 37 Single Loudspeakers in a Room, Pierre-Yohan Michaud, Mathieu Lavandier, Sabine Meunier and Philippe Herzog, Acta Acustica **vol. 101** (Number 3), 2015, pp. 603-615

DOI
https://doi.org/10.3813/AAA.918856

Acta Acustica united with Acustica
Vol. 101 (2015) 603 – 615

DOI 10.3813/AAA.918856

Objective Characterization of Perceptual Dimensions Underlying the Sound Reproduction of 37 Single Loudspeakers in a Room

Pierre-Yohan Michaud[1)], Mathieu Lavandier[2)], Sabine Meunier[1)], Philippe Herzog[1)]

[1)] LMA, C.N.R.S., UPR 7051, Aix-Marseille Université, Centrale Marseille, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France. pierre-yo@hotmail.fr

[2)] Université de Lyon, Ecole Nationale des Travaux Publics de l'Etat, Laboratoire Génie Civil et Bâtiment, Rue M. Audin, 69518 Vaulx-en-Velin Cedex, France

Summary

This study is focused on the relationships between perceptual and physical evaluations of the reproduction of timbre by loudspeakers in monophonic reproduction. A protocol compatible with both approaches was proposed by Lavandier et al.(2008) to evaluate a set of 12 loudspeakers. It consisted in recording the sound radiated by the loudspeakers in a room. The recorded sounds were submitted to listening tests under headphones where listeners were asked to judge their dissimilarity, highlighting the perceptual dimensions involved using a multidimensional scaling analysis. The recorded sounds were also submitted to signal analysis in order to propose an acoustical attribute describing each perceptual dimension. The present study further evaluated the reproduction of timbre by testing a set of 37 loudspeakers in a different room. Evaluating a larger set of loudspeakers led to constraints concerning the perceptual evaluation. Therefore, two perceptual tasks, free sorting and similarity picking with permutation of references, were compared. The perceptual results and the objective attributes of these experiments consolidate the two timbral dimensions revealed by Lavandier et al. (bass/treble balance, emergence of medium) and highlight a third dimension associated with a spatial sensation (feeling of space) despite a monophonic reproduction. An influence of the listening task was also observed, indicating that the free sorting paradigm might not be the most appropriate to evaluate dissimilarity.

PACS no. 43.38.Md, 43.66.Lj

1. Introduction

This study focuses on quantifying the perceived differences between the reproduction of musical excerpts through different loudspeakers. To our knowledge, most existing listening tests aiming to characterize the sound reproduction by loudspeakers are based on quality (preference) judgments [1, 2, 3, 4, 5]. A few studies deal with the perceived differences between sounds reproduced by different loudspeakers using absolute ratings on different semantic scales [6, 7, 8, 9] or the evaluation of relative differences [10, 11, 12, 13]. The study presented here focuses on quantifying the perceived differences between the reproduction of musical excerpts through different loudspeakers. Most of the studies cited above involve a limited number of loudspeakers, although a large panel of stimuli would allow to obtain a better description of the perceptual space with with potentially more dimensions involved and a weaker stimulus-context dependence. Thus here, a large set of thirty-seven single loudspeakers is considered in order to reinforce the description of the perceptual space and to highlight potential new dimension(s).

In several papers, Gabrielsson and his colleagues investigated perceived differences as well as global sound quality (fidelity, pleasantness). They extracted 8 rating scales from a factor analysis on the correlation between 55 rating scales [6]. The scalings were made for 9 reproducing systems composed of three different loudspeakers placed at different positions in front of the listener and different signal distortions of one of the loudspeaker. The restitution was monophonic. The rating scales were: "clearness / distinctness", "sharpness / hardness-softness", "brightness-darkness", "fullness-thinness", "feeling of space", "nearness", "disturbing sounds" and "loudness". They found positive correlations between the rating of the different perceptual scales and the quality judgments (based on rating scales of fidelity and pleasantness) [7]. Gabrielsson *et al.* [11] measured the perceived similarities between 5 loudspeakers in one experiment and between 6 different reproductions from one loudspeaker (varying the sound level and the frequency response) in another experiment. The loudspeakers were also placed in a room at different positions in front of the listener. They extracted 2 dimensions from a multidimensional scaling (MDS) analy-

Received 16 July 2013,
accepted 06 January 2015.

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

sis of the similarities: "clarity" and "brightness" (balance between bass and treble). In these studies, the perceptual dimensions were not quantified by any objective measure.

Several studies were focused on the link between perceptual dimensions and objective measures [4, 5, 10, 13, 9, 14]. They mostly consisted of "qualitative links" between the shapes of frequency responses of loudspeakers and their evaluations along an overall fidelity scale or a particular perceptual scale [4, 9, 15]. In order to provide quantitative measures that corresponds to perceptual dimensions, Olive [5] proposed attributes based on the spectrum of the signal, whereas Klippel [10], Bramslow [14] and Lavandier *et al.* [13] proposed attributes based on an auditory model.

Olive [5] asked the listeners to evaluate 13 loudspeakers on a preference scale from "really dislike" to "really like" and to rate (after training) the spectral balance and the perceived level of nonlinear distortion. An automated speaker shuffler allowed for fast positioning of four monophonic loudspeakers in order to test all loudspeakers at the same position in the room. He showed that the spectral balance is an important attribute of preference unlike the perception of non linear distortion. He proposed to quantify the perceived spectral balance by the frequency response of the loudspeaker on its measured listening window (see [5] for details). A correlation of 0.95 was obtained between the perceptual and objective evaluations of spectral balance. No objective measure could be found to describe the perceived level of distortion.

Klippel [10] tested from 4 to 8 loudspeakers in a set of 7 different tests, in monophonic or stereophonic reproduction using several musical excerpts. The loudspeakers were placed at different positions in front of the listener in order to perform paired comparisons (PC) between loudspeakers to evaluate the similarities between the sounds they produced. These similarities were analyzed using a multidimensional scaling technique. Listeners also evaluated the ratio between the sounds of each pair on different perceptual scales (i.e., a listener may have judged a sound twice "clearer" than another), the results being analyzed using a factor analysis. Each test revealed 2 or 3 perceptual dimensions. Combining the 7 tests, 7 dimension were defined: "clearness", "treble stressing" ("sharpness"), "general bass emphasis" ("volume"), "low bass emphasis", "feeling of space", "clearness in basses" and "brightness". Klippel [10] defined objective attributes in order to describe these perceptual dimensions. He calculated the frequency response of the loudspeakers in a room from the measurement of their anechoic responses and a simulation of the response of the room (from reverberation time and dimensions). He used an auditory model (based on the loudness model from Zwicker *et al.* [16]) to calculate the objective attributes. Each perceptual dimension could then be significantly correlated with an objective measure.

Bramsløw [14] studied 64 different signals (speech, music with or without background noise with different processing). The goal of his work was not to study reproducing systems but sound quality in general. The signals were rated over six of the perceptual scales highlighted by Gabrielsson *et al.* [6] (loudness, clearness, sharpness, fullness, spaciousness and overall impression). Perceived sound quality could be described by 4 dimensions with 2 dominant scales (clearness/overall impression, sharpness/fullness). A neural network was built to predict the sound quality ratings. This neural network connected each quality rating to an output of an auditory model (based on roex auditory filters). This neural network remains a black box, which can be used to predict sound quality, but which cannot highlight an individual link between a perceptual dimension and a corresponding objective attribute.

In the studies mentioned above, "the acoustical measurements (spectrum, loudness model) were chosen a priori among several possibilities" [12, 13]. Lavandier *et al.* [12] compared different objective measures to quantify the dissimilarities between the signals radiated by 12 loudspeakers in monophonic reproduction. The corresponding perceptual dissimilarities were evaluated in a PC experiment. Using a MDS analysis, Lavandier *et al.* [13] revealed two perceptual dimensions underlying these dissimilarities. The temporal mean of time-varying specific loudness (based on loudness models of non-stationary sounds [17, 18]) was used to define objective attributes correlated with the perceptual dimensions: the bass/treble balance and the emergence of medium.

Summarizing the above references, two main dimensions emerge to explain the perceived differences between the sounds radiated by loudspeakers. They are mainly linked to the frequency response of the loudspeakers. The first one was emphasized in most studies [5, 6, 10, 11, 13, 14] and can be called "brightness", or "balance between bass and treble", "sharpness", "spectral balance", "bass/treble balance". The second one was found in [6, 10, 11, 13, 14] and can be called "clarity", or "clearness", "emergence of medium".

A third dimension was highlighted in one study which was linked with the spatial aspect of the reproduction and was called "feeling of space" by Klippel [10].

Most previous studies involve a limited number of loudspeakers. To give a few examples, in the study by Staffeldt [4], 5 pairs of loudspeakers were evaluated. Klippel [10] designed listening tests on different sets of 4 and 8 loudspeakers. Gabrielsson *et al.* [11, 6] conducted several experiments involving 5 to 9 loudspeakers. Gabrielsson and Lindström [7] used 20 pairs of loudspeakers tested at different positions in different sessions. However, the loudspeakers were not directly compared, as the experiment used absolute evaluations along predefined perceptual scales. These scales had been highlighted in their previous studies using up to 9 loudspeakers. To our knowledge, 12 loudspeakers involved in the same PC experiment was the maximum of what can be found in the literature [12]. Evaluating perceptual dissimilarities on larger sets of loudspeakers might potentially lead to a more homogeneous repartition of the stimuli along the perceptual dimensions. This would help to obtain a better descrip-

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

tion of the perceptual space. Moreover, more dimensions might be revealed when evaluating more stimuli as listener might use more dimensions to make their dissimilarity judgments. Furthermore, the rating of a loudspeaker on a perceptual scale can depend on the other loudspeakers tested in the same experiment [7]. Then, in order to study dimensions which characterize the perception of sound reproduction in general and not the particular sample of loudspeakers under test, studying a large set of loudspeakers is important.

The aim of the present study was then to extend the characterization of the reproduction of timbre with a larger set of loudspeakers. Thirty-seven single loudspeakers are considered in the present study, using the same experimental protocol as Lavandier *et al.*. It consists in recording the sound radiated by the loudspeakers in the room, so that loudspeakers can be compared at the same position, one just after the other with very short time intervals in between (important aspect due to our auditory memory limited to about 2 to 5 s. [19]), with a unique configuration for the listener and microphones. The recorded sounds are then submitted to both signal analysis and listening tests under headphones.

The first part of the paper details the recordings involving 37 single loudspeakers in a room. Then, the second part concerns the listening experiments leading to the perceptual dimensions underlying the perceived dissimilarities between the 37 sound excerpts. These listening tests led to a three dimensional perceptual space, adding a new dimension to the previous study [13]. In the third part of the paper, objective attributes are proposed to describe quantitatively the perceptual dimensions, including the new one. The comparison between perceptual dimensions and objective attributes leads to high correlation coefficients for all three identified dimensions. The paper ends with a discussion of the results obtained from this unusually large set of loudspeakers, which contributes to tighten the link between perceptual and physical measures.

2. The stimuli: recordings of 37 loudspeakers at the same position in a listening room

The stimuli involved in this study correspond to the recordings of 37 single loudspeakers in a listening room. Although the actual loudspeakers used for the recordings are not the focus of the present work, they play a major role by ensuring a realistic spread of the recorded stimuli over the relevant perceptual dimensions involved in reproduced musical sounds. We therefore tried to select a panel of loudspeakers with enough diversity on one hand, but homogeneous enough on the other hand. Most of these loudspeakers were lent by various manufacturers or distributors who selected items they found representative of their current range. The brand and model names may not be displayed in this paper, so the loudspeakers are hereafter described by their main features and identified by a number between 1 and 37. No reliable data about their anechoic frequency responses could be obtained, but this was not considered mandatory as the topic of the study was to compare perceptual and objective results obtained with the same sound excerpts. Provided that they are representative of a wide range of realistic sound systems, the loudspeakers themselves have little importance. Among those 37 loudspeakers, 11 had been part of the 12-loudspeaker panel tested by Lavandier *et al.* [12, 13]. The description of the 37 loudspeakers is given in Table I.

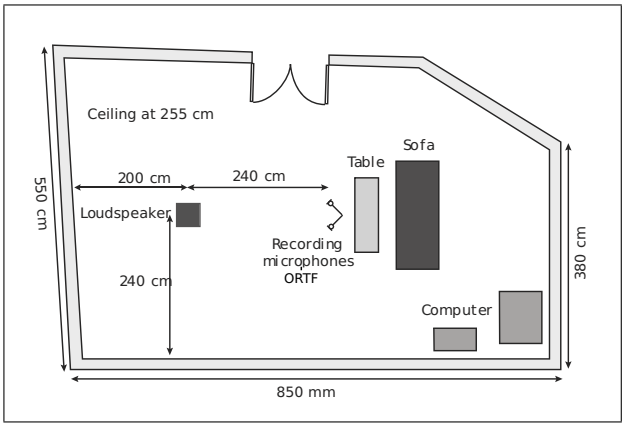


Figure 1. Recording set-up and listening room configuration.

The recordings were conducted in a room used by a loudspeaker manufacturer to test its products. This room was 8.5-m long, 5.5-m wide and 2.55-m high (Figure 1). The stereophonic ORTF recording technique was used, with two cardioid microphones placed 17 cm apart and angled 110° from each other (AKG Blue Line CK-91, SE-300B, preamplifier Tascam MX-4). The microphones were placed one meter above the floor to mimic the height of listener ears when seated. Loudspeakers were all placed at the same position in the room, 2.40 m in front of the microphones (Figure 1). Loudspeakers designed to stand directly on the floor were used so, while the vertical position of smaller units was adjusted to put the point midway between the medium and the tweeter at the microphones height.

The reverberation time was evaluated with an omnidirectional measuring microphone set between the listening position and a Genelec 1031A loudspeaker at the reference position. It was evaluated as the time required for the sound level to decrease by 60 dB after turning off a third-octave filtered pink noise (Table II). This measurement was performed after the recordings and listening tests, at a single position, just to get an overall idea of the reverberation time. The resulting values are close to the standards for loudspeaker quality evaluation [20, 21] thus confirming the realism of the resulting sound excerpts.

A single short musical excerpt — Mc Coy Tyner, "Miss Bea", 3.3 s — was used for the recordings (piano, saxophone, double bass and drums). It contains acoustical instruments rather than synthetic music and was selected among high-quality commercial recordings, according to the recommendations concerning listening tests on loud-

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

Table I. General characteristics of the 37 loudspeakers used to record the stimuli (PC stands for Personal Computer, PA for Public Address and DML for Distributed Mode Loudspeaker).

ID	Use	Box	Speakers	Filter
1	Nearfield monitor	Bookshelf, closed	1 boomer + 1 tweeter	2 ways
2	PC 2.1 System	Vented subwoofer + satellite	1 boomer + 6 wide range drivers	2 ways
3	HiFi - Satellite	Compact, closed	1 wide range	1 way
4	HiFi - Medium range	Floorstand,closed	2 boomers + 1 wide range + 1 tweeter	3 ways
5	HiFi - Medium range	Bookshelf, vented	1 boomer + 1 tweeter	2 ways
6	HiFi - High end	Floorstand,vented	1 boomers + 2 wide range + 1 tweeter	3 ways
7	HiFi - Medium range	Bookshelf, vented	1 boomer + 1 tweeter	2 ways
8	HiFi - Medium range	Bookshelf, closed	1 boomer (tweeter not connected)	1 (/2) way
9	HiFi - Medium range	Floorstand,closed	1 boomer + 1 tweeter	2 ways
10	HiFi - Medium range	Bookshelf, vented	1 boomer + 1 tweeter	2 ways
11	HiFi - Medium range	Floorstand,vented	1 boomer + 1 wide range + 1 tweeter	3 ways
12	HiFi - Medium range	Floorstand,vented	1 boomer + 1 wide range	2 ways
13	HiFi - Satellite	Compact, closed	1 wide range	1 way
14	HiFi - High end	Floorstand,vented	2 boomers + 1 wide range + 1 tweeter	3 ways
15	Miniature monitor	Compact, closed	1 wide range	1 way
16	Nearfield monitor	Bookshelf, vented	1 boomer + tweeter	2 ways
17	Nearfield monitor	Bookshelf, vented	1 boomer + tweeter	2 ways
18	Nearfield monitor	Bookshelf, vented	1 boomer + tweeter	2 ways
19	Studio monitor (80's)	Bookshelf, vented	1 boomer + 1 medium + 1 tweeter	3 ways
20	Studio monitor (80's)	Bookshelf, vented	1 boomer + 1 medium + 1 tweeter	3 ways
21	HiFi - High end	Floorstand,vented	2 boomers + 1 wide range + 1 tweeter	4 ways
22	HiFi - Medium range	Bookshelf, closed	1 boomer + tweeter	2 ways
23	Home-Studio	Bookshelf, vented	1 boomer + tweeter (horn)	2 ways
24	HiFi - High end	Floorstand,closed	1 woofer + 1 electrostatic panel	2 ways
25	HiFi - Medium range	Bookshelf, vented	1 boomer + tweeter	2 ways
26	HiFi - Medium range	Floorstand,closed	1 boomer + 1 wide range + 1 tweeter	3 ways
27	HiFi - High end	Floorstand,vented	2 boomers + 1 wide range + 1 tweeter	4 ways
28	P.A. - Medium range	Compact, vented	1 boomer + 1 wide range	2 ways
29	HiFi - Budget	Bookshelf, closed	1 boomer + tweeter	2 ways
30	Nearfield monitor	Bookshelf, vented	1 coax (boomer+tweeter)	2 ways
31	Nearfield monitor	Bookshelf, vented	1 coax (boomer+tweeter)	2 ways
32	HiFi - Medium range	Bookshelf, vented	1 boomer + tweeter	2 ways
33	HiFi - Medium range	Floorstand,vented	1 boomer + tweeter	2 ways
34	HiFi - High end	Bookshelf, vented	1 coax (boomer+tweeter)	2 ways
35	HiFi - Medium range	Vented subwoofer + satellite	1 woofer + 1 DML panel	2 ways
36	Miniature monitor	Compact, vented	1 wide range	1 way
37	HiFi - High end	Floorstand,vented	1 wide range	1 way

speakers [20, 21]. The monaural excerpt was obtained by considering only the right channel of the original stereophonic signal. This musical excerpt was among the three tested by Lavandier *et al.* [12, 13], who showed that listeners used the same perceptual dimensions to discriminate the loudspeakers with different musical excerpts.

Recordings were performed at reproduction levels within the linear dynamic range of each loudspeaker. The recording level was therefore different for each unit. To avoid a bias related to the listening level, the overall loudness of the recordings was subjectively equalized by the experimenters before the listening tests and validated using a loudness estimation model [17]. The reproduction level was then set to a level considered as comfortable, around 65 phons. This level equalization indeed led to different signal-to-noise ratio between sound excerpts, but the background noise of the recordings was very low and did not appear as a noticeable artifact during the listening tests. These listening tests were run using Stax SR Lambda Professional headphones in an isolated soundproof room.

3. Perceptual experiments

3.1. Free sorting

3.1.1. Procedure

The 37 recordings were presented as 37 crosses randomly distributed on a computer screen. Listeners could move each cross freely and listen to the corresponding signal as often as necessary by clicking on the cross. They were asked to group the recordings based on similarity: recordings found similar should be placed in the same class, whereas recordings found dissimilar should be placed in different classes. Listeners were specifically instructed to avoid grouping based on preference as "good" or "bad", in order to try make listeners sort the recordings mainly from their similarity. No anchor-point or repetition were

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

Table II. Reverberation time (TR60) in each third-octave band (center-frequency fc).

fc (Hz)	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
TR60 (s)	0.39	0.40	0.31	0.38	0.43	0.38	0.35	0.41	0.41	0.49	0.44

involved as the evaluation of 37 stimuli is quite demanding. However two loudspeakers of the same model (19-20) were involved in the experiment to test the consistency of the perceptual evaluation.

The number of classes to be used was free. The degree of similarity between two recordings for them to be placed in the same class was therefore left to the judgment of each listener. The partition of each listener was then transformed into an individual binary matrix: the coefficient for two recordings belonging to the same class was set to zero, while for two recordings belonging to different classes it was set to one. The final dissimilarity matrix was obtained by averaging the individual binary matrices, leading to an indirect estimation of the dissimilarity between stimuli: two recordings often grouped together by listeners are supposed to be similar, and lead to a lower value of their averaged coefficient. It took the listeners between ten and eighty minutes to complete the task, with an average of twenty-eight minutes per listener.

3.1.2. Listeners

Fifty-six listeners took part in the experiment, including twenty-two women and thirty-four men. Thirty-two had their hearing tested as normal. Among the remaining 24 listeners, eighteen were less than 31 years old and reported normal hearing. Listeners were between twenty-two and fifty-three years old, with an average of thirty-two years. They were members of the laboratory or students, and some of them had some musical and sound listening experience. Even with this listening experience, they were considered as untrained for loudspeakers comparison.

3.1.3. Results

Before averaging the perceptual dissimilarities over the panel of listeners, it was verified that this panel was not composed of groups with different strategies (which would have required a by-group analysis). A cluster analysis of the between-listener Rand indices [22] was conducted. There would have been groups among listeners if their partitions were similar (high Rand index) within each group but also different (low Rand index) across groups. Such groups could not be isolated while considering the cluster analysis. There was no clear difference of strategy between listeners, who might have simply used different degrees of similarity between the recordings to define their partitions. The perceptual dissimilarities were then averaged across listeners.

These mean perceptual dissimilarities were analyzed using a metric multidimensional scaling analysis (MDS). Because the free sorting (FS) test does not provide individual dissimilarity matrices. we processed only the average one, using the classical MDSCAL model. The SMACOF algorithm ("scaling by majorizing a complicated function") has been detailed by Borg and Groenen [23]. Stimuli are represented in an Euclidean space of N orthogonal dimensions, the most appropriate value of N being chosen by the experimenter relying on stress majorization without weighting. This space can be rotated or dilated; only relative distances between stimuli are fixed. To get a unique description, the axes are selected to coincide with the directions of maximum variance, determined by a principal component analysis, and the ranking of dimensions is fixed by the amount of variance explained by each of them.

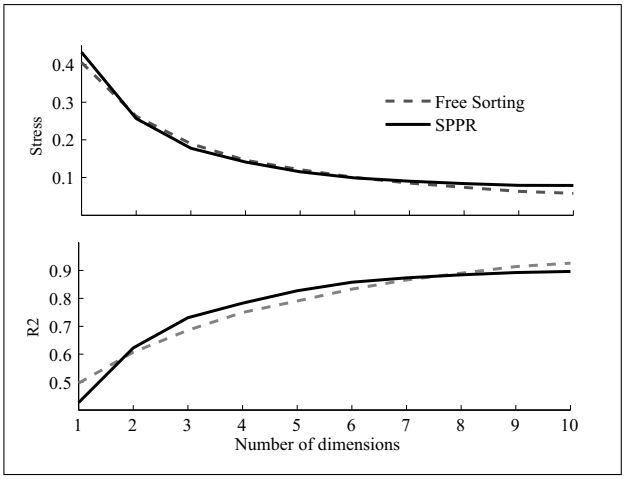


Figure 2. Stress and correlation index plots for FS and SPPR experiments.

The stress function is minimized iteratively for a given number of dimensions. Its mathematical definition is given in Chapter 8 of reference [23]. The R2 index indicates the proportion of variance explained for this number of dimensions. This index is a complementary information about the appropriateness of the MDS model. Obtaining a R2 above 0.6 is considered as an acceptable score [24]. Figure 2 presents the stress and correlation index as a function of the number of dimensions for the MDS analysis of the FS experiment (dashed lines). For this analysis, three dimensions were retained for a good description of the data, based on R2 curve (see Figure 2): the fourth dimension would provide only a marginal increase of R2. Also, when listening along the dimensions, only three of them were perceived as relevant and could be associated with an acoustical attribute (see Section 4).

Figure 3 presents the corresponding perceptual MDS space (each number corresponds to a recording / loudspeaker, see Table I), which highlights the three main perceptual dimensions which have been used by the listeners to differentiate the thirty-seven recordings.

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

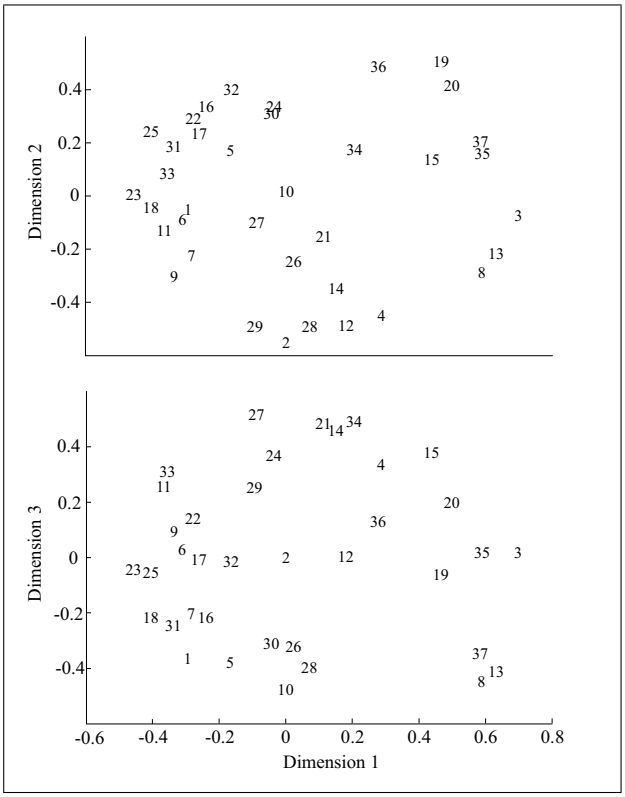


Figure 3. Three-dimensional space resulting from the MDS analysis of the dissimilarities obtained with the FS method.

3.1.4. Preliminary discussion

Listening to recordings while considering their position in the MDS space, a first qualitative interpretation of these three dimensions could be proposed by the experimenters. This interpretation will be verified/consolidated using objective attributes (Section 4).

As in reference [13], the predominant (first) dimension seemed strongly linked to the bass/treble balance of the recordings. This first dimension thus corresponds to previous dimensions highlighted in the literature: "balance between bass and treble" or "brightness-darkness" [11, 6], "amount of low frequencies compared to medium and high frequencies" [1], "emphasized treble" and "emphasized bass" [4] or "ratio between low and high frequencies", an interpretation by Bramsløw of the sharpness and fullness scales [14].

The second dimension apparently linked to "reverberance" could result from the interaction between loudspeakers and room. It seems similar to the "feeling of space" [6, 10] or "spaciousness" [14]. This result might contradict Staffeldt statement saying that "the room impression and width of the stereophonic sound image are of less importance to the subjective evaluation of the loudspeakers" [4]. It should however be noted that Staffeldt was dealing with preference, not similarity. Moreover, he was studying stereophonic reproduction and he referred to the width of the sound image, not a feeling of space in a monophonic reproduction. Our result is also somewhat in contradiction with the listening tests of Olive [5] that showed that 94% of the comments of his listeners were linked to timbre, whereas only 3% seemed characteristic of the spatial component of sound reproduction. The spatial dimension was highlighted in these studies, but with a lower importance than in the present experiment. However, the methods used to extract the dimensions are different and the results can only be compared qualitatively.

The third dimension obtained here seems related to the level of medium frequencies or a feeling of clarity, and seems similar to the second dimension revealed in [13]. This third dimension could correspond to the "distinctness" [11, 6] or "clearness" [14, 10].

This first experiment raises questions concerning the exact meaning of the dimension 1 obtained with the FS experiment. When ranking recordings with increasing coordinates on dimension 1 for the 37 loudspeakers, we find "lack of bass" recordings, then "boost of bass" recordings and finally "well-balanced" recordings; whereas along the dimension 1 in [13], the recordings appeared with a "lack of bass", then "well-balanced" and finally with a "boost of bass". In both experiments (12 and 37 loudspeakers), the first dimension is linked to the bass/treble balance, but when evaluating 37 loudspeakers, it does not seem monotonic with the coordinate identified for 12 loudspeakers. In order to investigate the character of the dimension 1, another listening test was thus conducted using a different method, the Similarity Picking with Permutation of References method [25].

3.2. Similarity Picking with Permutation of References

3.2.1. Procedure

The SPPR method [25] is an indirect dissimilarity evaluation method which consists in presenting to the listener four sounds on each trial. One of the sounds is considered as the reference stimulus for the current trial and the other three are comparison stimuli. The listener is asked to select among the comparison stimuli which one is the most similar to the current reference. Each stimulus becomes a reference during successive trials.

For a SPPR listening test involving a set of 37 stimuli, every stimulus is used as a reference and is compared to the other 36. Those 36 stimuli are randomly distributed in 12 triplets without replacement. As every stimulus is used as a reference, the entire test is composed of a total of 444 trials (12 triplets x 37 reference sounds). The answers of a listener are used to fill an individual binary matrix. Before the test, each value of this individual binary matrix is set to one, except for the diagonal values which are set to zero. When a stimulus j is judged within the comparison stimuli as the most similar to the reference i, the ij-th element of the individual binary matrix is reset to zero. The way each individual matrix is filled and the answers given by each listener thus lead to sparse individual binary matrices. A global dissimilarity matrix is then estimated by averaging individual binary matrices. For further details about the SPPR method, the reader is referred to [25].

During the test, listeners were presented with a graphical user interface comprising four square buttons corre-

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

sponding to the four stimuli (reference + three comparison sounds). First, these four stimuli were successively played. Then the listeners could listen to any of the stimuli, as many times as needed, by clicking on the corresponding button. The trial ended when the listener had choosen which comparison stimulus seemed most similar to the reference. The entire test was divided into 5 sessions with a rate of one session per day during 5 days. In average, the listeners spent between 30 to 40 minutes to complete each session. The total average duration for the entire SPPR test was about 3 hours.

3.2.2. Listeners

Twenty-seven participants (five women and twenty-two men) took part in the listening test. Audiometric measurements were performed in order to guarantee the normal hearing of the listeners. They were members of the laboratory or students. Four of them had participated in the FS listening test and some of the other listeners had a musical/sound listening experience but they were considered as untrained to loudspeakers comparison.

3.2.3. Results

The mean dissimilarity matrix was analyzed with the same metric MDS technique used for the FS experiment. This MDS analysis led to a three-dimensional perceptual space. Figure 2 shows the stress plot and correlation index (continuous lines) as a function of the number of dimensions. Figure 4 presents the perceptual three-dimensional space retained from the analysis. As 11 of the 37 loudspeakers were previously evaluated by Lavandier *et al.* in a 12-loudspeaker experiment [13], their corresponding position in the MDS space resulting from the 12-loudspeaker experiment using the SPPR method [25] is also plotted (grey underlined symbols). The comparison is discussed in the next section.

Similarly to the FS experiment, a rough qualitative interpretation of the dimensions could be proposed from informal listenings (they will be verified from the objective attributes in Section 4). The first dimension seems linked to the bass/treble balance of the recordings. Contrary to what was observed in the FS experiment, when ranking the recordings with increasing first coordinate, the recordings present a "lack of bass", then they are "well-balanced" and finally they present a "boost of bass". The recordings along the second dimension vary with the amount of medium frequencies, which could be associated with the notion of sound clarity. The third dimension seems linked to "reverberance" or "feeling of space".

3.3. Comparison between FS and SPPR experiments

When evaluating the set of 37 loudspeakers, three dimensions were revealed with both the FS and the SPPR experiments. In both experiments the dimension corresponding to the bass/treble balance appeared as the predominant one. The second and third dimensions appeared linked to similar aspects, but with an inverse significance rank (the MDS technique used in this work provides a multidimensional space with dimensions ordered from the most significant to the least one). The task that listeners had to conduct seems to have influenced the relative importance of the dimensions used to differentiate the recordings.

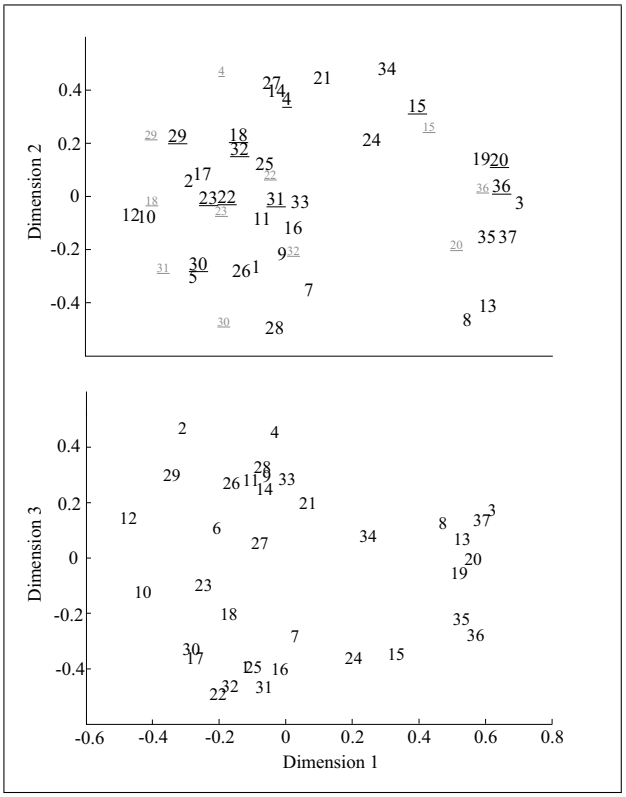


Figure 4. Three-dimensional space resulting from the MDS analysis of the dissimilarities obtained with the SPPR method. The black underligned numbers correspond to the 11 loudspeakers previously evaluated by Lavandier *et al.* [13], their position in the MSD space resulting from the 12-loudspeaker experiment using the SPPR method [25] are reported here for comparion (grey underlined numbers).

In the FS experiment, the validity of the dissimilarity evaluation was questioned because of the behavior of the dimension associated with the bass/treble content. The rank of the recordings along this dimension is different from an experiment to the other. For the SPPR experiment, this dimension appears monotonic with the bass/treble balance, as in previous studies [13, 25]. The dimension obtained with the FS method could be characteristic of the "degree of discoloration" of the sound reproduction [10], if too much bass could be assumed to be perceived more "well-balanced" than a reproduction lacking of bass. Instead of considering only similarity during the FS experiment, listeners might thus have used notions of quality or preference to group the recordings: reproductions lacking of bass would have been considered as bad ones, whereas the ones with too much bass would have been of intermediate quality.

Loudspeakers 8 and 13, two very band-limited loudspeakers, might have favored this bias: they correspond to a reproduction of poor quality and their recordings are the most often grouped together, whereas they do not seem to be the most similar ones. They could have been grouped because they were both very different from the others

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

rather than similar to each other. These loudspeakers might have impaired the homogeneity of our set of stimuli and this might have biased the dissimilarity evaluation using the FS task [26]. As this task offers more freedom to listeners, it is more difficult for the experimenter to control what is really under evaluation, possibly biasing the dissimilarity evaluation. The forced choice task used in the SPPR experiment may avoid this drawback and lead to a better dissimilarity evaluation. This hypothesis receives further support when considering the two loudspeakers of the same model (19-20) in the perceptual spaces: these loudspeakers are almost collocated in the SPPR space (Figure 4), whereas they are further apart in the FS space (Figure 3). The suitability of the FS task to evaluate correctly the dissimilarity for our set of stimuli seems then questionable. More experimental data and listening task comparisons would however be required to give a definite answer to this question.

In the present study, the first and second dimensions obtained with the SPPR method for 37 loudspeakers are similar to those obtained with the PC [13] and SPPR [25] methods for 12 loudspeakers. Eleven loudspeakers employed in the present study were part of the loudspeakers tested in [25]. Figure 4 allows for a visual comparison of the relative positions along the first and second dimensions obtained in [25] and in the present study. The correlation between the coordinates resulting from both tests is 0.90 for the first dimension and 0.76 for the second dimension. This result indicates that the two dimensions obtained with 37 loudspeakers are similar to those obtained with 12 loudspeakers even if the recordings were done in different rooms. This indication is verified using the objective attributes described in the following section.

The present data pushed further the investigation of the reliability of the SPPR method, after its validation with a test involving 12 stimuli [25]. Even if there is no reference method with which SPPR could be compared for large sets of stimuli (the PC method can hardly be run on such sets), the nature of the dimensions highlighted and their compatibility with previous results using the PC method suggest that SPPR is suitable for the evaluation of dissimilarity in such cases, as hypothesized by Michaud *et al.* [25]. The advantage of the SPPR method compared to the PC method is the presence of a reference stimulus during each trial which allows the experimenter to divide any long test into several sessions. Conversely our SPPR experiment was more time consuming for each participant than the FS experiment. The next section is dedicated to the objective characterization of the perceptual dimensions. As the perceptual dimensions revealed with the SPPR method were considered as more reliable, these dimensions were used for the elaboration of our acoustical attributes.

4. Acoustical attributes

In order to consolidate the interpretation of the perceptual dimensions, objective attributes were calculated directly on the recordings used for the listening tests. Our aim was to provide acoustical attributes perceptually relevant which could quantify the dimensions revealed by these tests.

In the present study the first two perceptual dimensions corresponded well to those reported in [13] and therefore they might be properly described by the same acoustical attributes. These attributes were based on an auditory model designed to calculate the loudness of non-stationary sounds [17, 18] taking into account time-frequency masking. For the third perceptual dimension interpreted as "feeling of space" or "reverberance", a measure associated with the interaural characteristics of the stereophonic recordings was tested.

4.1. "Bass/treble balance" dimension

The first acoustical attribute, which is referred to as BassTreb, is based on a ratio calculated over approximated critical bands in order to quantify the balance between low and high frequencies. As in [13], the ratio was computed using the specific loudness taken within the bands 1 to 3 corresponding to the frequency range [20-280 Hz] and the specific loudness taken within the bands 12 to 21 corresponding to the frequency range [1.8–15.5 kHz]. The equation expressing the BassTreb attribute is given in Equation (1) with b the approximated critical band number and $\mathrm{Dens_m}$ the temporal mean of time-varying specific loudness,

$$\mathrm{BassTreb} = \frac{\sum_{b=1}^{3} \mathrm{Dens_m}(b)}{\sum_{b=12}^{21} \mathrm{Dens_m}(b)}. \quad (1)$$

The BassTreb attribute was calculated separately for the left and right channels of every recording. The Pearson correlation coefficient was then computed between each (left and right) BassTreb values and the corresponding coordinates of the recordings along the first perceptual dimension. Table III presents these correlation values obtained with the SPPR and FS experiments from the present study (involving 37 loudspeakers) and, as a comparison, for 12 loudspeakers with the PC [13] and the SPPR [25] methods[1]

For the set of 12 loudspeakers, the correlation between the BassTreb attribute and the perceptual dimension is above 0.80 for both PC and SPPR methods. For the set of 37 loudspeakers, the SPPR experiment leads to a correlation above 0.85 whereas the BassTreb attribute is less correlated with the perceptual dimension resulting from the FS experiment (around 0.65). Indeed, in this last case the relation between the perceptual dimension and the attribute appears to be non-monotonic as discussed previously (Section 3.3). In conclusion, for both sets of recordings, the perceptual dimension obtained with the SPPR method is well described by the BassTreb attribute, in very good agreement with [13].

[1] The correlations presented in Tables III and IV were not published in [13], as in their study, correlations were calculated between the attribute and the coordinates along the dimension of a "physical" space obtained from objective metrics.

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

Table III. Correlation between the BassTreb attribute (calculated for the left (L) and right (R) channels of the recordings) and the coordinates along the first perceptual dimension for the set of 12 [13, 25] and 37 loudspeakers and for each method used for the dissimilarity evaluation. A star indicates a significant correlation ($p<0.05$).

	12 loudspeakers		37 loudspeakers	
	PC	SPPR		FS
(L)	0.82*	0.82*	0.86*	0.64*
(R)	0.85*	0.86*	0.87*	0.65*

Table IV. Correlation between the Med1 and Med2 attributes (calculated for the left (L) and right (R) channels of the recordings) and the coordinates along the perceptual dimension associated with the medium emergence for the set of 12 [13, 25] and 37 loudspeakers and for each method used for the dissimilarity evaluation. A star indicates a significant correlation ($p<0.05$).

		12 loudspeakers		37 loudspeakers	
		PC	SPPR		FS
Med1	(L)	0.88*	0.85*	0.52*	0.58*
	(R)	0.91*	0.89*	0.78*	0.68*
Med2	(L)	0.75*	0.84*	0.77*	0.69*
	(R)	0.80*	0.86*	0.83*	0.77*

4.2. "Emergence of medium" dimension

The emergence of medium could be linked to a notion of sound clarity. The clarity C80 widely used to objectively describe sound perception in rooms [27] might however correspond to a different notion/perceptual dimension: room and loudspeaker evaluations are not necessarily equivalent. Moreover C80 is based on a temporal analysis of the room impulse response. In the present study, objective attributes are computed directly from recorded musical signals, so that C80 does not seem appropriate. The corresponding "emergence of medium" dimension is then described using the attribute proposed in [13]. Like BassTreb this attribute, which is referred to as Med1, is based on the temporal mean of time-varying specific loudness of the recorded signals. This attribute is obtained by integrating the specific loudness over the approximated critical bands 5 to 9 (corresponding to the frequency range [355–1120 Hz]). Its expression is given by Equation (2) with b the approximated critical band number and $\mathrm{Dens_m}$ the temporal mean of time-varying specific loudness,

$$\mathrm{Med1} = \sum_{b=5}^{9} \mathrm{Dens_m}(b). \qquad (2)$$

The Med1 attribute calculated separately for the left and right channels of the 37 recordings was compared to the coordinates along the perceptual dimension associated with the "emergence of medium". The Pearson correlation coefficients between this attribute and the coordinate along the perceptual dimension are presented in Table IV. One should recall that, for the 37-loudspeaker experiments, this dimension linked to sound clarity appeared as the second dimension in the SPPR experiment and as the third dimension in the FS experiment. For the 12-loudspeaker experiments, the attribute was compared to the second dimension obtained both with the PC method [13] and the SPPR method [25] [1].

The attribute Med1 leads to correlations above 0.85 with the perceptual dimension obtained for the set of 12 loudspeakers with both the PC and SPPR methods. However, this attribute is less correlated with the coordinates of the perceptual dimension obtained for the 37 loudspeakers. In order to be of general use, an acoustical attribute should not depend on the particular set of loudspeakers involved in the test [5, 6]. The definition of the attribute expressing the "emergence of medium" was then refined to obtain better correlations with the perceptual dimension obtained with both set of loudspeakers (12 and 37).

The second attribute, which is referred to as Med2, was defined as the ratio of the combination of low and high frequencies over medium frequencies. The low and high frequencies corresponded to the bands 1 to 3 [20–280 Hz] and 12 to 21 [1.8-15.5 kHz]. The medium frequency range was taken over the bands 4 to 11 corresponding to the range [280–1800 Hz]. The expression of Med2 is given in Equation (3) with b the approximated critical band number and $\mathrm{Dens_m}$ the temporal mean of time-varying specific loudness. The inverse of Med2 was tested but did not lead to good correlations with the perceptual dimension,

$$\mathrm{Med2} = \frac{\sum_{b=1}^{3} \mathrm{Dens_m}(b) + \sum_{b=12}^{21} \mathrm{Dens_m}(b)}{\sum_{b=4}^{11} \mathrm{Dens_m}(b)}. \qquad (3)$$

Compared to the frequency range used for Med1, the medium frequency range chosen for the attribute Med2 was slightly extended. In Med2, the "medium" frequency range now complements the frequency domain involved in the calculation of the "bass" and "treble" loudnesses (also used for the BassTreb attribute), by which the "medium" loudness is normalized. The correlation between Med2 values and the coordinates of the perceptual dimension for each set of loudspeakers and each evaluation method is presented in Table IV.

Compared to the attribute Med1, the attribute Med2 leads to slightly lower correlations for the set of 12 loudspeakers. However, the correlations for the 37 loudspeakers increase significantly with this redefined attribute for both the SPPR experiment and the FS experiment. With the attribute Med2, the correlations remain quite constant among the two sets of loudspeakers: apart from one lower correlation obtained with the FS experiment, every other correlation exceeds 0.75.

4.3. "Feeling of space" dimension

The dimension "feeling of space" or "reverberance" might be explained by the interaction of the single loudspeakers with the room. The sound reaching the listener's ears (the recording microphones here) will contain a direct sound and reflections on the room surfaces. These reflections are

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

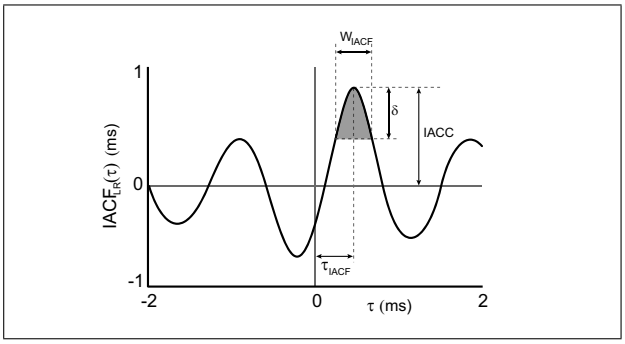


Figure 5. InterAural Cross-correlation Function (IACF) and schematic definitions of IACC, W_{IACF} and τ_{IACF}.

delayed and filtered versions of the original source signal that are influenced by the directivity pattern of each loudspeaker, and are different at the two ears - leading to dissimilarity between the left and right signals.

In order to provide an attribute describing the "feeling of space" based on the stimuli involved in the listening tests, we defined an attribute based on the stereophonic recordings by calculating the InterAural Cross-correlation Function (IACF). This function measures the similarity between the signals reaching the left and right ears of a listener in a sound field. Figure 5 presents three parameters usually extracted from the IACF: the maximum value of the peak (IACC), the width of the peak (W_{IACF}) and the interaural delay (τ_{IACF}) corresponding to this peak. For W_{IACF}, the width of the peak is measured for a given δ value. In this study, the IACF is computed for temporal delays within the range -2 and 2 ms between the two channels of the stereophonic signals.

None of these three usual parameters of the plain IACF computed from our recordings is significantly correlated with the coordinates of the "feeling of space" dimension obtained with either the SPPR or the FS method. These correlations are improved when the IACF is computed after low-pass filtering the sound samples in the frequency range within which interaural time delay is a predominant cue for localization (<1500 Hz) [28] and especially in the lower frequency range considered as important by several authors to describe the perceptual dimension "spaciousness" (<500 Hz) [29, 30]. This lower frequency range is also considered as critical by some authors investigating the coupling of woofers with rooms [31]. Widening the frequency range to 1500 Hz reduced correlation values. With a 500-Hz low-pass cut-off frequency the highest correlation is 0.67, obtained when considering the W_{IACF} with a δ taken 10% lower than the IACC as proposed by Ando *et al.* [32]. These authors showed that the "Apparent Source Width" (ASW) was correlated to IACC and W_{IACF}. They also proposed a formula combining both IACC and W_{IACF} as an attribute describing "spaciousness". This combined attribute did not however improve much the correlation with our test results.

Expanding over the ASW definition, we propose to compute the area (A_{IACF}) under the IACF peak down to its half height. This corresponds to a combination of the IACF peak height (IACC) and its width at half maximum (W_{IACF} with δ set to 50%) but keeping more information about the actual shape of the peak. The resulting attribute A_{IACF} is displayed on Figure 5 as the shaded area below the IACF. It is still computed after low-pass filtering the recordings below 500 Hz. The correlation between these A_{IACF} values and the "feeling of space" dimension reach 0.79 for the SPPR experiment, but only 0.59 for the FS experiment. These are the highest correlation values we could obtain for the objective attributes tested in this study.

5. Discussion and conclusion

For the evaluation of 12 single loudspeakers in a room, the Similarity Picking with Permutation of References experiment [25] and the Paired Comparison experiment [13] led to a two-dimensional perceptual space. These two dimensions were linked to the "bass/treble balance" and the "emergence of medium" frequency bands, which were also highlighted by other researchers [11, 6, 10, 14]. One important aspect of the present study was to extend the perceptual evaluation to 37 single loudspeakers.

This extension to a larger set of loudspeakers raised methodological issues. Indeed, the relationship between the BassTreb attribute and the coordinates of the loudspeakers on the corresponding perceptual dimension revealed by the FS method is not monotonic whereas this relationship is monotonic when the SPPR method is used. Therefore, the validity of the dissimilarity evaluation with the FS task is questionable and the SPPR task appears more relevant to evaluate the dissimilarities among the large set of sounds involved in the present study. Moreover, evaluating a larger set of 37 loudspeakers with the FS and the SPPR methods allowed us to highlight three dimensions. The two dimensions found for the set of 12 loudspeakers appear among the three dimensions found for the set of 37 loudspeakers. The description of these two dimensions is therefore more reliable as more stimuli are spread along the dimensions. Furthermore, compared to the experiment involving 12 loudspeakers, the present study investigated the reproduction of 37 loudspeakers in a different room with the same musical excerpt, reinforcing the validity of these two dimensions. The third dimension revealed in the present study seems linked to the interaction between the room and the single loudspeakers - an unexpected significance ranking given the fact that our protocol deals with loudspeakers in monophonic reproduction.

Acoustical attributes have been defined in order to describe the dimensions revealed by our listening tests. Concerning the "bass/treble balance" and "emergence of medium" dimensions, Gabrielsson *et al.* [8] and Bech [15] proposed qualitative attributes based on visual observations of the frequency response of loudspeakers. Gabrielsson *et al.*[8] described their dimension "brightness versus dullness" with the level of bass and treble. Bech [15] stated that the frequency range between 500 Hz and 2 kHz is important for changes in the restitution of timbre. Gabriels-

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

son *et al.* [8] linked their dimension "clarity" to an amplification of medium - between 500 and 4000 Hz. In the study by Klippel [10], the attribute linked with his perceptual scale "brightness" (corresponding to our bass/treble balance dimension) was derived from the sharpness defined by von Bismark [33]. The attribute corresponding to "clearness" was calculated as a degree of discolouration of the signal from a specific loudness representation. Bramsløw [14] used time-varying specific loudness to define objective attributes predicting two dimensions ("clearness" and "sharpness"). But as Bramsløw used a neural network to predict sound quality, he did not link each perceptual dimension with an objective attribute.

In the present study, the attributes corresponding to the bass/treble balance and the emergence of medium dimensions are adapted from Lavandier *et al.* [13]. These two attributes are based on the computation of time-varying specic loudness as it has been shown that such a loudness model improves the correlation between perceptual and objective dissimilarities [12]. The current work thus upgrades and consolidates the objective characterization already provided by previous work.

The third perceptual dimension highlighted in the present study is linked to the "feeling of space" or "spaciousness". It has been shown that the loudspeaker position influences the timbre of reproduced sound [2, 34] and preference ratings [3]. The dimension "feeling of space" highlighted here and in different studies [6, 10, 13] refers mainly to the sound image as showed in Olive's study [5]. Indeed, this dimension was extracted from a principal component analysis on listener comments and linked with verbal expressions such as: "narrow image", "distant", "image is not focused", which does not refer to timbre. In the studies in which loudspeaker position varied in the room [6, 7, 10, 11], the effect of position might have been included in the dimensions related to timbre. As pointed out in [35, 36], perceptual characterization of sound reproduction has mainly been concerned with perceived timbral modifications with loudspeakers in monophonic reproduction. The feeling of space has been deeply treated in room acoustics and most of the literature on spaciousness is focused on understanding the feeling of space in concert halls. Griesinger [37] points out the difference between music reproduction in small rooms and music played in concert halls. For small rooms, most of the reflected sound arrives sooner than 50 ms after the direct one - which does not create a perception of surround. In this context, the spatial impression called "Early Spatial Impression" by Griesinger [37] is principally frontal. The feeling of space or spaciousness is divided into two forms [37]. The first is associated with the "envelopment" or "Listener EnVelopment" (LEV) which expresses the general impression of being surrounded by the music [38]. The second is linked to the "Apparent Source Width" (ASW) also called "Auditory Source Width" which expresses the perceived source broadening [39, 32, 40]. The ASW is directly linked to the interaction of the sound source with the listening environment. This interaction leads to the perception of a wide reproduced sound scene or a very frontally-localized scene. Bradley and Soulodre [38] have shown that ASW is principally related to the early lateral reflections. The third dimension resulting from our listening test is very close to the description of the ASW. Morevover, the attribute proposed in this paper to describe the spatial dimension is adapted from an ASW estimation given in [32]. The fact that the attribute correlated to the perceptual dimension is computed on the signals below 500 Hz indicates that listeners might be more sensitive to the interaction of the loudspeakers and the room within this frequency range. Indeed, in this frequency range, the modal response of small rooms is important and can modify the sound reproduced by loudspeakers [31]. The low frequencies are considered to be more important in determining the ASW than middle and high frequencies [30]. This result is in agreement with the study of Griesinger [29] who found that the low frequencies under 500 Hz are important for the perception of spaciousness. Morimoto [41] worked on the effects of low frequencies on auditory spaciousness and also found that the feeling of space increased for low frequencies below 510 Hz.

The observations and the attribute chosen to describe the spatial dimension are in good agreement with the remarks made by Toole [42] (p.110): "Considering only acoustical interaction between these frontal-located loudspeakers and the adjacent room boundaries, it is reasonable to assume that full-scale envelopment is not possible with conventional stereo and multichannel mixes. So in practical terms, we are talking about perceptions of image size, ASW, image broadening, impression of height, distance, depth, and perhaps, some early spatial impression in the frontal hemisphere". Moreover, Toole defines the perceived width as "the perception that a sound source is wider than the physical extent of the source". Toole [42] (p. 99) also mentions concerning the spatial impression that it "can be a broadening of the sound from a single loudspeaker or the broadening of a sound stage created by a stereo pair, or by a multichannel trio of left, center and right front loudspeakers". A monophonic reproduction can indeed lead to the feeling of space.

The protocol used in the present study was not designed to investigate spatial attributes or loudspeaker/room interaction. We were interested in the restitution of timbre by loudspeakers in monophonic reproduction. Even with such recordings and listening tests conducted under headphones, the feeling of space or the perception of the interaction with the room appeared as an important perceptual dimension used by listeners to discriminate the loudspeakers. We anticipate that it should be even more important in stereo or multichannel reproduction, but this is beyond the scope of the present work. Future work should assess the validity of our objective attributes concerning the reproduction of timbre by single loudpseakers, which involved a large set of stimuli and a fair number of listeners, but were obtained with a single audio excerpt. This study could also be extended to more complex reproduction scenarii in order to describe the spatial dimension with an appropriate

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

protocol as proposed in [35, 36]. Following the same multidimensional approach, future studies might help in consolidating or upgrading the proposed preliminary objective attribute describing the feeling of space associated with reproduced sound in a room.

Acknowledgments

The authors wish to thank Jeremy Marozeau for providing his program running the FS experiment, Philippe Penna and the Mosquito Group for lending their listening room, Mickael Lefebvre for his help during recording sessions, the manufacturers, audio professionals and laboratories who lent us their loudspeakers (BC Acoustique, Cabasse, Conservatoire National des Arts et Métiers, Copper et Cobalt, France Telecom R&D, Genesis, Mosquito Group, Relief Sonore, Supravox), and finally all listeners for participating in the experiments.

References

[1] F. Toole: Subjective measurements of loudspeaker : sound quality and listener performance. J. Audio Eng. Soc. **33** (1985) 2–32.

[2] S. Bech: Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position. J. Audio Eng. Soc. **42** (1994) 999–1007.

[3] S. Olive, P. Schuck, S. Sally, M. Bonneville: The effect of loudspeaker placement on listener preference ratings. J. Audio Eng. Soc. **42** (1994) 651–669.

[4] H. Staffeldt: Correlation between subjective and objective data for quality loudspeakers. J. Audio Eng. Soc. **22** (july/august 1974) 402–415.

[5] S. Olive: A multiple regression model for predicting loudspeaker preference using objective measurements: Part 1 - Listening test results. AES 116th Convention, may 2004.

[6] A. Gabrielsson, H. Sjögren: Perceived sound quality of sound-reproducing systems. J. Acoust. Soc. Am. **65** (1979) 1019–1033.

[7] A. Gabrielsson, B. Lindström: Perceived sound quality of high-fidelity loudspeakers. J. Audio Eng. Soc. **33** (1985) 33–53.

[8] A. Gabrielsson, B. Hagerman, T. Bech-Kristensen, G. Lundberg: Perceived sound quality of reproductions with different frequency responses and sound level. J. Acoust. Soc. Am. **88** (1990) 1359–1366.

[9] A. Gabrielsson, B. Lindström, O. Till: Loudspeaker frequency response and perceived sound quality. J. Acoust. Soc. Am. **90** (1991) 707–719.

[10] W. Klippel: Multidimensional relationship between subjective listening impression and objective loudspeaker parameters. Acustica **70** (march 1990) 45–54.

[11] A. Gabrielsson, U. Rosenberg, H. Sjögren: Judgments and dimension analyses of perceived sound quality of sound-reproducing systems. J. Acoust. Soc. Am. **55** (1974) 854–861.

[12] M. Lavandier, P. Herzog, S. Meunier: Comparative measurements of loudspeakers in a listening situation. J. Acoust. Soc. Am. **123** (2008) 77–87.

[13] M. Lavandier, S. Meunier, P. Herzog: Identification of some perceptual dimensions underlying loudspeaker dissimilarities. J. Acoust. Soc. Am. **123** (2008) 4186–4198.

[14] L. Bramsløw: An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing. Acta Acustica united with Acustica **90** (november/december 2004) 1007–1018.

[15] S. Bech: Timbral aspects of reproduced sound in small rooms. 2. J. Acoust. Soc. Am. **99** (june 1996) 3539–3549.

[16] E. Zwicker, H. Fastl, C. Dallmayr: Basic-programme for calculating the loudness of sounds from their 1/3-oct band spectra according to iso 532b. Acustica **55** (1984) 63–67.

[17] E. Zwicker, H. Fastl: A portable loudness-meter based on ISO 532B. 11th International Congress on Acoustics, 1983, 135–137.

[18] E. Zwicker, H. Fastl: Psychoacoustics: facts and models. Springer, 1999.

[19] Z.-L. Lu, S. J. Williamson, L. Kaufman: Behavioral lifetime of human auditory sensory memory predicted by physiological measures. Science **258** (1992) 1668–1670.

[20] AES20: AES recommended practice for professional audio - Subjective evaluation of loudspeakers (Reaffirmed 2007). Tech. Rept. Audio Engineering Society, 1996.

[21] IEC 60268-13: Sound system equipment - part 13 : Listening test on loudspeakers. Tech. Rept. International Electrotechnical Commision, Geneva, Switzerland, 1998.

[22] A. Krieger, G. P.E.: A generalized rand-index method for consensus clustering of separate partitions of the same data base. Journal of classification **16** (1999) 63–89.

[23] I. Borg, P. Groenen: Modern multidimensional scaling, theory and applications, 2nd edition. Springer, New York, 2005.

[24] N. Jaworska, A. Chupetlovska-Anastasova: A review of multidimensional scaling (mds) and its utility in various psychological domains. Tutorials in Quantitative Methods for Psychology **5** (2009) 1–10.

[25] P. Y. Michaud, S. Meunier, P. Herzog, M. Lavandier, G. d'Aubigny: Perceptual evaluation of dissimilarity between auditory stimuli : an alternative to the paired comparison. Acta Acustica united with Acustica **99** (2013) 806–815.

[26] P. Susini, S. McAdams, S. Winsberg: A multidimensional technique for sound quality assessment. Acta Acustica united with Acustica **85** (1999) 650–656.

[27] ISO 3382: Acoustics - measurement of the reverberation time of rooms with reference to other acoustical parameters. International Organization for Standardization, Geneva (1997).

[28] J. Blauert: Spatial hearing: the psychophysics of human sound localization. MIT Press, Cambridge, 1983.

[29] D. Griesinger: Spaciousness and envelopment in musical acoustics. AES 101th Convention, 1996.

[30] T. Okano, L. Beranek, T. Hidaka: Relations among interaural cross-correlation coefficient (iacce), lateral fraction (lfe), and apparent source width (asw) in concert halls. J. Acoust. Soc. Am. **104** (1998) 255–265.

[31] S. Linkwitz: Why is bass reproduction from dipole woofer in a living room often subjectively more accurate than from a monopole woofer ? J. Audio Eng. Soc. **51** (2003) 1062–1063.

[32] Y. Ando, H. Sakai, S. Sato: Formulae describing subjective attributes for sound fields based on a model of the auditory-brain system. Journal of Sound and Vibration **232** (2000) 101–127.

[33] G. von Bismark: Sharpness as an attribute of the timbre of steady sounds. Acustica **30** (1974) 159–172.

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only

[34] S. Bech: The influence of the room and of loudspeaker position on the timbre of reproduced sound in domestic room. AES 12th International Conference: The Perception of Reproduced Sound, 1999.

[35] C. Guastavino, B. F. G. Katz: Perceptual evaluation of multi-dimensional spatial audio reproduction. The Journal of the Acoustical Society of America **116** (2004) 1105–1115.

[36] S. Choisel, F. Wickelmaier: Evaluation of multichannel reproduced sound: scaling auditory attributes underlying listener preference. Journal of the Acoustical Society of America **121** (2007) 388–400.

[37] D. Griesinger: The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces. Acta Acustica united with Acustica **83** (1997) 721–731.

[38] J. Bradley, G. Soulodre: Objective measures of listener envelopment. J. Acoust. Soc. Am. **98** (1995) 2590–2957.

[39] M. Barron, A. Marshall: Spatial impression due to early lateral reflections in concert halls: The derivation of a physical measure. Journal of Sound and Vibration **77** (1981) 211–232.

[40] R. Shimokura, L. Tronchin, A. Cocchi, Y. Soeta: Subjective diffuseness of music signals convolved with binaural impulse responses. Journal of Sound and Vibration **330** (2011) 3526–3537.

[41] M. Morimoto, Z. Maekawa: Effects of low frequency components on auditory spaciousness. Acta Acustica united with Acustica **66** (1988) 190–196.

[42] F. Toole: Sound reproduction: The acoustics and psychoacoustics of loudspeakers and rooms. Focal Press, Oxford, 2008.

Document downloaded by @DAEL, 2026-11-14 03:28:36 - personal use only