



HAL
open science

Comparative measurements of loudspeakers in a listening situation

Mathieu Lavandier, Philippe Herzog, Sabine Meunier

► **To cite this version:**

Mathieu Lavandier, Philippe Herzog, Sabine Meunier. Comparative measurements of loudspeakers in a listening situation. *Journal of the Acoustical Society of America*, 2008, 123 (1), pp.77-87. <10.1121/1.2816571>. <hal-00206058>

HAL Id: hal-00206058

<https://hal.science/hal-00206058v1>

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

Comparative measurements of loudspeakers in a listening situation

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

Laboratoire de Mécanique et d'Acoustique, C.N.R.S. UPR 7051, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France

(Received 25 May 2007; revised 28 September 2007; accepted 29 October 2007)

Comparison of loudspeakers is a major concern during design or product selection. There are several standards for the measurement of loudspeaker characteristics, but none of them provides hints for a rigorous comparison between devices. In this study, different ways of evaluating acoustical dissimilarity between loudspeakers were compared. Several methods of signal analysis were used, and for each method a metric evaluating the dissimilarity between two signals was defined. The correlation between the different dissimilarity evaluations over a significant panel of loudspeakers led to identified classes of measurements. A specific aspect of this work is that measurements were performed in a standard listening environment, rather than in an anechoic or reverberant one. It allowed the use of the recorded signals for a simple listening test, providing a perceptual metric which was compared to the acoustical ones. It also allowed the introduction of auditory models in the computation of some acoustical metrics, so defining a new class of measurements which gave results close to the perceptual ones. © 2008 Acoustical Society of America.

[DOI: 10.1121/1.2816571]

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

Pages: 77–87

I. INTRODUCTION

There are situations when studying loudspeakers where relatively slight differences must be quantified: parametric studies, classifications, or ranking of experimental or design factors. Dealing with loudspeakers requires that such comparative measurements use a rigorous protocol, especially when dealing with small changes, but also that their results correlate well with the differences which would be perceived by a listener. The need for such a tool might increase nowadays, as there is a trend to build active loudspeakers with embedded amplifiers and dedicated electronics, including linear filtering, adaptation to radiation conditions, and potentially dynamics or nonlinear processing. Studying such complex systems which involve many parameters usually requires the comparison of numerous samples (different loudspeakers and/or tunings). Associated measurements should be as fast as possible, and hopefully require widely available facilities.

This paper focuses on the comparison of loudspeakers by acoustical measurements, with a special interest for differences which can be perceived by average listeners. The long term objective of our work is to build a measurement tool for assessing the relevance of modifications in design or modeling of loudspeakers, in relation to their perceived influence. As a first step, our main interest was the timbre-related dissimilarities between loudspeakers, generally investigated in monophonic reproduction.

Timbre-related accuracy is a multidimensional characteristic of the perception of reproduced sound. It was studied in different ways depending on the aims of the researchers. Some researchers investigated the perceptual dimensions underlying the judgements of listeners.^{1–5} They either used dissimilarity evaluations and multidimensional scaling analysis,⁶ or absolute evaluations on chosen semantic scales and factor analysis.⁷ Other researchers investigated the fidelity judgments or preferences of listeners as a global absolute evaluation of loudspeakers,^{8–13} in order to look for their link with individual perceptual dimensions or acoustical measurements (see Toole¹⁴ for a review of this approach). Again, both our acoustical and perceptual approaches were focused on the relative differences between loudspeakers, but not on their absolute quality or the preference of listeners. The study presented in this paper identifies the most relevant acoustical discrimination of loudspeakers regarding perception, without evaluation along a quality scale or any particular perceptual scale.

Researchers and manufacturers have at their disposal different standards to evaluate their loudspeakers, physically by using acoustical measurements,^{15,16} or perceptually by using listening tests.^{17–21} The goal of a standard is that an absolute evaluation of a loudspeaker performed at a given time and location could be compared to the same measurement performed at another time or elsewhere. For this reason, standards define measurement environments, test signals, and measuring equipment. However, they allow different combinations of those as long as the actual measurement conditions are specified together with the results. This adds to the uncertainty of the measurement results, especially when they must be compared: some of the differences could be related to differences in the measurement protocol rather than to actual differences between devices. When relative

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

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

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

differences between loudspeakers are the main goal of the measurement, it is therefore better to use a single measurement setup so that changing the device under test is, to some extent, the only parameter that varies.

An important aspect of the measurements of loudspeakers is that their indirect goal is to quantify the quality of the sound reproduction, which is a very complicated notion, mixing (at least) physical, psychological, and cultural aspects. Listening tests on loudspeakers require facilities that are close to ordinary rooms. Such facilities are designed so that they provide a usual “look and feel” to the listener, and thus avoid bias listening habits. They must, however, ensure a limited noise and a controlled reverberation environment. More generally, they should not add a specific character to the sound reproduction. Beyond this need for a “natural” reproduction environment, a standardized listening environment ensures that the loudspeaker itself has the same behavior during its test and its normal use. The room may have a significant influence on the vibration of loudspeaker parts, and it has a major influence on the spatial repartition of the acoustic field.^{22,23} The frequency and dynamic content of the program excerpts also has some influence on the test results, as it may interact with nonlinear and memory effects.²⁴ Hence, the test signals have to be speech or musical excerpts chosen with great care. When dealing with a comparison of loudspeakers which are very similar (i.e., prototypes with slight design changes), selecting a realistic environment is more relevant than choosing one of the usual standardized acoustical measurement environments.

Searching for a relationship between measurements and perception is not a new topic. There is a general agreement among researchers on the fact that “the frequency response of the loudspeaker is the most important factor related to perceived sound quality,”²⁵ but there is less agreement on the most relevant way to measure this frequency response in order to link it with perception.²⁶ The timbre-related accuracy of a loudspeaker might also be influenced by its nonlinearities and its directivity, which mediates its interaction with the room.¹⁴ 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 to greater or lesser extents and only a posteriori. Measurements were done in listening rooms,^{2,27} or this environment was resimulated more or less completely from anechoic measurements.^{4,25,28} After taking into account the listening room, some measurements considered the influence of the musical excerpt,^{4,5,27} and the properties of the listener auditory system.^{2,4,5} In previous studies, either acoustical measurements were compared but their link with perceptual evaluation was not quantified,^{2,27,28} or the link between acoustical and perceptual evaluations was quantified but acoustical measurements were chosen a priori among several possibilities without comparing these different possibilities.^{4,5,13} The aim of our study was to compare different acoustical discriminations of loudspeakers, and quantify their link with a parallel perceptual evaluation.

In order to find a link between our acoustical and perceptual evaluations, and in front of the difficulty to do it a posteriori, we tried to keep these two approaches as close as

possible one to the other. The acoustical measurements had to be done in the same environment as the listening tests, and preferably at the same time, so we could be sure to measure the same sound field along both approaches. To comply with the constraints on listening tests,¹⁷⁻²¹ loudspeakers had to be compared in the same room, for the same positions of receiver (microphone/listener) and loudspeaker. To evaluate their relative differences, they also had to be compared one just after the other, due to our short auditory memory.¹¹ An experimental protocol compatible with both the acoustical and perceptual approaches was designed. It first consisted of recording the sound radiated by loudspeakers in a room, and then submitting the recorded sounds to signal analysis as well as to listening tests using headphones.

Of course this protocol suffers from restrictions. For example, the spatial dimension of reproduced sound and the interaction between loudspeaker and room cannot be reliably investigated using recordings and headphones; neither could an absolute evaluation of the sound reproduction such as preference or fidelity be undertaken with such protocol. We did not intend to evaluate all the potential dissimilarities between the loudspeakers tested, as some of these dissimilarities might not be present in the final recordings. Our aim was to study the dominant remaining dissimilarities, with the advantage of complying with the constraints mentioned above, of being sure that the remaining dissimilarities were associated with the loudspeakers and not another experimental parameter, and of having a direct access to the signals presented to the listeners. Previous studies dealing with the perception of reproduced sound compared listening tests realized live or using headphones.^{11,29-31} Even if the studies had various aims, they showed that the two types of listening tests gave very similar results concerning timbre-related accuracy.

Several methods of signal analysis were tested on our recordings, and for each method a metric evaluating the dissimilarity between two signals was defined. These acoustical dissimilarities were computed for the recordings, and were systematically compared to the perceptual dissimilarities resulting from the listening tests, in order to identify the most relevant method for the acoustical discrimination. We chose to base the acoustical approach on signal analysis done directly on the recordings used for the listening tests, rather than considering estimations of the responses of loudspeakers, in order to remain as close as possible to the signals heard and judged by listeners. Instead of measuring separately the frequency response, the directivity, and the nonlinear characteristics of the loudspeakers in the room, we considered a measurement involving a combination of these effects. Once the link with perception is established, separate measurements might be considered. Our approach followed the trend encountered in the literature, which showed that taking into account listening conditions helped to bridge the gap between measurement and perception. The listening conditions were then directly recorded.

This paper first presents 13 ways of calculating the acoustical dissimilarity between two signals. Our acoustical measurements of loudspeakers are then detailed. They are used to compare the different techniques of acoustical dis-

TABLE I. Definition of the acoustical dissimilarities ($\langle \rangle_{t,f,b}$: arithmetic mean over time t , frequency f , or Bark scale b , $\| \cdot \|$: modulus of a complex value, $\text{Min}\{ \}$: minimum value, $\text{Max}\{ \}$: maximum value, DFT: discrete Fourier transform, STFT: short-time Fourier transform, PSD: power spectral density, PSD_w : weighted power spectral density, Dens_o : overall specific loudness, Dens : time-varying specific loudness, Dens_m : temporal mean of time-varying specific loudness).

Method of analysis	Acoustical dissimilarity between signals $x(t)$ and $y(t)$
(a) Time signal	$\text{Min}\{\langle [x(t)-y(t)]^2 \rangle_t, \langle [x(t)+y(t)]^2 \rangle_t\}$
(b) Spectrum (complex)	$\text{Min}\{\langle \ \text{DFT}(x)-\text{DFT}(y)\ ^2 \rangle_f, \langle \ \text{DFT}(x)+\text{DFT}(y)\ ^2 \rangle_f\}$
(c) Time-frequency transform	$\text{Min}\{\langle \ \text{STFT}(x)-\text{STFT}(y)\ ^2 \rangle_{t,f}, \langle \ \text{STFT}(x)+\text{STFT}(y)\ ^2 \rangle_{t,f}\}$
(d) Modulus of spectrum	$\langle \ \text{DFT}(x)-\ \text{DFT}(y)\ ^2 \rangle_f$
(e) Modulus of time-frequency transform	$\langle \ \text{STFT}(x)-\ \text{STFT}(y)\ ^2 \rangle_{t,f}$
(f) Power spectral density (PSD)	$\langle [\sqrt{\text{PSD}(x)}-\sqrt{\text{PSD}(y)}]^2 \rangle_f$
(g) A weighted PSD	$\langle [\sqrt{\text{PSD}_w(x)}-\sqrt{\text{PSD}_w(y)}]^2 \rangle_f$
(h) 70 phone contour weighted PSD	$\langle [\sqrt{\text{PSD}_w(x)}-\sqrt{\text{PSD}_w(y)}]^2 \rangle_f$
(i) Overall specific loudness	$\langle \text{Max}\{\text{Dens}_o(x), \text{Dens}_o(y)\} / \text{Min}\{\text{Dens}_o(x), \text{Dens}_o(y)\} - 1 \rangle_b$
(j) Time-varying specific loudness 1	$\langle \langle \text{Max}\{\text{Dens}(x), \text{Dens}(y)\} / \text{Min}\{\text{Dens}(x), \text{Dens}(y)\} - 1 \rangle_b \rangle_t$
(k) Temporal mean of time-varying specific loudness 1	$\langle \text{Max}\{\text{Dens}_m(x), \text{Dens}_m(y)\} / \text{Min}\{\text{Dens}_m(x), \text{Dens}_m(y)\} - 1 \rangle_b$
(l) Time-varying specific loudness 2	$\langle \langle \text{Max}\{\text{Dens}(x), \text{Dens}(y)\} / \text{Min}\{\text{Dens}(x), \text{Dens}(y)\} - 1 \rangle_b \rangle_t$
(m) Temporal mean of time-varying specific loudness 2	$\langle \text{Max}\{\text{Dens}_m(x), \text{Dens}_m(y)\} / \text{Min}\{\text{Dens}_m(x), \text{Dens}_m(y)\} - 1 \rangle_b$

crimination. The acoustical dissimilarities are finally compared to the perceptual dissimilarities resulting from the listening tests. The multidimensional analysis of the measured dissimilarities is not considered here. It will be described in detail in a following paper.

II. ACOUSTICAL DISSIMILARITIES

A. Definitions

Thirteen evaluations of the acoustical dissimilarity between two signals were defined. Different methods of signal analysis were assessed. The temporal, spectral, and time-frequency domains were investigated. Two spectral weightings were tested. Different auditory models were also used. For each method of analysis, we defined a metric evaluating the acoustical dissimilarity between two signals. The different types of acoustical dissimilarities are presented in Table I.

Except for the analyses involving auditory models, the acoustical dissimilarities were all quadratic regarding the compared variables. The corresponding metrics were all defined in the same way in order to focus on the methods of analysis. Research is still ongoing testing other types of metrics. Each metric was defined to evaluate a global acoustical dissimilarity between two signals, giving a single value from the differences between the analyses of these two signals. To get a single scalar, dissimilarities calculated at each sample period or each frequency were integrated. As we could not suppose that one sample period or frequency region was more important than another, this integration was realized by an arithmetic mean, giving the same weight to all dissimilarities.

1. Time, spectral, and time-frequency domains

The spectral domain was investigated using the discrete Fourier transform (DFT) of the signals (b and d) and their power spectral density (PSD) (f, g, and h). The calculation of the power spectral density involved successive windows of 8192 samples with nonoverlapping Hanning windows. We checked that the dissimilarities obtained were largely inde-

pendent of the frequency resolution used for the computation. The time-frequency plane was obtained by calculation of the short-time Fourier transform (STFT) of the signals (c and e). It involved successive windows of 1024 samples, with overlapping Hanning windows.

We considered the spectral and time-frequency domains with (b and c) and without (d, e, f) phase information. When phase information was kept, and also in the time domain (a), a potential phase inversion between the two compared signals x and y was taken into account by considering y and $-y$.

2. Spectral weightings

Spectral weightings were applied to the power spectral density of our signals, to model the ear sensitivity. They were the A weighting (g) and a weighting based on the normal equal-loudness contour³² at 70 phones (h), which is close to the sound level used during our listening tests.

3. Auditory models

Auditory models are sometimes considered as perceptual analyses, because they were built using human perception. However, they are considered here as signal representations, because they are used like any other signal analysis technique, without any listener being involved. The only dissimilarities considered as perceptual will be those directly resulting from the listening tests.

Two auditory models were involved in our evaluations of acoustical dissimilarity. These models were proposed by Zwicker and Fastl³³ to calculate loudness, and are based on auditory masking. They were used to analyze our signals in terms of specific loudness which is the density of loudness along the Bark scale, which models the perceptual frequency scale. The first model^{34,35} 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 our signals, which is the specific loudness determined over the entire signal taken at once. The second model was designed for nonstationary sounds.^{33,36} It takes into account both frequency and temporal masking, calculating specific loudness every 10 ms. Two

time-frequency patterns of specific loudness were examined for our signals. 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 (j). For the second pattern, the second auditory model was applied. We called the resulting analysis time-varying specific loudness 2 (l). Finally, the temporal means of time-varying specific loudnesses 1 (k) and 2 (m) were considered. Compared to overall specific loudness, they took into account the fact that auditory masking depends on the spectral content of signals, and that this content varied over time for nonstationary musical signals.

Two loudnesses have to be compared by their ratio and not by their difference.³³ Thus, acoustical dissimilarities between two signals were computed by taking the ratio of their specific loudnesses (Table I). 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 ratios, a threshold was introduced in order to avoid artificially high dissimilarities where there was no useful signal but noise. The threshold was arbitrarily chosen at 0.03 sone/Bark. We also subtracted one from each ratio to make sure that two identical values of specific loudness would lead to a dissimilarity equal to zero. Integrations of ratios over time and Bark scales were done by taking the arithmetic mean of the dissimilarities over these two scales. Figure 1 illustrates our calculation of the dissimilarity between two specific loudnesses, taking the example of the overall specific loudness.

The definition of a metric from the output of the auditory models involved the estimation of a single scalar to quantify the difference between two signals. For the stationary analysis (i), the comparison data were straightforwardly integrated over the Bark scale. For the time-varying analyses, a temporal integration had to be performed. The auditory masking effects depend on the spectral content of the analyzed signals, and this content may vary over time. The metric may thus vary significantly when changing the order of the integrations. We tested the two possible orders. The two methods involving the time-varying specific loudnesses therefore differed only by the fact that the temporal integration was done as the first step for the temporal means of time-varying specific loudness (k) and (m), and as the last step for time-varying specific loudnesses (j) and (l). The combination of the two auditory models and two methods thus led to four different metrics, in addition to the stationary one.

B. Loudness equalization

Before computing the acoustical dissimilarities, signals had to be equalized in level, to prevent any sound level difference from creating an artificial dissimilarity between them. In the same way, any loudness differences during the listening tests would create uninteresting perceptual dissimilarities potentially masking more subtle ones.⁸ Before the listening tests and the acoustical analyses, the overall loudness of our signals was set to 70 phons. This equalization

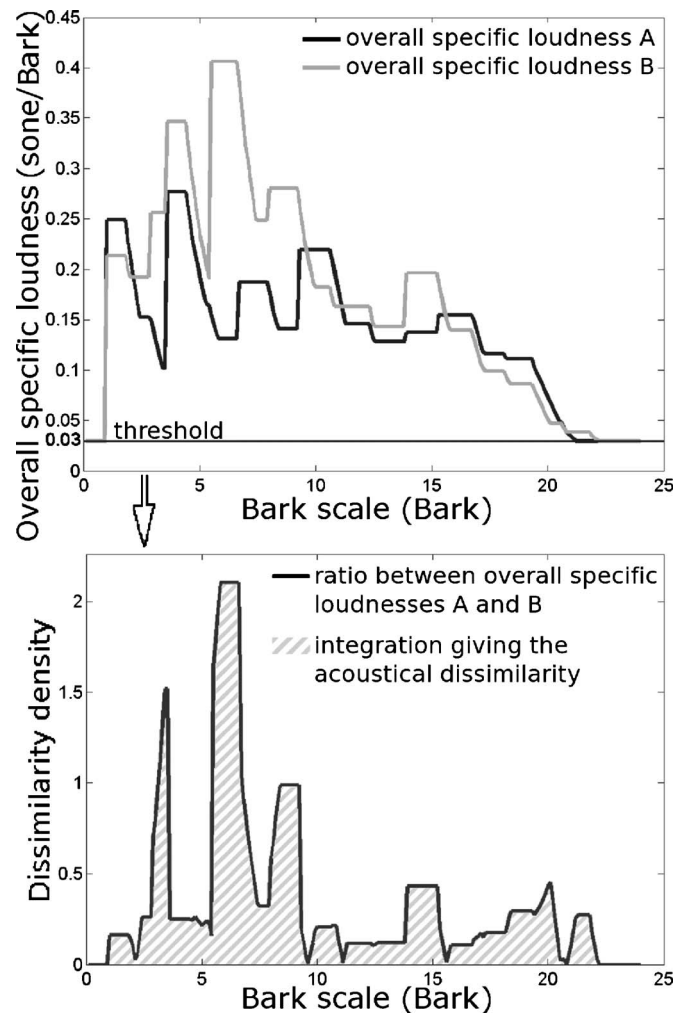


FIG. 1. Calculation of the acoustical dissimilarity between two overall specific loudnesses A and B.

was realized subjectively by the experimenters, and verified by computing the overall loudness of the equalized stimuli using a loudness estimation model.^{33,36}

C. Synchronization

In the same way as loudness was equalized to avoid uninteresting dissimilarities, our signals had to be synchronized. Time delays of a few sample periods between signals, although being crucial for the acoustical analyses, are not perceived during listening tests. Acoustical analyses should copy this behavior, and signals thus had to be synchronized in order to eliminate this irrelevant information. The evaluation of the time delay between two signals was not trivial. Our signals were not identical signals simply delayed in time: the musical stimuli used for the recording were filtered by the different loudspeakers and the room before being recorded by the microphones. The evaluation of the time delay between two signals depended on the frequency range considered, and so on the frequency responses of the loudspeakers involved. The reverberation and the first reflections recorded in the room further complicated the task.

Three methods evaluating the time delay between two signals were tested. The first one evaluated the pure delay

TABLE II. 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

contained in the transfer function between the two signals.³⁷ The minimum phase component of this transfer function was extracted by considering its complex cepstrum.³⁸ The time delay between the signals was estimated by a linear regression on the phase excess. The frequency range considered for this regression was restricted to the range containing useful signal, i.e., to the range for which the coherence between the signals was close to unity.³⁷ Even when the criterion of high coherence was verified, the time delays obtained by this method appeared to be very dependent on the frequency range considered, and the method was abandoned. The second method estimated the time delay between two signals by maximizing their cross-correlation function.³⁷ The third method consisted of minimizing their dissimilarity in the time domain (a, Table I). This dissimilarity is minimum for synchronized signals. Computing the dissimilarity for the first signal and delayed versions of the second signal allowed identifying the delay minimizing the dissimilarity, and compensating for this delay, thus synchronizing the two signals. The delay applied to the second signal was varied by steps of one sample period, testing positive and negative values of delay. Any potential phase inversion between the two signals was taken into account in the definition of the metric.

None of the tested methods allowed a perfect global synchronization of all our signals at the same time. We found some incoherent results such as the delay A-C between signals A and C being different from the sum of the delays A-B and B-C. The observed differences were limited to a few sample periods for the last two methods mentioned above. We could have chosen one signal as a reference, and synchronized all the signals with this unique reference, but then the signals might not have been perfectly synchronized with all the others signals. So, we chose to synchronize the signals taking them by pairs. The time delay between two signals was determined within one sample period by minimizing their dissimilarity in the time domain for each pair of signals, independently of the other pairs. Acoustical dissimilarities were then evaluated within each synchronized pair.

III. ACOUSTICAL MEASUREMENTS

The acoustical dissimilarities presented in the section above were tested on acoustical measurements of loudspeakers. These measurements consisted of recording the sound radiated by loudspeakers in a room, using different short musical excerpts. As the spatial component of sound reproduction could not be reliably investigated with our protocol, the loudspeakers were used in monophonic reproduction. “Timbre-related accuracy is much more easily heard in single-loudspeaker listening.”¹⁸

A. Loudspeakers

Twelve single loudspeakers were measured. We tried to use a wide range of models coming from different manufactures: high-fidelity column or bookshelves, studio monitoring, and computer monitoring. To get a reference for comparison, two loudspeakers of the same model were measured.

B. Room

The room used for the measurements 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.³⁹ The microphones were at 1 m from the floor and the point between the medium and tweeter of the loudspeakers was approximately at this same height. The floor was entirely carpeted. The reverberation time was evaluated at the position of the recording microphones, with the source at the loudspeaker position. The measurement microphone was omnidirectional and the source was a Genelec 1031A loudspeaker reproducing pink noise. The reverberation time was evaluated by measuring the time required for the sound level to decrease by 60 dB after the offset of the noise. The measured signals were filtered by third-octave bands and the reverberation time was evaluated for each band individually (Table II). Its value was found around 0.5 s at midrange frequencies, satisfying standard requirements.^{17,19}

C. Signals

Three musical excerpts were used for the recordings and we only kept a short part of them for the computation of the acoustical dissimilarities: (1) McCoy Tyner (“Miss Bea”, right channel, jazz, 3.3 s), (2) Kan’nida (“Konsyans”, left channel, percussion, 1.7 s), and (3) Vivaldi (“L’Europa Galante”, left channel, symphonic orchestra, 4.7 s). Such short excerpts were also suitable for the evaluation of perceptual dissimilarities between loudspeakers.^{40–42}

D. Procedure

Several recording techniques were tested and three were submitted to listeners: the stereophonic AB ORTF, stereophonic MS, and monophonic omnidirectional techniques.³⁹ These three types of recordings were chosen for their differences among the available recordings. The binaural technique using an artificial head was tested, but it did not allow a better externalization of our recordings involving frontal monophonic loudspeakers. In order to externalize frontal sources, individualized head related transfer functions (HRTFs) should be used as well as a head tracker to follow head movements. As our goal was to propose a sufficiently

easy way to compare loudspeakers, this technique was abandoned. As our study was focused on the timbre-related accuracy of the sound reproduction rather than its spatial component, the use of binaural techniques seemed less critical. The dissimilarity results from the three recording techniques submitted to listeners were very similar, so that the other techniques were not involved in listening tests, and only the results from the ORTF technique (arbitrarily chosen) are presented here. This technique involved 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 a 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 carried out directly with an audio workstation featuring a RME DIGI9652 sound card and an external Fostex VC-8 analog to digital (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 prior to the acoustical analyses and listening tests (see Sec. II B).

IV. COMPARISON OF ACOUSTICAL DISSIMILARITIES

The 13 types of acoustical dissimilarities (Table I) were computed on the recordings presented above. The right and left channels of these stereophonic recordings were analyzed separately. For each musical excerpt and each channel, we determined the acoustical dissimilarities among the 12 recordings corresponding to the 12 measured loudspeakers, leading to 13 66-dissimilarity vectors. The different acoustical dissimilarities were compared by evaluating the coefficient of correlation between these dissimilarity vectors.

A. Results

Figure 2 presents the coefficients of correlation between the acoustical dissimilarities calculated on the right and left channels of the recordings, for the three musical excerpts tested. The significant correlations are indicated in bold and marked with an asterisk. The individual significance level was fixed at 0.0005, following the conservative Bonferroni correction,⁴³ in order to keep the family wise significance level below 0.05 for the 78 tests of correlation realized on each recording channel. The right and left channels of the stereophonic recordings did not lead to exactly the same results, but the correlations followed the same trend with both channels. A perfect correlation was observed between the dissimilarities calculated in the time, spectral, and time-frequency domains [(a), (b) and (c); or (d), (e) and (f)], whereas weak or no correlations were obtained when comparing these domains with or without phase information [(a), (b), (c) and (d), (e), (f)]. The dissimilarities using phase information were not correlated with any of the other types of dissimilarity for the excerpt McCoy Tyner. The dissimilarities based on the two spectral weightings (g) and (h) were almost perfectly correlated. They were strongly correlated with the dissimilarities involving the spectral and time-

frequency domains without weightings nor phase information [(d), (e), and (f)] for the excerpt Vivaldi, but not for the two others musical excerpts. The dissimilarities based on the different specific loudnesses were strongly correlated [(i), (j), (k), (l) and (m)].

B. Discussion

The comparison of the acoustical dissimilarities showed that some dissimilarity evaluations were identical, but also that there were large differences among the tested methods, leading to identified classes of measurements.

It confirmed that the time, spectral, and time-frequency domains contained the same information. As our metrics were all defined in the same way from these different representations—quadratic regarding the compared variables (Table I)—they led to perfectly correlated dissimilarities. Conversely, taking into account or discarding phase information led to large differences in the discrimination of signals. Again, the calculations based on the power spectral density of signals led to the same results as the calculations based on the modulus of their spectrum or time-frequency transform: these are three representations of the same information. The A weighting and the weighting based on the normal equal-loudness contour at 70 phons gave almost identical results. For the excerpt Vivaldi, taking these weightings into account did not seem to change the information contained in the dissimilarities with no weighting. The different metrics using auditory models led to dissimilarities close to each others, but not identical, indicating that they might not contain exactly the same information.

The comparison of the acoustical dissimilarities showed that different types of dissimilarity contained different information, but did not show which one, if any, contained relevant information regarding perception. The acoustical dissimilarities were then compared to perceptual ones, in order to determine the analyses and metrics showing the differences heard by listeners.

V. RELEVANCE OF ACOUSTICAL DISSIMILARITIES REGARDING PERCEPTION

A. Perceptual dissimilarities

1. Procedure

The perceptual dissimilarities between the recordings were gathered during three listening tests, one for each musical excerpt. These tests were paired comparisons. 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.” This evaluation of the perceptual dissimilarities between the recordings was completely similar to the evaluation of their acoustical dissimilarities presented above.

Before the listening tests, the overall loudness of the recordings was set to 70 phons. This prevented loudness

McCoy Tyner

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
(a)	1.00*			0.22	0.23	0.21	0.03	0.04	0.07	0.10	0.11	0.13	0.14
(b)	1.00*	1.00*		0.22	0.23	0.21	0.02	0.04	0.07	0.10	0.11	0.13	0.14
(c)	1.00*	1.00*	1.00*	0.22	0.23	0.21	0.02	0.04	0.07	0.10	0.11	0.13	0.14
(d)	0.03	0.03	0.03	1.00*	1.00*	1.00*	0.18	0.23	0.42*	0.52*	0.55*	0.51*	0.58*
(e)	0.08	0.08	0.08	0.98*	1.00*	1.00*	0.21	0.26	0.43*	0.52*	0.54*	0.53*	0.57*
(f)	0.04	0.04	0.04	0.99*	0.98*	1.00*	0.18	0.23	0.42*	0.50*	0.54*	0.50*	0.57*
(g)	0.08	0.08	0.08	0.65*	0.68*	0.59*	1.00*	0.97*	0.60*	0.66*	0.48*	0.73*	0.43*
(h)	0.08	0.08	0.08	0.65*	0.68*	0.58*	0.98*	1.00*	0.61*	0.69*	0.52*	0.75*	0.46*
(i)	0.05	0.05	0.05	0.62*	0.61*	0.56*	0.74*	0.73*	1.00*	0.95*	0.96*	0.90*	0.92*
(j)	0.06	0.06	0.06	0.71*	0.71*	0.65*	0.80*	0.81*	0.96*	1.00*	0.94*	0.95*	0.90*
(k)	0.07	0.07	0.07	0.65*	0.63*	0.60*	0.70*	0.69*	0.98*	0.96*	1.00*	0.87*	0.98*
(l)	0.13	0.13	0.13	0.71*	0.75*	0.66*	0.84*	0.84*	0.89*	0.96*	0.88*	1.00*	0.83*
(m)	0.06	0.06	0.06	0.67*	0.64*	0.63*	0.71*	0.70*	0.96*	0.94*	0.99*	0.87*	1.00*

Kan'nida

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
(a)	1.00*			0.54*	0.53*	0.52*	0.35	0.44*	0.57*	0.54*	0.48*	0.61*	0.54*
(b)	1.00*	1.00*		0.54*	0.53*	0.52*	0.35	0.44*	0.57*	0.54*	0.48*	0.61*	0.54*
(c)	1.00*	1.00*	1.00*	0.54*	0.53*	0.52*	0.35	0.44*	0.57*	0.54*	0.48*	0.61*	0.54*
(d)	0.48*	0.49*	0.49*	1.00*	1.00*	1.00*	0.23	0.38	0.57*	0.51*	0.51*	0.66*	0.62*
(e)	0.48*	0.48*	0.49*	1.00*	1.00*	1.00*	0.23	0.38	0.57*	0.50*	0.50*	0.65*	0.61*
(f)	0.47*	0.47*	0.47*	1.00*	1.00*	1.00*	0.24	0.38	0.60*	0.53*	0.53*	0.67*	0.64*
(g)	0.30	0.30	0.30	0.18	0.18	0.19	1.00*	0.98*	0.65*	0.79*	0.76*	0.77*	0.68*
(h)	0.39	0.39	0.39	0.34	0.33	0.34	0.98*	1.00*	0.70*	0.81*	0.79*	0.82*	0.73*
(i)	0.52*	0.52*	0.52*	0.41	0.40	0.43*	0.73*	0.77*	1.00*	0.92*	0.90*	0.90*	0.87*
(j)	0.50*	0.50*	0.50*	0.39	0.38	0.41	0.73*	0.76*	0.98*	1.00*	0.99*	0.96*	0.96*
(k)	0.47*	0.47*	0.47*	0.39	0.38	0.40	0.70*	0.73*	0.97*	0.99*	1.00*	0.94*	0.97*
(l)	0.55*	0.55*	0.55*	0.57*	0.56*	0.58*	0.75*	0.79*	0.94*	0.96*	0.94*	1.00*	0.95*
(m)	0.48*	0.48*	0.48*	0.53*	0.52*	0.55*	0.65*	0.70*	0.92*	0.96*	0.97*	0.95*	1.00*

Vivaldi

	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)	(m)
(a)	1.00*			0.36	0.36	0.40	0.36	0.37	0.30	0.34	0.38	0.31	0.37
(b)	1.00*	1.00*		0.36	0.36	0.40	0.36	0.37	0.30	0.34	0.38	0.31	0.37
(c)	1.00*	1.00*	1.00*	0.36	0.36	0.40	0.36	0.37	0.31	0.34	0.38	0.31	0.37
(d)	0.48*	0.48*	0.48*	1.00*	1.00*	1.00*	0.96*	0.96*	0.72*	0.83*	0.75*	0.85*	0.76*
(e)	0.49*	0.49*	0.49*	1.00*	1.00*	1.00*	0.96*	0.96*	0.72*	0.84*	0.76*	0.85*	0.76*
(f)	0.48*	0.48*	0.48*	1.00*	1.00*	1.00*	0.97*	0.97*	0.71*	0.83*	0.75*	0.84*	0.76*
(g)	0.45*	0.45*	0.45*	0.96*	0.96*	0.97*	1.00*	0.99*	0.72*	0.80*	0.74*	0.81*	0.74*
(h)	0.48*	0.48*	0.49*	0.97*	0.97*	0.97*	0.99*	1.00*	0.74*	0.82*	0.76*	0.82*	0.77*
(i)	0.54*	0.54*	0.54*	0.68*	0.69*	0.67*	0.60*	0.63*	1.00*	0.94*	0.96*	0.92*	0.95*
(j)	0.49*	0.49*	0.49*	0.75*	0.76*	0.73*	0.69*	0.70*	0.96*	1.00*	0.98*	0.99*	0.97*
(k)	0.52*	0.52*	0.52*	0.63*	0.64*	0.62*	0.56*	0.58*	0.97*	0.97*	1.00*	0.95*	0.99*
(l)	0.47*	0.47*	0.47*	0.78*	0.79*	0.76*	0.72*	0.73*	0.94*	0.99*	0.94*	1.00*	0.95*
(m)	0.54*	0.54*	0.54*	0.69*	0.70*	0.68*	0.63*	0.65*	0.96*	0.96*	0.99*	0.94*	1.00*

from creating uninteresting dissimilarities, potentially masking more subtle ones. Listening tests were run in an isolated soundproof room, using a Tucker&Davis workstation and Stax SR Lambda Professional headphones.

2. Participants

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. They were all otologically normal. They were members of the laboratory or students. None of them had significant previous experience in loudspeaker comparison.

3. Results

For each 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 the listeners. This verification was realized by a cluster analysis of the similarities between listeners, with the similarity between two listeners being evaluated by the correlation of their judgments.⁴⁴ Multidimensional analyses of the perceptual dissimilarities are not presented here, as the topic of this paper is to evaluate the relevance of the acoustical dissimilarities regarding perception. Perceptual dissimilarities are used here as a reference for this evaluation.

Figure 3 presents the perceptual dissimilarities measured for the 66 pairs of recordings and the three musical excerpts. These dissimilarities ranged from 0.22 to 0.83 (McCoy Tyner), 0.13 to 0.90 (Kan'nida), and 0.21 to 0.84 (Vivaldi), indicating that the recordings of the chosen loudspeakers and musical excerpts led to a broad range of dissimilarities. This was further verified for each listening test by comparing the different values of dissimilarity. Paired-sample *t* tests were

FIG. 2. Correlations between the acoustical dissimilarities calculated on the right and left channels of the recordings, for the three musical excerpts (McCoy Tyner, Kan'nida, Vivaldi): (a) time signal; (b) spectrum; (c) time-frequency transform; (d) modulus of spectrum; (e) modulus of time-frequency transform; (f) power spectral density (PSD); (g) *A* weighted PSD; (h) 70 phone contour weighted PSD; (i) overall specific loudness; (j) time-varying specific loudness 1; (k) temporal mean of time-varying specific loudness 1; (l) time-varying specific loudness 2; and (m) temporal mean of time-varying specific loudness 2. The significant correlation coefficients are in bold, marked with an asterisk.

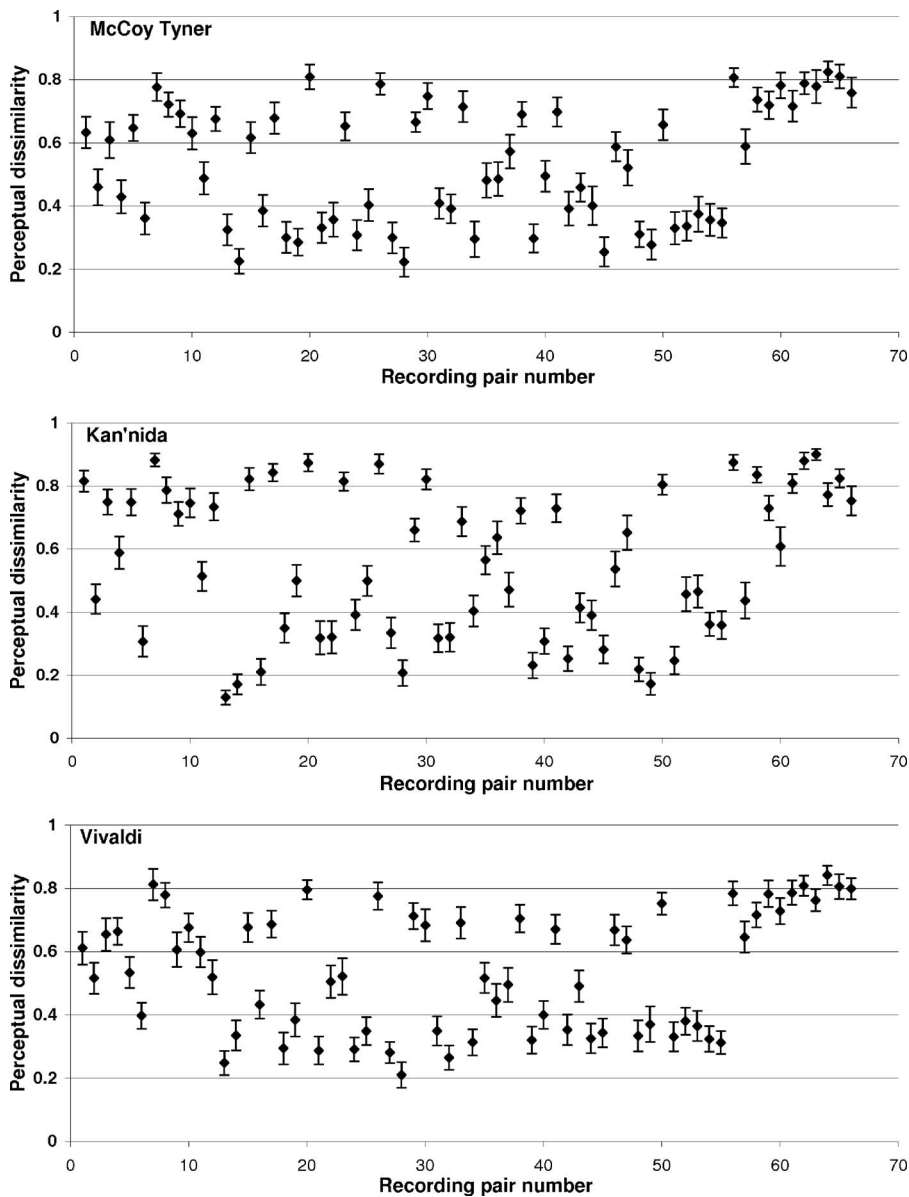


FIG. 3. Mean perceptual dissimilarities with standard errors measured for the 66 pairs of recordings and the three musical excerpts (McCoy Tyner, Kan'nida, Vivaldi).

used to investigate the significance of the differences between these values. The individual significance level of each t test was fixed at 0.00002, following the Bonferroni correction, in order to keep the family wise significance level below 0.05 for the 2145 comparisons realized for each musical excerpt. These statistical analyses confirmed that 621 dissimilarities were significantly different for the excerpt McCoy Tyner. The number of significantly different dissimilarities was 876 for the excerpt Kan'nida and 672 for the excerpt Vivaldi.

B. Correlation between acoustical and perceptual dissimilarities

For each musical excerpt, the perceptual dissimilarities averaged over all listeners resulted in a 66-dissimilarity vector. This vector was used as a reference to assess the relevance of the acoustical discrimination methods regarding perception. This assessment was realized by evaluating the coefficient of correlation between the acoustical and perceptual dissimilarity vectors. Although the absolute values of

these correlations might be contaminated by noise and non-linearities, the values obtained with the different acoustical methods could be compared to determine which acoustical dissimilarities were the closest to the judgements of listeners.

1. Results

Figure 4 presents the coefficients of correlation between the mean perceptual dissimilarities resulting from the listening tests and the acoustical dissimilarities calculated on the right and left channels of the recordings using each of the 13 acoustical discrimination methods, for the three musical excerpts involved. The significant correlations are marked with an asterisk. The individual significance level was fixed at 0.001, following the Bonferroni correction, in order to keep the family wise significance level below 0.05 for the 26 tests of correlation realized for each musical excerpt. The correlations obtained with the acoustical dissimilarities calculated on the right and left channels of the recordings followed the same trend. The effects were also similar for the three musical excerpts. Dissimilarities based on the time (a), spectral

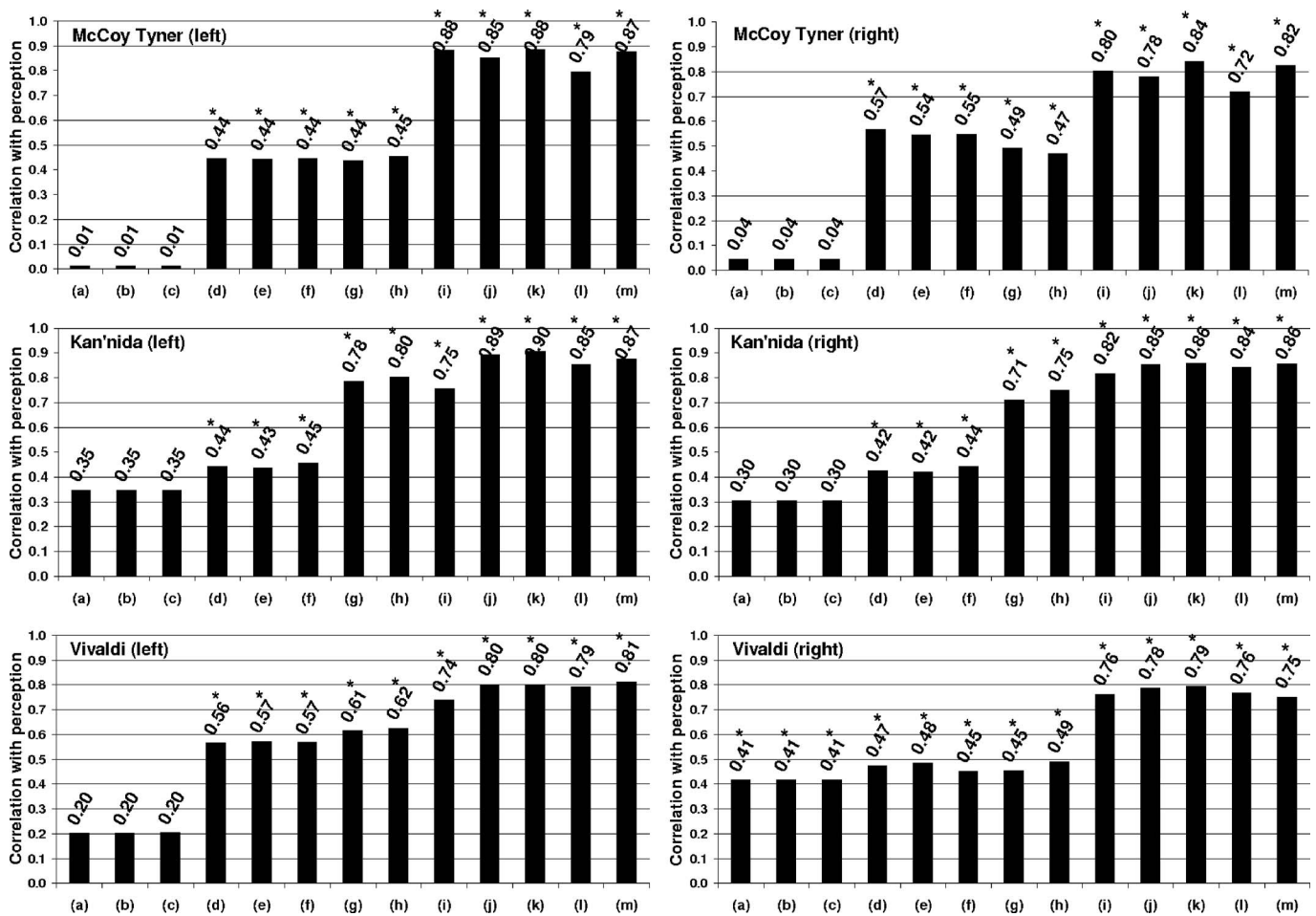


FIG. 4. Correlations between the mean perceptual dissimilarities resulting from the listening tests and the acoustical dissimilarities calculated on the right and left channels of the recordings, for the three musical excerpts (McCoy Tyner, Kan'nida, Vivaldi): (a) time signal; (b) spectrum; (c) time-frequency transform; (d) modulus of spectrum; (e) modulus of time-frequency transform; (f) power spectral density (PSD); (g) A weighted PSD; (h) 70 phone contour weighted PSD; (i) overall specific loudness; (j) time-varying specific loudness 1; (k) temporal mean of time-varying specific loudness 1; (l) time-varying specific loudness 2; and (m) temporal mean of time-varying specific loudness 2. The significant correlation coefficients are marked with an asterisk.

(b), and time-frequency (c) domains were never correlated with perceptual dissimilarities, except for the dissimilarities calculated on the right channel of the excerpt Vivaldi. When acoustical dissimilarities were evaluated using the modulus of the spectrum (d), the one of the time-frequency transform (e), or the power spectral density (f), the correlations were significant but weak. The weightings applied to the power spectral density [(g) and (h)] led to an increase of the correlation for the musical excerpt Kan'nida, but this increase was not confirmed for the two other excerpts. Using specific loudnesses to evaluate acoustical dissimilarities greatly enhanced the correlation with perceptual dissimilarities [(i), (j), (k), (l), and (m)].

2. Discussion

The acoustical discrimination methods which were almost perfectly correlated (Fig. 2) gave almost identical results when compared to perceptual dissimilarities. Discarding phase information improved the correlation with perceptual dissimilarities, whereas the tested spectral weightings did not seem to offer any further improvement, except for the excerpt Kan'nida. The high correlations obtained with the methods involving auditory models showed that the cor-

responding acoustical dissimilarities contained at least part of the relevant information regarding perception. It should be noted that perfect correlations would have meant that acoustical dissimilarities accounted for the noise in the evaluation of perceptual dissimilarities, which would be highly undesirable. The present data did not discriminate between the different methods using auditory models. Overall and mean specific loudnesses seemed to contain as much useful information towards perception as time-varying specific loudnesses. The temporal evolution of information did not seem to be crucial to explain our perceptual dissimilarities.

VI. GENERAL DISCUSSION

By defining metrics associated with several signal analyses, we calculated acoustical dissimilarities between musical signals radiated by loudspeakers. These acoustical dissimilarities were compared to perceptual dissimilarities resulting from listening tests, in order to identify the most relevant method of acoustical discrimination. The correlation between acoustical and perceptual dissimilarities was used to quantify the link between the two types of evaluation. This criterion of correlation was severe, as we could not suppose a priori that a linear link would exist. Yet, the high values of corre-

lation obtained revealed that such a link was established for the cases tested. The acoustical dissimilarities based on auditory models were close to the perceptual ones and contained at least part of the relevant information regarding perception. They were closer to the perceptual dissimilarities than to the other acoustical dissimilarities tested, bridging to some extent the gap between acoustical measurements and perception.

Other signal analyses and metrics should be tested to refine these results. For example, the acoustical dissimilarities considered in this study only used monaural information, as the left and right channels of the recordings were treated separately. The potential phase differences between the two ears/microphones were not incorporated in the dissimilarity evaluation. Such differences might be important to take into account if perceptual dimensions such as “feeling of space”^{3,4} or “spaciousness”⁵ were used by listeners to discriminate the recordings.

The contribution of auditory models was quantified by an increment in correlation, so we went a little further than Staffeldt,^{2,10,45} who justified the use of such a model by visual comparisons between the results of his listening tests and the shape of the frequency responses of his loudspeakers. Bramsløw⁵ and Klippel⁴ based their acoustical evaluation of sound reproduction on the specific loudness of their signals. They chose this method of analysis a priori, by 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 to test their hypothesis.

The correlation between acoustical and perceptual dissimilarities led to the same trends with three different musical excerpts. They remained relatively independent of the experimental parameters tested in following experiments considering other recording techniques,³⁹ another room, up to 37 loudspeakers, and another listening task.⁴⁶

Other comparisons of acoustical and perceptual dissimilarities should be investigated. A linear link has been found by evaluating their correlation, but this relationship might be refined. A multidimensional comparison of our acoustical and perceptual dissimilarities was therefore performed,^{39,47} and will be fully described in a following paper. It allowed a comparison without an implicit linear relationship, but it only relied on visual comparisons of multidimensional spaces, without quantitative evaluation of their resemblance to one another. This comparison indicated that our acoustical analyses using specific loudnesses led to dimensions very similar to the perceptual ones used by listeners to discriminate our recordings during the listening tests.

Whereas specific loudnesses 1 and 2 and their temporal mean gave similar results, the overall specific loudness did not seem sufficient to fully describe the perceptual dissimilarities.³⁹ This difference between specific loudnesses determined on the overall signal or as a function of time could not be quantified by the correlations presented in this paper (Fig. 4). Klippel⁴ and Staffeldt² only considered overall specific loudness, while Bramsløw⁵ only used a time-varying specific loudness. More experimental data are required to make a clear difference between these acoustical evaluations. This question is of great importance to assess

the necessity of taking into account the temporal dependency of auditory masking effects. When Staffeldt^{2,10,45} applied an auditory model directly 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 direct examination of their frequency response might be questioned.

VII. CONCLUSION

Different ways of evaluating the acoustical dissimilarity between loudspeakers were compared. Twelve single loudspeakers were measured in a listening situation. The correlation between the comparative measurements led to identified classes of measurements. The recorded signals were involved in listening tests, providing a perceptual metric used to evaluate the relevance of the acoustical discrimination methods regarding perception. Our first results showed the importance of using auditory models in order to discriminate the loudspeakers in the same way as listeners did.

ACKNOWLEDGMENTS

We wish to thank 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 on a first version of this paper. 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.

¹A. Gabrielsson, U. Rosenberg, and H. Sjögren, “Judgments and dimension analyses of perceived sound quality of sound-reproducing systems,” *J. Acoust. Soc. Am.* **55**, 854–861 (1974).

²H. Staffeldt, “Correlation between subjective and objective data for quality loudspeakers,” *J. Audio Eng. Soc.* **22**, 402–415 (1974).

³A. Gabrielsson and H. Sjögren, “Perceived sound quality of sound-reproducing systems,” *J. Acoust. Soc. Am.* **65**, 1019–1033 (1979).

⁴W. Klippel, “Multidimensional relationship between subjective listening impression and objective loudspeaker parameters,” *Acustica* **70**, 45–54 (1990).

⁵L. Bramsløw, “An objective estimate of the perceived quality of reproduced sound in normal and impaired hearing,” *Acta. Acust. Acust.* **90**, 1007–1018 (2004).

⁶I. Borg and P. Groenen, *Modern Multidimensional Scaling. Theory and Applications* (Springer, New York, 1997).

⁷W. R. Dillon and M. Goldstein, *Multivariate Analysis. Methods and Applications* (Wiley, New York, 1984).

⁸A. Gabrielsson and B. Lindstrom, “Perceived sound quality of high-fidelity loudspeakers,” *J. Audio Eng. Soc.* **33**, 33–53 (1985).

⁹F. E. Toole, “Subjective measurements of loudspeaker: sound quality and listener performance,” *J. Audio Eng. Soc.* **33**, 2–32 (1985).

¹⁰H. Staffeldt, “Differences in the perceived quality of loudspeaker sound reproduction caused by the loudspeaker-room-listener interactions,” *Proceedings AES 90th Convention*, 1991, p. 3046.

¹¹S. E. Olive, P. L. Schuck, S. L. Sally, and M. E. Bonneville, “The effect of loudspeaker placement on listener preference ratings,” *J. Audio Eng. Soc.* **42**, 651–669 (1994).

¹²S. E. Olive, “Differences in performance and preference of trained versus untrained listeners in loudspeaker tests: a case study,” *J. Audio Eng. Soc.* **51**, 806–825 (2003).

- ¹³S. E. Olive, "A multiple regression model for predicting loudspeaker preference using objective measurements: Part 1 - Listening test results," *Proceedings AES 116th Convention*, 2004, p. 6113.
- ¹⁴F. E. Toole, "Loudspeakers and rooms for sound reproduction—A scientific review," *J. Audio Eng. Soc.* **54**, 451–476 (2006).
- ¹⁵IEC Publication 60268-5, "Sound system equipment - Part 5: Loudspeakers," *International Electrotechnical Commission*, Geneva, Switzerland, 1989.
- ¹⁶IEC Publication 60581-7, "High fidelity audio equipment and systems. Minimum performance requirements - Part 7: Loudspeakers," *International Electrotechnical Commission*, Geneva, Switzerland, 1986.
- ¹⁷IEC Publication 60268-13, "Sound system equipment - Part 13: Listening tests on loudspeakers," *International Electrotechnical Commission*, Geneva, Switzerland, 1998.
- ¹⁸AES20-1996, "AES recommended practice for professional audio - Subjective evaluation of loudspeakers," *J. Audio Eng. Soc.* **44**, 382–400 (1996).
- ¹⁹ITU-R Recommendation BS.1116-1, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," *International Telecommunication Union*, Geneva, Switzerland, 1997.
- ²⁰ITU-R Recommendation BS.1534, "Method for the subjective assessment of intermediate quality levels of coding systems," *International Telecommunication Union*, Geneva, Switzerland, 2003.
- ²¹ITU-R Recommendation BS.1284, "General methods for the subjective assessment of sound quality," *International Telecommunication Union*, Geneva, Switzerland, 2003.
- ²²R. V. Waterhouse, "Radiation impedance of a source near reflectors," *J. Acoust. Soc. Am.* **35**, 1144–1151 (1963).
- ²³T. Salava, "Acoustic load and transfer functions in rooms at low frequencies," *J. Audio Eng. Soc.* **36**, 763–775 (1988).
- ²⁴W. Klippel, "Nonlinear large-signal behaviour of electrodynamic loudspeakers at low frequencies," *J. Audio Eng. Soc.* **40**, 402–415 (1992).
- ²⁵S. E. Olive, "A multiple regression model for predicting loudspeaker preference using objective measurements: Part 2 - Development of the model," *Proceedings AES 117th Convention*, 2004, p. 6190.
- ²⁶F. E. Toole, "Loudspeaker measurements and their relationship to listener preferences: Part 1," *J. Audio Eng. Soc.* **34**, 227–235 (1986).
- ²⁷A. Gabriëlsson, B. Lindström, and O. Till, "Loudspeaker frequency response and perceived sound quality," *J. Acoust. Soc. Am.* **90**, 707–719 (1991).
- ²⁸F. E. Toole, "Loudspeaker measurements and their relationship to listener preferences: Part 2," *J. Audio Eng. Soc.* **34**, 323–348 (1986).
- ²⁹F. E. Toole, "Binaural record/reproduction systems and their use in psychoacoustic investigations," *Proceedings AES 91st Convention*, 1991, p. 3179.
- ³⁰S. Bech, "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 (2002).
- ³¹J. A. Pedersen and A. Mäkiavirta, "Requirements for low-frequency sound reproduction, Part 2: Generation of stimuli and listening system equalization," *J. Audio Eng. Soc.* **50**, 581–593 (2002).
- ³²British Standard ISO 226 2003, "Acoustics - Normal equal-loudness level contours," BSI, London, 2003.
- ³³E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models* (Springer, New York, 1999).
- ³⁴E. Paulus and E. Zwicker, "Programme zur automatischen bestimmung der lautheit austerzpegeln oder frequenzgruppenpegeln (Computer programs for calculating loudness from third octave band levels or from critical band levels)," *Acustica* **27**, 253–266 (1972).
- ³⁵E. Zwicker, H. Fastl, and C. Dallmayr, "BASIC-Program for calculating the loudness of sounds from their 1/3-oct band spectra according to ISO 532 B," *Acustica* **55**, 63–67 (1984).
- ³⁶E. Zwicker and H. Fastl, "A portable loudness-meter based on ISO 532B," *Proceedings 11th International Congress on Acoustics*, 1983, pp. 135–137.
- ³⁷J. Max and J.-L. Lacoume, *Méthodes et Techniques de Traitement du Signal (Methods and Techniques for Signal Processing)* (Dunod, Paris, 2000).
- ³⁸A. V. Oppenheim and R. W. Schaffer, *Digital Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ, 1975).
- ³⁹M. Lavandier, P. Herzog, and S. Meunier, "Perceptual and physical evaluation of loud-speakers," *Proceedings AES 117th Convention*, 2004, p. 6240.
- ⁴⁰S. Bech, "Timbral aspects of reproduced sound in small rooms. 1," *J. Acoust. Soc. Am.* **97**, 1717–1726 (1995).
- ⁴¹S. Bech, "Timbral aspects of reproduced sound in small rooms. 2," *J. Acoust. Soc. Am.* **99**, 3539–3549 (1996).
- ⁴²B. C. J. Moore and C. T. Tan, "Perceived naturalness of spectrally distorted speech and music," *J. Acoust. Soc. Am.* **114**, 408–419 (2003).
- ⁴³G. Keppel and T. D. Wickens, *Design and Analysis. A Researcher's Handbook*, 4th ed. (Pearson Prentice-Hall, Englewood Cliffs, NJ, 2004).
- ⁴⁴M. Lavandier, "Différences entre enceintes acoustiques: Une évaluation physique et perceptive (Differences between Loudspeakers: A physical and perceptual evaluation)," Ph.D. thesis, Université de la Méditerranée - Aix-Marseille II, 2005, (<http://tel.archives-ouvertes.fr/tel-00087414>).
- ⁴⁵H. Staffeldt, "Measurement and prediction of the timbre of sound reproduction," *J. Audio Eng. Soc.* **32**, 410–414 (1984).
- ⁴⁶M. Lavandier, S. Meunier, and P. Herzog, "Perceptual and physical evaluation of differences among a large panel of loudspeakers," *Proceedings Forum Acusticum 2005*, 2005, pp. 1689–1694.
- ⁴⁷M. Lavandier, P. Herzog, and S. Meunier, "Physical and perceptual estimation of differences between loudspeakers," *C. R. Mec.* **334**, 732–736 (2006).