



HAL
open science

Identification of some perceptual dimensions underlying loudspeaker dissimilarities

Mathieu Lavandier, Sabine Meunier, Philippe Herzog

► To cite this version:

Mathieu Lavandier, Sabine Meunier, Philippe Herzog. Identification of some perceptual dimensions underlying loudspeaker dissimilarities. *Journal of the Acoustical Society of America*, 2008, 123 (6), pp.4186-4198. <hal-00345218>

HAL Id: hal-00345218

<https://hal.science/hal-00345218v1>

Submitted on 5 Dec 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Copyright - All rights reserved

Identification of some perceptual dimensions underlying loudspeaker dissimilarities

Mathieu Lavandier,^{a)} Sabine Meunier,^{b)} and Philippe Herzog^{c)}

Laboratoire de Mécanique et d'Acoustique, CNRS UPR 7051, 31 Chemin Joseph Aiguier,
13402 Marseille Cedex 20, France

(Received 19 July 2007; revised 9 January 2008; accepted 7 April 2008)

This study investigated the dimensions underlying perceived differences between loudspeakers. Listeners compared the sound reproduction of 12 loudspeakers in a room, using three musical excerpts. For the loudspeakers to be compared one just after the other in exactly the same conditions, the sounds radiated by the loudspeakers were recorded in a listening room, and the recorded sounds were submitted to paired comparisons using headphones. The resulting perceptual dissimilarities were analyzed by using a multidimensional scaling technique, revealing two main perceptual dimensions used by listeners to discriminate the loudspeakers. These dimensions were identical for the three musical excerpts. As the signals heard by listeners were directly accessible, they were used to define acoustical attributes describing the perceptual dimensions. Instead of arbitrarily choosing one acoustical analysis to define these attributes, several analyses were compared. The temporal, spectral, and time-frequency domains were investigated, and different auditory models were tested. These auditory models allowed the best description of the differences perceived by listeners, and were used to define two acoustical attributes describing our perceptual dimensions: the bass/treble balance and the medium emergence. © 2008 Acoustical Society of America. [DOI: 10.1121/1.2916688]

PACS number(s): 43.38.Md, 43.66.Lj, 43.20.Ye, 43.58.Kr [AJZ]

Pages: 4186–4198

I. INTRODUCTION

It is important to know why different loudspeakers sound differently both in terms of loudspeaker measurement and design and in terms of fundamentals of perception. Manufacturers have to control the acoustical parameters which can induce perceptual differences in the loudspeaker reproduction. Researchers interested in auditory perception try to identify the relevant acoustical criteria for discriminating complex sounds.

The perceptual characteristics of reproduced sound and their link with acoustical measurements were investigated by various authors, starting from the hypothesis that “perceived sound quality is constituted by a (limited) number of separate perceptual dimensions, and that it would be possible to give a perceptual description of sound-reproducing systems by stating their positions in such dimensions,” [Gabrielsson and Sjögren \(1979\)](#). Multidimensional scaling techniques can be used to reveal the perceptual dimensions underlying simple dissimilarity evaluations by listeners [[Borg and Groenen \(1997\)](#); [Susini et al. \(1999\)](#)]. They give access to the characteristics used to discriminate the stimuli, without listeners having to know or name what they are experiencing while listening. The differential semantic is another technique used to highlight perceptual dimensions [[Gabrielsson and Sjögren \(1979\)](#); [Guski \(1997\)](#)]. It involves absolute evaluations of

loudspeakers on chosen semantic scales labeled by adjectives. The number of scales is then reduced to independent factors by using factor analysis [[Dillon and Goldstein \(1984\)](#)].

The recommendations concerning listening tests on loudspeakers highlight three main categories of characteristics associated with the evaluation of sound reproduction: the timbre-related accuracy, sometimes called sound quality or fidelity and involving monophonic reproduction, the spatial quality involving stereophony and multichannel reproductions, and the dynamics and distortion characteristics related to the behavior of loudspeakers at different levels of solicitation [[AES20-1996 \(1996\)](#); [IEC Publication 60268-13 \(1998\)](#)]. Whereas recent studies were concerned with spatial quality and multichannel reproductions [[Choisel and Wickelmaier \(2007\)](#); [Guastavino and Katz \(2004\)](#); [Rumsey et al. \(2005a,b\)](#)], the main interest of our study was timbre-related accuracy. Several perceptual dimensions associated with timbre-related accuracy of sound-reproducing systems have been proposed [[Bramsløw \(2004\)](#); [Eisler \(1966\)](#); [Gabrielsson et al. \(1974\)](#); [Gabrielsson and Sjögren \(1979\)](#); [Klippel \(1990\)](#); [Staffeldt \(1974\)](#)]. These dimensions can be summarized as “clearness/distinctness,” “sharpness/hardness-softness” and “fullness-thinness,” “brightness-darkness,” “feeling of space” and “spaciousness,” “nearness,” “bass emphasis” or “treble emphasis,” and “disturbing sounds.” As noted by [Gabrielsson and Sjögren \(1979\)](#), these dimensions might be redundant and other dimensions might exist.

The perception of the sound radiated by a loudspeaker is greatly influenced by the room in which it is used and by the positions of both loudspeaker and listener [[Bech \(1994\)](#); [Olive et al. \(1994\)](#)]. The evaluation of loudspeakers also de-

^{a)}Electronic mail: lavandiermn@cardiff.ac.uk. Present address: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, United Kingdom.

^{b)}Electronic mail: meunier@lma.cnrs-mrs.fr

^{c)}Electronic mail: herzog@lma.cnrs-mrs.fr

depends on the loudness of their reproduction [Illényi and Korpásky (1981)] and on the musical excerpts used [Eisler (1966)]. Loudspeakers should then be compared in the same room, for the same positions of listener and loudspeaker, at the same sound level and with the same musical excerpt [AES20-1996 (1996); IEC Publication 60268-13 (1998)]. To evaluate their relative differences, they also have to be compared one just after the other, due to our short auditory memory [Olive *et al.* (1994)].

Because of the practical difficulty to compare many loudspeakers, the studies mentioned above tested loudspeakers at different positions in the listening room, and only a limited number of loudspeakers were compared in the same test. To our knowledge, a maximum of eight was reported by Klippel (1990). When more loudspeakers were tested in an experiment, only few comparisons were performed [Eisler (1966); Gabrielsson and Sjögren (1979)]. The perceptual dimensions involved in a listening test depend on the number and nature of the loudspeakers under evaluation. Gabrielsson and Lindstrom (1985) also showed that the evaluation of a loudspeaker on a particular scale was dependent on the other loudspeakers being tested in the experiment. To consider dimensions which are characteristic of the perception of sound reproduction in general, and not only on the particular loudspeakers under test, listening tests should involve as many different loudspeakers as possible.

The previous studies on the perception of reproduced sound often faced the difficulty of finding acoustical measurements of loudspeakers explaining the perceptual dimensions revealed by the listening tests. Three main approaches may be distinguished in the literature. The first one consisted in visual comparisons between the shapes of the frequency responses of loudspeakers and their evaluations along a given perceptual scale [Gabrielsson *et al.* (1991); Staffeldt (1974); Toole (1986)]. The link between perceptual and acoustical evaluations was not quantified and relied on the experimenter interpretation of the frequency responses, this interpretation involving many simultaneous criteria [Toole (1986)]. As a consequence, the most suitable measurement explaining perceptual evaluations differed from one author to the other [Olive (2004a)]. “Black box” models constituted a second approach [Bramsløw (2004); Olive (2004b)]. By taking acoustical measurements as inputs, they generate a rating as output, being as close as possible to a corresponding perceptual one. Olive (2004b) developed a multiple regression model predicting the loudspeaker preference of listeners from many different acoustical measurements. He was able to quantitatively compare the acoustical and perceptual ratings of preference, and found a very good agreement between them. However, this model does not individually treat the different perceptual dimensions, neither does it elucidate their link with the different acoustical measurements. A third approach looked for acoustical attributes derived from measurement curves, which would correlate well with perceptual dimensions revealed by listening tests [Klippel (1990); Olive (2004a)]. In order to bridge the gap between standardized measurements and perception, researchers tended to define acoustical measurements which took into account the listening conditions, but this was done to greater or lesser extents

and only a posteriori. To find acoustical attributes well correlated with his perceptual dimensions, Klippel (1990) had to resimulate the sound reproductions involved in his listening tests. This simulation was realized by using the original musical excerpts, the statistics of the room, and the anechoic measurements of the loudspeakers.

The aim of the work presented here was to investigate the perceptual space associated with the timbre-related accuracy of loudspeakers in a listening room and to find out the best way to describe the resulting perceptual dimensions by acoustical attributes. Our goal was to compare many loudspeakers in order to obtain perceptual dimensions which would be as much as possible independent of the tested panel. The loudspeakers were used in monophonic reproduction and their relative differences were evaluated. To be able to compare many loudspeakers one just after the other at the same position in the room, we decided to record their sound reproductions and submit the recordings to paired comparisons by using headphones. Previous studies dealing with the perception of reproduced sound compared listening tests realized live or using headphones [Bech (2002); Olive *et al.* (1994); Pedersen and Mäkivirta (2002); Toole (1991)]. Even if the studies had various aims, they all showed that the two types of listening test gave very similar results concerning timbre-related accuracy. We did not intend to evaluate all the potential dissimilarities between the tested loudspeakers. Some of these dissimilarities might not have been captured in the final recordings. Our aim was to study the remaining dominant dissimilarities, with the advantage of being able to compare many loudspeakers at the same position, one just after the other, making sure that the dissimilarities were associated with the loudspeakers and not another experimental parameter.

As our protocol used recordings of the sound reproductions rather than live listening tests, we also had a direct access to the signals compared by listeners. These signals were used to define acoustical attributes describing the perceptual dimensions involved in the listening tests. We chose to base our acoustical approach on a signal analysis directly done on the recordings, rather than considering estimations of the loudspeakers responses, in order to remain as close as possible to the signals judged by listeners. Instead of arbitrarily choosing one signal analysis, several analyses were compared. The most suitable one was identified as the analysis leading to differences between recordings which were the closest to the differences evaluated by listeners.

II. LISTENING TESTS

The sounds radiated by different loudspeakers in a listening room were recorded and the listening tests were conducted by using headphones. As the spatial component of the sound reproduction could not be reliably investigated with such protocol, the loudspeakers were used in monophonic reproduction, as “timbre-related accuracy is much more easily heard in single-loudspeaker listening,” AES20-1996 (1996).

TABLE I. Reverberation time RT measured by third-octave bands in the listening room used for the recordings as a function of the central frequency f of the third-octave band.

f (Hz)	200	250	315	400	500	630	800	1000	1250	1600	2000	2500	3150	4000	5000
RT (s.)	0.78	0.80	0.62	0.58	0.49	0.36	0.49	0.46	0.53	0.68	0.55	0.45	0.49	0.59	0.54

A. Stimuli

The listening room used for the recordings was 8.7 m long, 4.5 m wide, and 2.9 m high. The recording microphones and the loudspeaker were placed along the median axis of the room, with the loudspeaker at 1.5 m from the 4.5 m wall and the microphones at 2.20 m in front of the loudspeaker. The microphones were at 1 m from the floor, which approximately corresponds to the point between the medium and the tweeter of the loudspeakers. This room was chosen among a few which were available over the duration of the study, its characteristics being almost compatible with the standard for loudspeaker quality evaluation [IEC Publication 60268-13 (1998)]. Although this may not be considered as mandatory for estimating relative dissimilarities, we felt that keeping listening conditions close to an average living room was consistent with our goal. This room does not fully comply with the cited standard, especially at low frequencies, where its dimensions lead to a noticeable modal behavior. This is typical of many living rooms in contemporary urban flats, and we decided to consider this feature as part of the “usual” listening conditions for our comparisons, although it would not be quite appropriate for absolute quality evaluation. As stated in the standard, the reverberation time may not be measured at frequencies for which the mode density is low, and we therefore did not include such data below 200 Hz, although we checked that no mode resonance was much stronger than average in the frequency range of the tested loudspeakers. The reverberation time was measured at the position of the recording microphones, with the source at the loudspeaker position. Its value was found around 0.5 s at midrange frequencies (Table I).

Twelve single loudspeakers were involved in the recordings. As we were interested in evaluating their perceived differences, we chose loudspeakers leading to a broad range of dissimilarities: loudspeakers which would obviously be very dissimilar, others which should be very similar, and

finally loudspeakers which should lead to intermediate levels of dissimilarities. We tried to use a wide range of models coming from different manufacturers. They were commercial units, available in our laboratory or lent by partners and manufacturers. Because we were aware that our comparisons were done in nonstandard conditions, we agreed with our partners not to disclose the actual references of the loudspeakers, which are only designated by their number. Table II gives general characteristics about the tested panel. We chose units in the high-fidelity and small studio monitoring ranges, avoiding low-end or very high prices. We considered this panel as representative of loudspeakers dedicated to music listening for people with actual interest in sound quality, but neither professional nor audiophile. This is consistent with our panel of listeners. We only kept electrodynamic units with widely available acoustic loads (closed or vented boxes) and usual filtering options (wide band to three ways). All loudspeakers were different from each other, except numbers 7 and 8, which correspond to the two units of a stereo pair. They were included in the test to get an example of two very similar reproductions, in order to check the validity of our protocol.

Figure 1 provides an estimation of the frequency response of each loudspeaker, measured in the listening room used for the experiment, with the loudspeakers and the measuring microphone at the positions used for the recordings. The responses were estimated with pink noise, considering the root-mean-square average of 128 fast Fourier transforms (FFTs) of Hanning windowed sections of the signal with 50% overlap. They were calculated as the spectrum at the measurement microphone divided by the spectrum of the input voltage of the loudspeaker. The FFT bins were summed in each third-octave band. The 12 estimated responses were normalized by their value in the 1 kHz band to facilitate their comparison. Figure 1 gives a general view of the loudspeaker panel, but also illustrates the performance of the

TABLE II. General characteristics of the 12 tested loudspeakers.

No.	Use	Box	Speakers	Filter
1	HiFi-medium range	Column, closed	2 boomers+1 wide band+1 tweeter	3 ways
2	Miniature monitor	Compact, closed	1 wide band	1 way
3	Professional monitor (90's)	Shoebox, vented	1 boomer+tweeter	2 ways
4	HiFi-bookshelf	Shoe box, closed	1 boomer+tweeter	2 ways
5	Professional monitor (80's)	Shoebox, vented	1 boomer+1 medium+1 tweeter	3 ways
6	Home-Studio	Shoebox, vented	1 boomer+tweeter (horn)	2 ways
7	HiFi-budget	Shoebox, closed	1 boomer+tweeter	2 ways
8	HiFi-budget	Shoebox, closed	1 boomer+tweeter	2 ways
9	Small monitor	Shoebox, vented	1 coax(boomer+tweeter)	2 ways
10	Small monitor	Shoebox, vented	1 coax(boomer+tweeter)	2 ways
11	HiFi-bookshelf	Shoebox, closed	1 boomer+tweeter	2 ways
12	Miniature monitor	Shoebox, vented	1 wide band	1 way

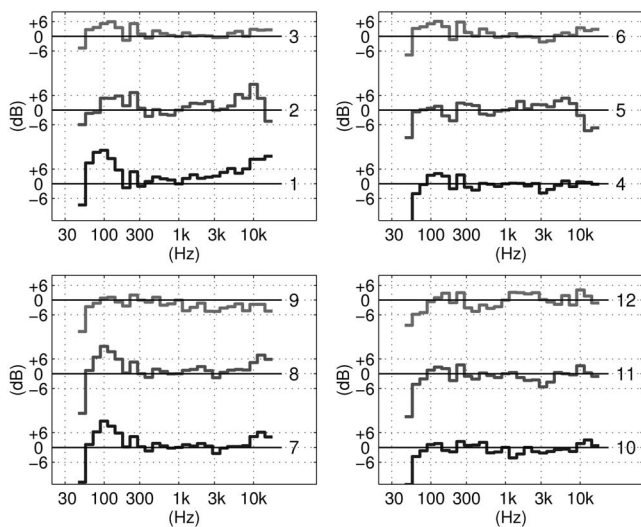


FIG. 1. Frequency responses of the 12 tested loudspeakers (Table II), measured in the configuration used for the recordings.

room: The response of loudspeaker 3 (considered as the higher-end unit) is kept within ± 6 dB, but presents a few accidents at low frequencies which are shared with many other loudspeakers, and may be attributed to the room.

The musical excerpts used for the recordings were chosen according to the recommendations concerning listening tests on loudspeakers [AES20-1996 (1996); IEC Publication 60268-13 (1998)]. They were of very different musical styles, involving acoustical instruments rather than synthetic music, and were selected among high-quality commercial recordings. Our monaural stimuli were obtained by considering only one channel of the original stereophonic excerpts. Three musical excerpts were recorded and we only kept a short part of them for the listening tests: McCoy Tyner (Miss Bea, Best of Chesky Jazz, Vol. 2, Chesky CD:68, from 01:01 to 01:05, right channel), Kan'nida (Konsyans, Kyenzenn, Indigo LBLC 2566, from 00:07 to 00:09, left channel), and Vivaldi (L'Europa Galante, Fabio Biondi, Opus 111, OPS 30-86, Concerto pour violino e organo in re minore, from 00:08 to 00:13, left channel). Such short excerpts were reported as suitable for the evaluation of perceived differences in timbre-related accuracy [Bech (1995, 1996); Moore and Tan (2003)].

The stereophonic ORTF recording technique was used, with two cardioid microphones placed 17 cm and 110° apart from each other (AKG Blue Line CK-91, SE-300B, preamplifier Tascam MX-4). The musical excerpts were stored on a compact disk and reproduced on the loudspeakers by high-grade CD player and amplifier (Vecteur I-4.2, Vecteur L-3.2, Behringer Ultralink Pro MX882 as a preamplifier for the studio monitoring loudspeakers). The recordings were directly carried out with an audio workstation featuring an RME DIGI9652 sound card and an external Fostex VC-8 A/D converter. The sampling frequency was 44 100 Hz. During the recordings, reproduction levels were roughly adjusted to normal listening conditions. A more accurate loudness equalization was undertaken by the experimenters prior to the listening tests.

B. Procedure

Perceptual dissimilarities between recordings were gathered during three listening tests, one test for each musical excerpt. The listening tests were paired comparisons where the listener had to directly evaluate the dissimilarities between the stimuli. For each test, the 12 recordings were presented by pairs to the listener in random order. The listener had to quantify the overall dissimilarity within each pair by adjusting a cursor on a line whose end points were labeled "very similar" and "very dissimilar," this line being displayed on a computer terminal screen in front of the listener. A numerical value was linearly assigned to the position of the cursor. The resulting perceptual dissimilarity varied from 0 for very similar to 1 for very dissimilar. Participants could listen to each pair of recordings as many times as they wanted before doing their dissimilarity evaluation, but they were encouraged to base this evaluation on their first impression and avoid multiple listenings. Listening tests lasted about 20 min (Kan'nida), 25 min (McCoy Tyner), or 35 min (Vivaldi).

Before the tests, the overall loudness of the recordings was set to the same level of 70 phons, as judged by the experimenters and verified by a loudness estimation model [Zwicker and Fastl (1983, 1999)]. This prevented loudness from creating uninteresting dissimilarities potentially masking more subtle ones. Listening tests were run in an isolated soundproof room by using a Tucker&Davis workstation and Stax SR Lambda Professional headphones. The frequency responses of the recording microphones and headphones were not compensated for, but all comparisons were made with the same headphones and between recordings involving the same microphones.

Twenty-seven listeners (12 women, 15 men) took part in the experiment, each of them participating in the three listening tests in random order. They were between 14 and 53 years old, with an average age of 30. The listeners were otologically normal. Twenty five had hearing thresholds less than 25 dB HL from 250 to 8000 Hz. Two had slight presbycusis (thresholds between 25 and 40 dB HL above 4 kHz), normal for their age. The results from these two listeners did not differ from the data of the group, indicating that their presbycusis did not affect their judgments. They were members of the laboratory or students. None of them had a significant previous experience in loudspeaker comparison.

C. Results

For each listening test, the final perceptual dissimilarities were obtained by averaging individual dissimilarities. Individual dissimilarities were not scaled before the averaging, but it was verified that there was no group with different judgement strategies among listeners. This verification was based on a cluster analysis of the similarities between listeners, the similarity between two listeners being evaluated by the correlation of their judgments. Figure 2 presents the dendrogram resulting from this analysis for the listening test involving the musical excerpt McCoy Tyner, which was here arbitrarily chosen as an example. There would be classes of

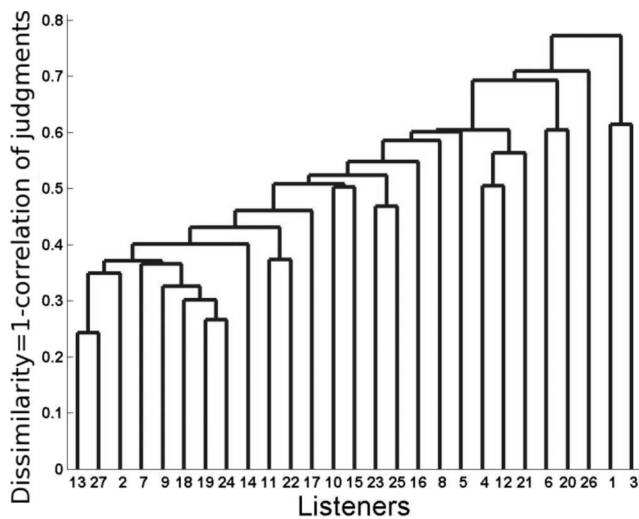


FIG. 2. Dendrogram resulting from the cluster analysis of the similarities between listeners (listening test involving the musical excerpt McCoy Tyner). Each number corresponds to one of the 27 listeners. The height of the node connecting two listeners corresponds to the dissimilarity of their judgments, whereas the position of the listeners on the horizontal axis is irrelevant.

listeners with different judgement strategies, if one could isolate at least two groups of listeners having well correlated judgments within each group, and uncorrelated judgments between groups. No such group appears in Fig. 2. Any partition of the listeners “continuously” depends on the degree of similarity chosen to discriminate them. No classes of listeners were found in any of the three listening tests.

The final mean perceptual dissimilarities between recordings were ranging between 0.22 and 0.83 (McCoy Tyner), 0.13 and 0.90 (Kan'nida), or 0.21 and 0.84 (Vivaldi), indicating that the chosen loudspeakers and musical excerpts led to a broad range of dissimilarities [Lavandier *et al.* (2008)]. These dissimilarities were submitted to multidimensional scaling analysis (MDS). The classical model MDSCAL was used. The algorithm SMACOF, for “scaling by majorizing a complicated function,” has been detailed by Borg and Groenen (1997). 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 measurements. The algorithm does not fix the axes of the multidimensional space. This space can be rotated or dilated; only relative distances between stimuli are fixed. By default, the axes 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 dimension.

Interpretation of MDS results requires much caution: Even if the analysis allows to locate stimuli in a space with continuous orthogonal dimensions, the MDS does not mean that such dimensions correspond to actual perceptual attributes. Even heterogeneous stimuli—involving different categories with no perceptual dimension relating them—lead to a continuous MDS representation. Our experiments apparently involved homogeneous stimuli, so it seemed relevant to look for continuous perceptual dimensions [Susini *et al.* (1999)]. Nevertheless, our MDS analyses were always asso-

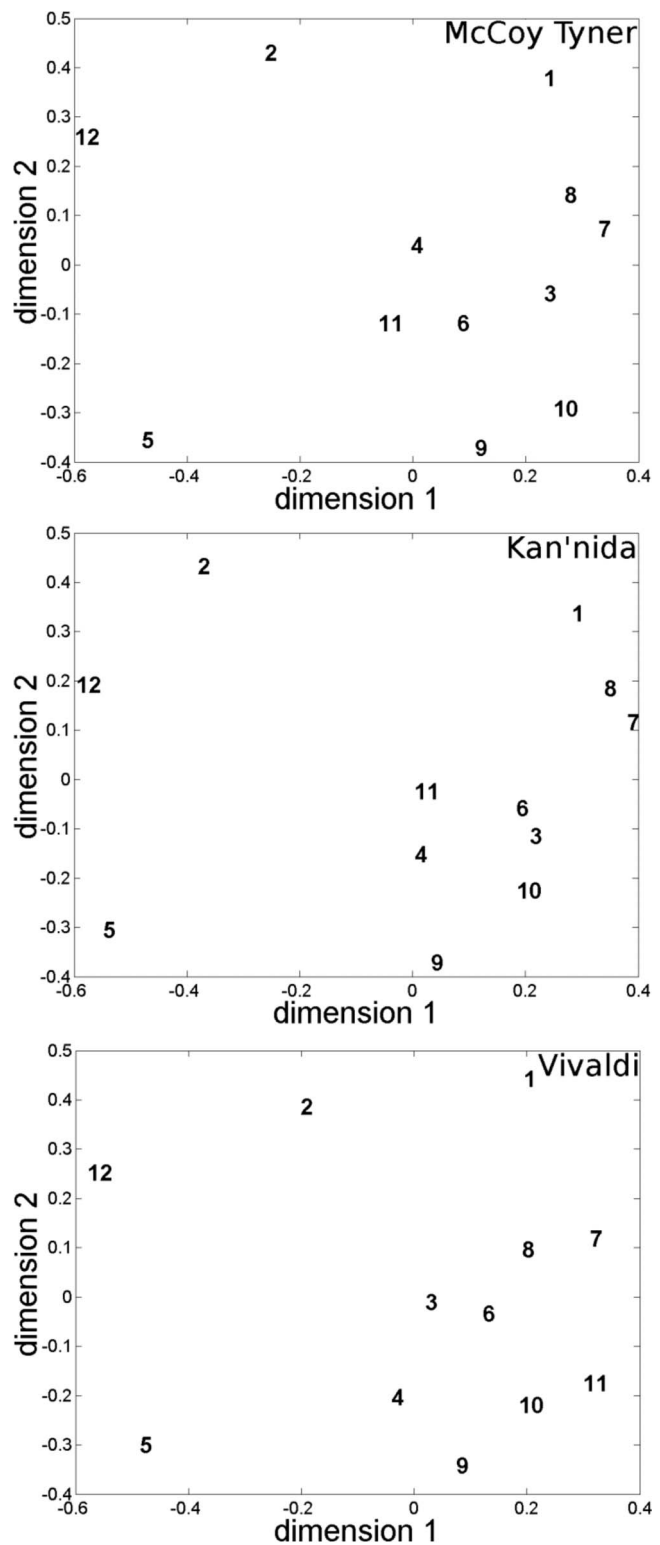


FIG. 3. Two-dimensional spaces resulting from the MDS analysis of the perceptual dissimilarities obtained with each musical excerpt (McCoy Tyner, Kan'nida, and Vivaldi). Each loudspeaker is identified by its number (Table II).

ciated with cluster analyses applied to the same dissimilarity data, in order to verify the homogeneity of our stimuli by highlighting the potential presence of classes among them. These cluster analyses never revealed any classes among our recordings.

Figure 3 presents the perceptual spaces resulting from

the three listening tests. In each case, a two-dimensional space was obtained. Each number in these spaces corresponds to a loudspeaker (Table II). To be compared, the perceptual spaces were all rotated and dilated to fit as much as possible the space involving the musical excerpt Kan'nida, which was arbitrarily chosen as a reference with its default orientation.

The perceptual spaces obtained with the three musical excerpts were very similar. The influence of the musical excerpt on the evaluation of the perceptual dissimilarities was small, as shown by the rather small changes in the relative positions of loudspeakers between the three spaces. Listeners used the same two main perceptual dimensions to discriminate the recordings during the three listening tests, despite the change of musical excerpt. These two dimensions would then be characteristic of the loudspeakers under evaluation. One can notice that loudspeakers 7 and 8, the two loudspeakers of the same model, stand very close together in the perceptual spaces.

We listened to the recordings while looking at the perceptual spaces and used the free comments of listeners to propose an interpretation of the dimensions. The predominant dimension 1 seemed linked to the bass/treble balance of the recordings. The dimension 2 would be linked to the amount of medium frequencies, and could be associated with the notion of sound clarity. These interpretations were consolidated by acoustical attributes presented in the following.

III. ACOUSTICAL ANALYSES

As our protocol used recordings of the sound reproductions of loudspeakers rather than live listening tests, we had a direct access to the signals compared by listeners. These signals were used to define acoustical attributes describing the perceptual dimensions revealed by the listening tests. First, a suitable signal analysis of the recordings had to be chosen. For example, the recordings could be compared in the time, spectral, or time-frequency domains. Instead of arbitrarily choosing one signal analysis, we decided to compare several analyses and choose the analysis that showed out differences between recordings which were similar to those evaluated by listeners. Several methods of signal analysis were tested, and for each method, a metric evaluating the acoustical dissimilarity between two signals was defined. The acoustical dissimilarities were computed for all recordings and compared to the perceptual dissimilarities, in order to identify the most relevant acoustical discrimination method. As the analyses and metrics were presented in detail in a previous paper [Lavandier *et al.* (2008)], they are only summarized here.

A. Procedure

1. Signal analyses

The temporal, spectral, and time-frequency domains were investigated. Two spectral weightings were tested. Different auditory models were also used. These auditory models are considered here as acoustical analyses and not perceptual ones, because they were used like any other signal analysis technique without any listener being involved.

The spectral domain was investigated by using the discrete Fourier transform of the signals and their power spectral density. The time-frequency plane was obtained by calculating their short-time Fourier transform. We considered the spectral and time-frequency domains with and without phase information. Spectral weightings were applied to the power spectral density of the signals to model the ear sensitivity. They were the *A*-weighting and a weighting based on the normal equal-loudness contour at 70 phons [British Standard ISO 226 2003 (2003)], which corresponds to the sound level used during our listening tests.

Two auditory models were tested. These models were proposed by Zwicker and Fastl (1999) to calculate loudness, and are based on auditory masking. They were used to analyze the signals in terms of specific loudness which is the density of loudness along the Bark scale modeling the perceptual frequency scale. The first model [Paulus and Zwicker (1972); Zwicker *et al.* (1984)] was originally designed to evaluate the loudness of stationary sounds. It takes into account only frequency masking, and was used to calculate what we called the overall specific loudness of the signals, which is the specific loudness determined over the entire signal taken at once. Two time-frequency patterns of specific loudness were examined. The first one was obtained by applying the first previous model to successive 100 ms sections of the signals. We called the resulting analysis time-varying specific loudness 1. To get the second time-frequency pattern of specific loudness, a second auditory model was used. This model was designed for nonstationary sounds [Zwicker and Fastl (1983, 1999)]. It takes into account both frequency and temporal masking, calculating specific loudness every 10 ms. We called the resulting analysis time-varying specific loudness 2. Finally, another way to analyze the signals by using these loudness models was to average the time-varying specific loudness over time. The temporal means of time-varying specific loudnesses 1 and 2 were considered. Compared to the overall specific loudness, they took into account the fact that auditory masking depends on the spectral content of signals and that this content varies over time for nonstationary musical signals.

2. Acoustical dissimilarities

For each method of analysis, we defined a metric evaluating an overall acoustical dissimilarity between two recordings, giving a single value from the differences between the analyses of these two recordings [Lavandier *et al.* (2008)]. To get a single scalar, dissimilarities calculated along the temporal, spectral, or Bark scales were integrated. As we could not suppose that a sample period or a frequency region was more important than another, this integration was realized by an arithmetic mean, giving the same weight to all dissimilarities.

For the analyses not involving the auditory models, the acoustical dissimilarity between recordings x and y was calculated following the general formula

$$\min\{\langle [Ax(v) - Ay(v)]^2 \rangle_v, \langle [Ax(v) + Ay(v)]^2 \rangle_v\}, \quad (1)$$

where Ax and Ay are the considered signal analysis of x and y (waveform, Fourier transform, short-time Fourier trans-

form, modulus of Fourier transform or short-time Fourier transform, or power spectral density with or without weightings), v is the variable associated with this analysis (time, frequency, or time and frequency), $\langle \rangle_v$ takes the arithmetic mean over the variable v , and $\min\{\}$ takes the minimum value. To get a single overall dissimilarity between x and y , dissimilarities calculated at each sample period or each frequency are integrated by taking the arithmetic mean over the considered scale. Because phase information was sometimes considered, a potential phase inversion between the two compared recordings x and y is taken into account by considering y and $-y$, and the smallest dissimilarity is kept.

The acoustical dissimilarity between the overall specific loudnesses or the temporal mean of time-varying specific loudnesses of recordings x and y was calculated following the formula

$$\langle \max\{\text{SL}(x), \text{SL}(y)\} / \min\{\text{SL}(x), \text{SL}(y)\} - 1 \rangle_B, \quad (2)$$

where SL is the specific loudness considered, $\langle \rangle_B$ takes the arithmetic mean over the Bark scale, $\min\{\}$ takes the minimum value, and $\max\{\}$ takes the maximum value. The dissimilarities were computed by taking the ratio of the specific loudnesses, as two loudnesses should be compared by their ratio not their difference [Zwicker and Fastl (1999)]. We took care of always dividing the largest value by the smallest one, as we would have taken the absolute value for a difference. Before calculating the ratios, a threshold was introduced in order to avoid artificially high dissimilarities where there was no useful signal but noise. The integration of the resulting dissimilarities over the Bark scale was done by taking their arithmetic mean. For the acoustical dissimilarities involving the time-varying specific loudnesses, the instantaneous dissimilarities were calculated following formula (2), and then integrated over time.

Before calculating the different acoustical dissimilarities, our signals were synchronized [Lavandier *et al.* (2008)]. The right and left channels of our stereophonic recordings were separately analyzed. For each musical excerpt and each channel, we determined the different acoustical dissimilarities between the 12 recordings corresponding to the 12 loudspeakers. Acoustical dissimilarities were then compared to the perceptual ones, in order to determine the acoustical analyses allowing to show out the differences heard by listeners.

3. Comparison of acoustical and perceptual dissimilarities

A previous comparison of acoustical and perceptual dissimilarities consisted in directly evaluating the correlation between them [Lavandier *et al.* (2008)]. The dissimilarities between the 12 recordings were gathered in 13 acoustical 66-value dissimilarity vectors—one vector per type of acoustical dissimilarity computed on the signals—and one perceptual 66-value dissimilarity vector. The different acoustical dissimilarity vectors were compared by evaluating their coefficient of correlation with the perceptual dissimilarity vectors. The acoustical analyses keeping phase information led to a correlation of about 0.2, whereas the correlation increased to 0.5 when phase information was not considered.

The spectral weightings only improved the correlation for one musical excerpt (Kan'nida). On the other hand, the auditory models greatly improved the correlation between acoustical and perceptual dissimilarities, for all tested signals, with a coefficient of correlation often above 0.8. These correlations indicated that the acoustical dissimilarities involving auditory models contained at least part of the information contained in the perceptual dissimilarities.

A second comparison of the acoustical and perceptual dissimilarities is presented here. Indeed, the acoustical and perceptual dissimilarities could be badly correlated because they are not linked at all, but also because their relation is nonlinear, or because the acoustical dissimilarities contain more or less underlying dimensions than the perceptual ones. The restricting criterion of linearity associated with correlation was abandoned, and the multidimensional nature of our perception of reproduced sound was taken into account. The acoustical dissimilarities were submitted to the same MDS analysis as the perceptual dissimilarities, and the resulting acoustical spaces were visually compared to the perceptual ones shown in Fig. 3. This type of comparison indicates if an acoustical analysis seems appropriate to highlight any of the perceptual dimensions. Even if the acoustical and perceptual spaces are different, it is interesting to know if some acoustical dimensions are relevant regarding perception. Compared to the evaluation of correlation between dissimilarities, it gives a better understanding of the information contained in acoustical and perceptual dissimilarities.

B. Results

Thirteen acoustical spaces resulted from the MDS analysis of the acoustical dissimilarities associated with the 13 different acoustical discrimination methods presented above. To facilitate their comparison, the acoustical spaces were rotated and dilated to fit the same reference as the perceptual spaces. In front of the large number of acoustical spaces obtained, only some examples are presented here. The analyses involving auditory models were the only acoustical analyses leading to two-dimensional spaces being close to the perceptual ones for the three musical excerpts. The other analyses led to one- or two-dimensional spaces, depending on the signals considered. No link could be found between the corresponding acoustical dimensions and the perceptual ones. The two acoustical spaces presented in Fig. 4 result from the analysis of the dissimilarities obtained with the waveform of the recordings (top) and with their power spectral density (bottom). They should be compared to the corresponding perceptual spaces of Fig. 3.

Figure 5 presents one acoustical space based on the overall specific loudness for each musical excerpt. The overall specific loudness led to acoustical spaces close to the perceptual space only for the musical excerpt Kan'nida. For the two other musical excerpts, the global analysis applied on the entire signal taken at once used while calculating overall specific loudness was not sufficient to fully describe the perceptual dissimilarities.

On the other hand, time-varying specific loudnesses and their temporal mean led to acoustical spaces very similar to

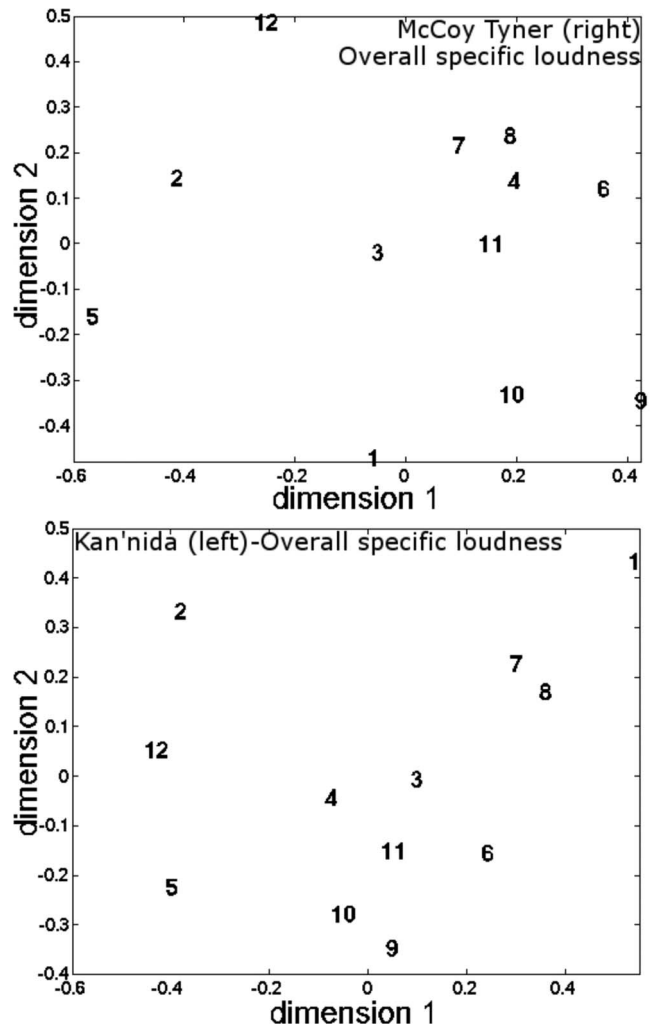
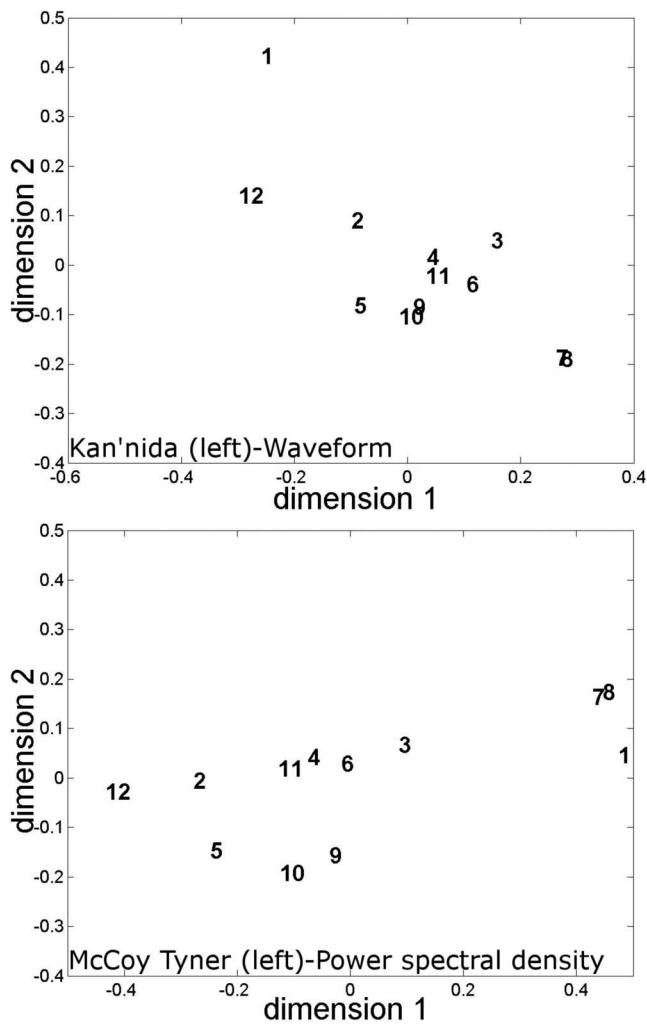


FIG. 4. Examples of acoustical spaces based on the waveform (top) and the power spectral density (bottom). They should be compared to the corresponding perceptual spaces of Fig. 3.

the perceptual ones for the three musical excerpts. The results were equivalent for models 1 and 2 and for time-varying specific loudnesses and their temporal means. Figure 6 presents two examples obtained with the signals leading to the best (Kan'nida, left channel) and the worst (Vivaldi, right channel) correspondence between acoustical and perceptual dissimilarities. The time-varying specific loudness 2 was preferred to the time-varying specific loudness 1 because it is based on a more widely used computation [Fastl (1997)]. We decided to present examples based on the temporal mean of time-varying specific loudness rather than on the time-varying specific loudness because the temporal mean significantly reduces the amount of data involved in the dissimilarity evaluation, while giving results which were found very close to those obtained without averaging.

C. Discussion

The comparison of acoustical and perceptual spaces confirmed the results based on the simple correlation between acoustical and perceptual dissimilarities [Lavandier *et al.* (2008)]: The acoustical discriminations by using auditory models were the closest to the perceptual evaluation. Even if

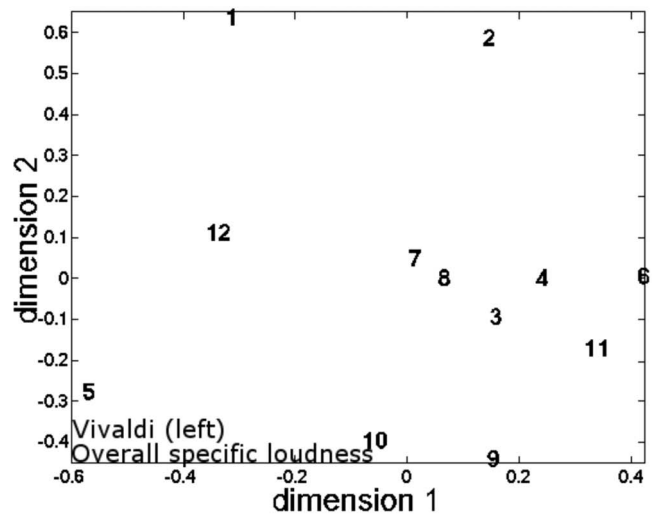


FIG. 5. Examples of acoustical spaces based on the overall specific loudness. They should be compared to the corresponding perceptual spaces of Fig. 3.

a criterion is still needed to quantify the similarity between two spaces, the visual comparison of spaces indicated that the time-varying specific loudnesses and their temporal means were suitable to describe the two perceptual dimensions revealed by the listening tests. The overall specific loudness might, however, not be sufficient to fully describe the perceptual dissimilarities.

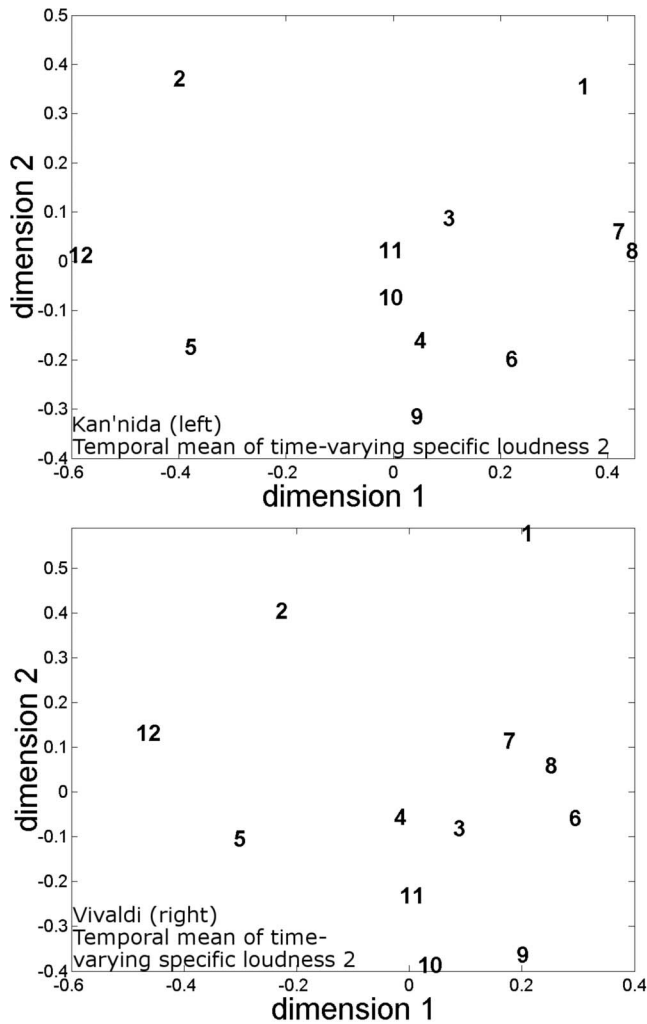


FIG. 6. Examples of acoustical spaces based on the temporal mean of time-varying specific loudness 2. They should be compared to the corresponding perceptual spaces of Fig. 3.

The temporal resolution of the auditory models did not seem crucial regarding the two dimensions involved in our listening tests. The two time-varying specific loudnesses gave equivalent results, although time-varying specific loudness 1 was calculated every 100 ms and time-varying specific loudness 2 was evaluated every 10 ms. Taking into account temporal masking in model 2 did not bring further improvement. This would support the hypothesis used by Bramsløw (2004) who neglected this last effect in his reproduced sound quality evaluation model. Time-varying specific loudnesses and their temporal mean gave equivalent results, suggesting that the two perceptual dimensions were linked to an average impression over the entire musical excerpt [Bramsløw (2004); Gabrielsson *et al.* (1990, 1991); Staffeldt

(1984, 1991)].

IV. ACOUSTICAL ATTRIBUTES

We defined acoustical attributes describing the perceptual/acoustical dimensions. The temporal mean of time-varying specific loudness 2 of the stimuli was chosen as a basis for this definition, as it was one of the most relevant analyses for discriminating the recordings. We described the acoustical dimensions as they were very similar to the perceptual dimensions (Figs. 3 and 6). Our first criterion to define acoustical attributes was to look for a monotonic relation between the dimension and the attribute. The criterion of correlation was also considered, even if it is more restrictive as it looks for a linear relationship between dimension and attribute.

A. Bass/treble balance

The first proposed acoustical attribute evaluates the bass/treble balance of the recordings. It is defined as the ratio between the loudnesses in the three first Bark bands (corresponding to frequencies between 20 and 280 Hz) and in the ten last Bark bands (corresponding to frequencies between 1.8 and 15.5 kHz). These two values of loudness are calculated by integrating the temporal mean of time-varying specific loudness 2 over the specified Bark bands. These Bark bands were chosen in order to give the best possible results to describe the first dimension of the spaces resulting from the three musical excerpts.

The relation between bass/treble balance and first dimension of the acoustical space was monotonic for almost all signals, and often linear. The correlation coefficient between attribute and dimension varied between 0.80 and 0.96 for the different musical excerpts, with an average value of 0.90 (Table III). Figure 7 presents two examples: The most favorable one with a linear relation (Kan'nida, left channel) and the least favorable one for which monotony was not preserved (Vivaldi, right channel).

B. Medium emergence

The second proposed acoustical attribute evaluates the emergence of medium frequencies in the recordings. The medium emergence is defined as the loudness in the Bark bands 5–9 (corresponding to frequencies between 355 and 1120 Hz). This loudness is calculated by integrating the temporal mean of time-varying specific loudness 2 over the specified Bark bands. As all recordings were equalized in overall loudness, the attribute evaluates the perceived emergence of this frequency range compared to the rest of the spectrum.

TABLE III. Correlation coefficient between the bass/treble balance and the dimension 1 of the acoustical space based on the temporal mean of time-varying specific loudness 2, for the left and right channels of the recordings and the three musical excerpts.

Musical excerpt	Kan'nida		McCoy Tyner		Vivaldi	
	Left	Right	Left	Right	Left	Right
Correlation	0.96	0.95	0.92	0.83	0.93	0.80

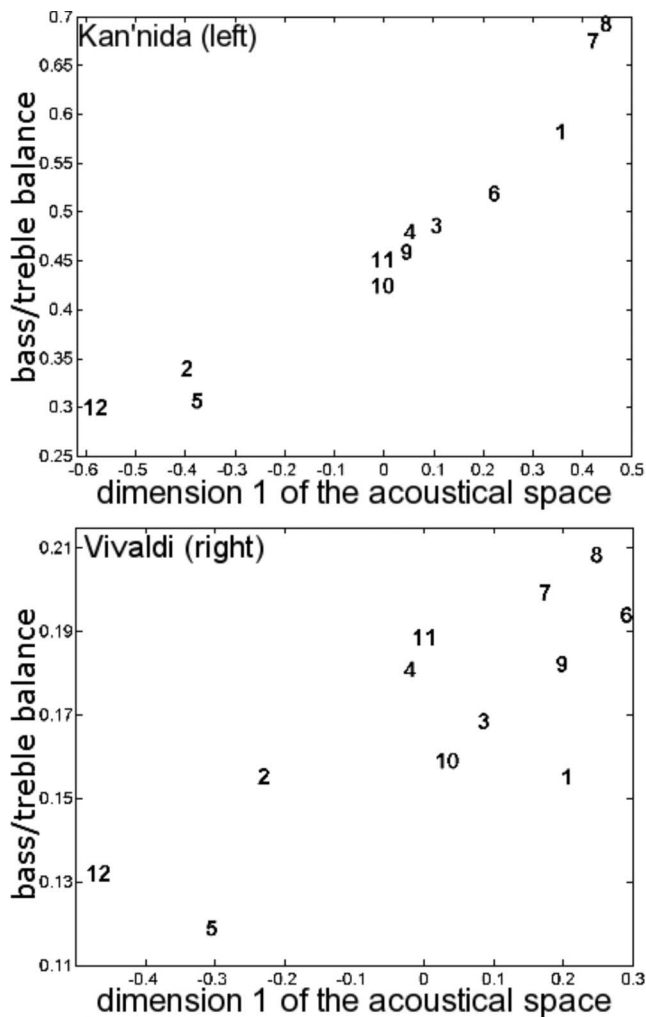


FIG. 7. Bass/treble balance as a function of the dimension 1 of the acoustical space based on the temporal mean of time-varying specific loudness 2. Top: case leading to the best correlation between dimension and attribute (Kan'nida, left channel). Bottom: case leading to the worst correlation between dimension and attribute (Vivaldi, right channel). Each loudspeaker is identified by its number (Table II).

The relation between medium emergence and second dimension of the acoustical space was monotonic for almost all recordings, and often linear. The correlation coefficient between attribute and dimension varied between 0.65 and 0.96 for the different musical excerpts, with an average value of 0.87 (Table IV). Figure 8 presents two examples: The most favorable one with a linear relation (McCoy Tyner, right channel) and the least favorable one (Kan'nida, left channel). This last case was an exception among the tested signals, for which the proposed attribute was not sufficiently discriminant.

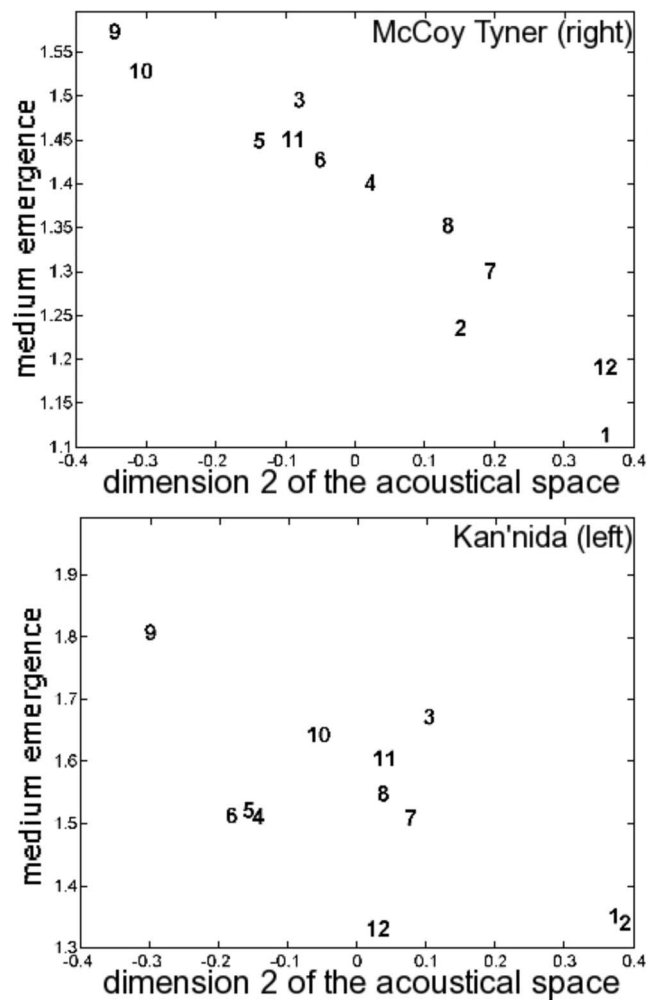


FIG. 8. Medium emergence as a function of the dimension 2 of the acoustical space based on the temporal mean of time-varying specific loudness 2. Top: case leading to the best correlation between dimension and attribute (McCoy Tyner, right channel). Bottom: case leading to the worst correlation between dimension and attribute (Kan'nida, left channel).

V. GENERAL DISCUSSION

The dissimilarities between 12 loudspeakers radiating in a listening room were investigated by using three musical excerpts. As our experimental protocol was based on recordings of the sound reproductions rather than live listening tests, loudspeakers could be compared one just after the other at the same position in the room. A MDS analysis of the resulting perceptual dissimilarities revealed two main dimensions used by listeners to discriminate the loudspeakers. Our protocol gave us direct access to the signals heard by listeners, so we were able to use these signals to define acoustical attributes describing the perceptual dimensions.

TABLE IV. Correlation coefficient between the medium emergence and the dimension 2 of the acoustical space based on the temporal mean of time-varying specific loudness 2, for the left and right channels of the recordings and the three musical excerpts.

Musical excerpt	Kan'nida		McCoy Tyner		Vivaldi	
	Left	Right	Left	Right	Left	Right
Correlation	0.65	0.87	0.88	0.96	0.93	0.95

As the perceptual dimensions did not depend on the particular musical excerpt used, we defined general attributes which were appropriate for all types of signal, giving a measure more characteristic of the loudspeakers and less dependent on the particular signal being reproduced. The specific loudness of the recordings was used to define two acoustical attributes well correlated with the perceptual dimensions: the bass/treble balance and the medium emergence.

These dimensions were in good agreement with the literature. The bass/treble balance can be compared to the “balance between bass and treble” or “brightness-darkness” found by [Gabrielsson et al. \(1974\)](#) or the “fullness-thinness” highlighted by [Gabrielsson and Sjögren \(1979\)](#). This last dimension was later defined by [Toole \(1985\)](#) as being associated with the amount of low frequencies compared to medium and high frequencies. [Gabrielsson et al. \(1990\)](#) linked their dimension “brightness versus dullness” to the levels of bass and treble, and the “fullness” was associated with the amplification of low frequencies. In [Gabrielsson et al. \(1990\)](#), the acoustical attributes were directly evaluated on the loudspeaker’s frequency responses, but the link between attribute and perceptual dimension was not quantified. It was only based on visual comparisons, that is to say on trends observed by experimenters on frequency responses compared to perceptual evaluations. In the same way, [Staffeldt \(1974\)](#) mentioned “emphasized treble” and “emphasized bass” to interpret his results. The listening tests of [Klippel \(1990\)](#) revealed the same kind of dimension, and he used the specific loudness of his signals to define acoustical attributes describing the results of his listening tests. These attributes relied on different weightings favoring different frequency bands. He obtained attributes well correlated with the dimensions “treble stressing” (“sharpness”), “general bass emphasis” (“volume”), and “low bass emphasis,” with correlation coefficients ranging from 0.87 to 0.90. He also described “brightness,” presented as an overall evaluation including “general bass emphasis” and “treble stressing,” by a linear combination of the two corresponding attributes. This last attribute seems closely related to our bass/treble balance. Our first dimension could also be compared to the “sharpness/fullness” highlighted by [Bramsløw \(2004\)](#), which was interpreted as the ratio between low and high frequencies. For [Olive \(2004a\)](#), the perceived spectral balance was an important factor relative to listener’s preferences. A semantic analysis of free comments from his listeners highlighted nine factors, from which he interpreted the two main ones as “bass quality, overall spectral balance, and smoothness” and “midtreble colorations like bright, midpeak (megaphone), and colored.” The first factor can be linked to our bass/treble balance, whereas the midtreble colorations can be linked to the second dimension revealed by our listening tests, the medium emergence. This last dimension also seems similar to the “distinctness” (“clarity” and “clearness”) of [Gabrielsson et al. \(1974\)](#) and the “clearness” of [Klippel \(1990\)](#) or [Bramsløw \(2004\)](#). Our attribute describing the medium emergence involved the frequency range 355–1120 Hz. [Bech \(1996\)](#) indicated that the frequency range 500–2000 Hz was crucial for changes in the restitution of

timbre. [Gabrielsson et al. \(1990\)](#) associated their dimension “clarity” with an amplification of medium frequencies between 500 and 4000 Hz.

Even if musical excerpts influence the perceptual evaluation of sound reproduction, the dimensions involved in our experiment were interpreted in the same way for the three tested musical excerpts, like in the listening tests of [Gabrielsson et al. \(1974\)](#). Our dimensions also remained unchanged when other recording techniques [[Lavandier et al. \(2004\)](#)], reproduction modes [[Lavandier et al. \(2005a\)](#)], and listening room [[Lavandier et al. \(2005b\)](#)] were tested. These perceptual dimensions seem then characteristic of the loudspeakers we intended to measure. The experiment presented here involving 12 loudspeakers was a first attempt to use our protocol to investigate the sound reproduction of loudspeakers in rooms. In order to consider dimensions which are characteristic of the perception of sound reproduction in general, listening tests should involve as many different loudspeakers as possible. To our knowledge, 12 loudspeakers involved in the same paired-comparison test is the maximum of what can be found in the literature. Paired comparisons rapidly limit the number of loudspeakers which can be included in the same test, as the listening test becomes too long for the participants to be able to handle the task. A listening test based on a free classification task allowed us to involve up to 37 loudspeakers [[Lavandier et al. \(2005b\)](#)]. The results were promising as four perceptual dimensions were revealed, two of them being very similar to the dimensions presented in this paper. The potential influence of the classification task has yet to be clarified.

Our approach to define acoustical attributes describing the perceptual dimensions consisted in identifying first the best method of analysis for a relevant description of the signals. Whereas the evaluation of the correlation between the different types of acoustical dissimilarities and the perceptual dissimilarities allowed us to quantify the link between measurement and perception [[Lavandier et al. \(2008\)](#)], the comparison of acoustical and perceptual spaces took into account the multidimensional nature of our perception of reproduced sound. [Pols et al. \(1969\)](#) used a similar approach, comparing acoustical and perceptual spaces resulting from a principal component analysis, to look for a method of analysis relevant to study the perception of vowels. Even if a criterion is still needed to quantify the similarity between two spaces, it seems easier to visually compare spaces with a low number of dimensions rather than loudspeaker responses taken as a whole like in [Staffeldt \(1974\)](#), [Toole \(1986\)](#), or [Gabrielsson et al. \(1991\)](#).

The comparison of the acoustical and perceptual spaces led to the same conclusion as the evaluations of the correlation between the acoustical and perceptual dissimilarities: By using an auditory model to discriminate recordings, we got closer to the perceptual dissimilarities, and to our goal aiming at a more relevant acoustical discrimination of loudspeakers. Among the tested acoustical analyses, only specific loudnesses could highlight our perceptual dimensions. [Klippel \(1990\)](#) and [Bramsløw \(2004\)](#) based their acoustical evaluation of sound reproduction on the specific loudness of their signals. They chose this method of analysis a priori,

assuming that it would lead to better results than other analyses. They obtained good results, but did not perform actual comparisons with other methods of analysis. Our results seem to validate their original hypothesis. Staffeldt (1974) tested loudspeaker measurements involving or not an auditory model. Even if he relied only on visual comparisons, using an auditory model led to a better fit between the shape of the measurement curves and the results of the listening tests. It seems crucial to take into account auditory masking in order to reveal the information used by listeners in their evaluation. The reason why measurements of frequency responses usually used to evaluate loudspeaker cannot be easily linked to the results of listening tests might be that these measurements cannot take into account these auditory phenomena.

It is difficult at this stage to precisely state which particular auditory phenomena should be taken into account or not. Thirteen acoustical analyses were tested with their associated metrics, but other signal analyses and metrics need to be investigated to refine the results. For example, the excitation pattern [Moore and Glasberg (1983)] of the recordings could be considered as an intermediate analysis between spectrum and specific loudness. The excitation pattern models the frequency resolution of the ear, without taking into account auditory masking effects. However, even if it remained to be tested, modeling only the frequency resolution of the ear might not be sufficient to fully describe the differences perceived by listeners. If it were the case, the overall specific loudness should have been successful in highlighting these differences. Overall specific loudness led to an acoustical space very similar to the perceptual one with the excerpt Kan'nida, but not with the two other musical excerpts (Fig. 5). The global analysis used while calculating overall specific loudness was not sufficient to accurately describe the perceptual dissimilarities. Our experiment indicated that it might be necessary to take into account the nonstationarity of auditory masking. We are not aware of any other study comparing different auditory models for the purpose of estimating acoustical dissimilarities. Klippel (1990) and Staffeldt (1974) only considered overall specific loudness, while Bramsløw (2004) only used a time-varying specific loudness. Further experiments are required to confirm our results. When Staffeldt (1974, 1984, 1991) directly applied an auditory model on the frequency response of loudspeakers to evaluate their sound reproduction, he did not take into account the fact that loudspeakers reproduce nonstationary musical signals and that the temporal evolution of these signals might influence the perception of listeners. If the importance of taking into account these nonstationary effects is confirmed in the future, the evaluation of loudspeakers by the direct examination of their frequency response might be questioned.

VI. CONCLUSION

This study investigated the perceived characteristics of the sound reproduction of 12 loudspeakers in a listening room. Paired comparisons were realized by using headphones and recordings of the sound reproductions. A MDS

analysis of the resulting perceptual dissimilarities revealed two main perceptual dimensions used by listeners to discriminate the loudspeakers. These dimensions were identical for three musical excerpts. The signals heard by the listeners were used to define acoustical attributes describing the perceptual dimensions. Instead of arbitrarily choosing one acoustical analysis, several analyses were compared. Only auditory models analyzing the signals in terms of their specific loudness allowed a good description of the differences perceived by the listeners. These models were used to define two acoustical attributes describing our perceptual dimensions: the bass/treble balance and the medium emergence.

ACKNOWLEDGMENTS

We wish to thank Dr. Jeremy Marozeau for providing us his MDS programme and for our very instructive discussions, Dr. Barrie Edmonds for his comments on an earlier version of this manuscript, the Mosquito Group and Genesis for lending us their loudspeakers, and all the listeners who took part in the experiment. The authors are grateful to the associate editor and the three anonymous reviewers for their helpful comments. The work of Mathieu Lavandier was supported by a grant from the Centre National de la Recherche Scientifique (C.N.R.S.) and the Région Provence-Alpes-Côtes d'Azur.

- AES20–1996 (1996). “AES recommended practice for professional audio—Subjective evaluation of loudspeakers,” *J. Audio Eng. Soc.* **44**, 382–400.
- Bech, S. (1994). “Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position,” *J. Audio Eng. Soc.* **42**, 999–1007.
- Bech, S. (1995). “Timbral aspects of reproduced sound in small rooms. I,” *J. Acoust. Soc. Am.* **97**, 1717–1726.
- Bech, S. (1996). “Timbral aspects of reproduced sound in small rooms. II,” *J. Acoust. Soc. Am.* **99**, 3539–3549.
- Bech, S. (2002). “Requirements for low-frequency sound reproduction, Part 1: The audibility of changes in passband amplitude ripple and lower system cutoff frequency and slope,” *J. Audio Eng. Soc.* **50**, 564–580.
- Borg, I., and Groenen, P. (1997). *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York).
- Bramsløw, L. (2004). “An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing,” *Acta. Acust. Acust.* **90**, 1007–1018.
- British Standard ISO 226 2003 (2003). “Acoustics—normal equal-loudness level contours,” BSi.
- Choi, S., and Wickelmaier, F. (2007). “Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference,” *J. Acoust. Soc. Am.* **121**, 388–400.
- Dillon, W. R., and Goldstein, M. (1984). *Multivariate Analysis: Methods and Applications* (Wiley, New York).
- Eisler, H. (1966). “Measurement of perceived acoustic quality of sound-reproducing systems by means of factor analysis,” *J. Acoust. Soc. Am.* **39**, 484–492.
- Fastl, H. (1997). “The psychoacoustics of sound-quality evaluation,” *Acta. Acust. Acust.* **83**, 754–764.
- Gabrielsson, A., Hagerman, B., Bech-Kristensen, T., and Lundberg, G. (1990). “Perceived sound quality of reproductions with different frequency responses and sound level,” *J. Acoust. Soc. Am.* **88**, 1359–1366.
- Gabrielsson, A., and Lindstrom, B. (1985). “Perceived sound quality of high-fidelity loudspeakers,” *J. Audio Eng. Soc.* **33**, 33–53.
- Gabrielsson, A., Lindström, B., and Till, O. (1991). “Loudspeaker frequency response and perceived sound quality,” *J. Acoust. Soc. Am.* **90**, 707–719.
- Gabrielsson, A., Rosenberg, U., and Sjögren, H. (1974). “Judgments and dimension analyses of perceived sound quality of sound-reproducing systems,” *J. Acoust. Soc. Am.* **55**, 854–861.
- Gabrielsson, A., and Sjögren, H. (1979). “Perceived sound quality of sound-reproducing systems,” *J. Acoust. Soc. Am.* **65**, 1019–1033.

- Guastavino, C., and Katz, B. F. G. (2004). "Perceptual evaluation of multidimensional spatial audio reproduction," *J. Acoust. Soc. Am.* **116**, 1105–1115.
- Guski, R. (1997). "Psychological methods for evaluating sound quality and assessing acoustic information," *Acta. Acust. Acust.* **83**, 765–774.
- IEC Publication 60268-13 (1998). "Sound system equipment—Part 13: Listening tests on loudspeakers," International Electrotechnical Commission, Geneva, Switzerland.
- Illényi, A., and Korpácssy, P. (1981). "Correlation between loudness and quality of stereophonic loudspeakers," *Acustica* **49**, 334–336.
- Klippel, W. (1990). "Multidimensional relationship between subjective listening impression and objective loudspeaker parameters," *Acustica* **70**, 45–54.
- Lavandier, M., Guyot, B., Meunier, S., and Herzog, P. (2005a). "The influence of stereophony on the restitution of timbre by loudspeakers," *Proceedings AES 119th Convention*, Paper No. 6619.
- Lavandier, M., Herzog, P., and Meunier, S. (2004). "Perceptual and physical evaluation of loudspeakers," *Proceedings AES 117th Convention*, Paper No. 6240.
- Lavandier, M., Herzog, P., and Meunier, S. (2008). "Comparative measurements of loudspeakers in a listening situation," *J. Acoust. Soc. Am.* **123**, 77–87.
- Lavandier, M., Meunier, S., and Herzog, P. (2005b). "Perceptual and physical evaluation of differences among a large panel of loudspeakers," *Forum Acusticum 2005*, Paper No. 430-0, pp. 1689–1694.
- Moore, B. C. J., and Glasberg, B. R. (1983). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *J. Acoust. Soc. Am.* **74**, 750–753.
- Moore, B. C. J., and Tan, C. T. (2003). "Perceived naturalness of spectrally distorted speech and music," *J. Acoust. Soc. Am.* **114**, 408–419.
- Olive, S. E. (2004a). "A multiple regression model for predicting loudspeaker preference using objective measurements: Part 1—Listening test results," *Proceedings AES 116th Convention*, Paper No. 6113.
- Olive, S. E. (2004b). "A multiple regression model for predicting loudspeaker preference using objective measurements: Part 2—Development of the model," *Proceedings AES 117th Convention*, Paper No. 6190.
- Olive, S. E., Schuck, P. L., Sally, S. L., and Bonneville, M. E. (1994). "The effect of loudspeaker placement on listener preference ratings," *J. Audio Eng. Soc.* **42**, 651–669.
- Paulus, E., and Zwicker, E. (1972). "Programme zur automatischen bestimmung der lautheit aus terzpegeln oder frequenzgruppenpegeln (Computer programs for calculating loudness from third octave band levels or from critical band levels)," *Acustica* **27**, 253–266.
- Pedersen, J. A., and Mäkivirta, A. (2002). "Requirements for low-frequency sound reproduction, Part 2: Generation of stimuli and listening system equalization," *J. Audio Eng. Soc.* **50**, 581–593.
- Pols, L. C. W., Van Der Kamp, L. J. T., and Plomp, R. (1969). "Perceptual and physical space of vowel sounds," *J. Acoust. Soc. Am.* **46**, 458–466.
- Rumsey, F., Zielinski, S., Kassier, R., and Bech, S. (2005a). "On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality," *J. Acoust. Soc. Am.* **118**, 968–976.
- Rumsey, F., Zielinski, S., Kassier, R., and Bech, S. (2005b). "Relationships between experienced listener ratings of multichannel audio quality and naive listener preferences," *J. Acoust. Soc. Am.* **117**, 3832–3840.
- Staffeldt, H. (1974). "Correlation between subjective and objective data for quality loudspeakers," *J. Audio Eng. Soc.* **22**, 402–415.
- Staffeldt, H. (1984). "Measurement and prediction of the timbre of sound reproduction," *J. Audio Eng. Soc.* **32**, 410–414.
- Staffeldt, H. (1991). "Differences in the perceived quality of loudspeaker sound reproduction caused by the loudspeaker-room-listener interactions," *Proceedings AES 90th Convention*, Paper No. 3046 (H-1).
- Susini, P., McAdams, S., and Winsberg, S. (1999). "A multidimensional technique for sound quality assessment," *Acta. Acust. Acust.* **85**, 650–656.
- Toole, F. E. (1985). "Subjective measurements of loudspeaker: sound quality and listener performance," *J. Audio Eng. Soc.* **33**, 2–32.
- Toole, F. E. (1986). "Loudspeaker measurements and their relationship to listener preferences: Part 2," *J. Audio Eng. Soc.* **34**, 323–348.
- Toole, F. E. (1991). "Binaural record/reproduction systems and their use in psychoacoustic investigations," *AES 91st Convention*, Paper No. 3179 (L-6).
- Zwicker, E., and Fastl, H. (1983). "A portable loudness-meter based on ISO 532B," *11th International Congress on Acoustics*, 135–137.
- Zwicker, E., and Fastl, H. (1999). *Psychoacoustics: Facts and Models* (Springer, New York).
- Zwicker, E., Fastl, H., and Dallmayr, C. (1984). "BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B," *Acustica* **55**, 63–67.