Third-octave analyses describing two perceptual dimensions of sound reproduction and the resulting overall perceived dissimilarity between loudspeakers or headphones

Nathan Szwarcberg, Mathieu Lavandier

Third-octave analyses describing two perceptual dimensions of sound reproduction and the resulting overall perceived dissimilarity between loudspeakers or headphones

Nathan Szwarcberg[1,2] and Mathieu Lavandier[1,a]

[1]*ENTPE, Ecole Centrale de Lyon, CNRS, LTDS, UMR5513, 69518 Vaulx-en-Velin, France*

[2]*Aix Marseille Univ, CNRS, Centrale Marseille, LMA, 13013 Marseille, France*

Many objective measurements have been proposed to evaluate sound reproduction, but it is often difficult to link measured differences with the differences perceived by listeners. In the literature, the best correlations with perception were obtained for measures involving an auditory model. The present study investigated simpler measurements to highlight the signal processing steps required to make the link with perception. It is based on dissimilarity evaluations from two previous studies: one comparing 12 single loudspeakers using three musical excerpts, one comparing 21 headphones using two musical excerpts; both studies highlighting two perceptual dimensions associated with the relative strengths of bass and midrange. The objective approach compared several signal analyses computing the dissimilarity between the spectra of the recorded sound reproductions. The results show that a third-octave analysis can accurately describe the overall dissimilarity between the loudspeakers or headphones, and the two underlying perceptual dimensions.

[a]mathieu.lavandier@entpe.fr

I. INTRODUCTION

The evaluation of sound reproduction can rely on listening tests involving a panel of listeners, but those are time consuming, thus expensive, and many biases need to be controlled for the tests to be valid (AES20-1996, 1996; Bech and Zacharov, 2006; IEC Publication 60268-13, 1998). For loudspeaker evaluation, seeing the loudspeakers could generate a bias, the listening room and the positions of the loudspeaker and listener are important (Bech, 1994; Olive *et al.*, 1994), sound level influences quality judgments (Illényi and Korpássy, 1981), the musical excerpts used for the evaluation also matter (Eisler, 1966; Gabrielsson *et al.*, 1974), and our short auditory memory further complicates the evaluation of relative differences between loudspeakers (Klippel, 1990; Olive *et al.*, 1994). Thus, many researchers have looked for objective measurements to evaluate sound reproduction. Several measures can be undertaken (ANSI/CTA Standard 2034-A, 2015; IEC Publication 60268-5, 1989; IEC Publication 60581-7, 1986), but it is difficult to link differences highlighted in these measures with the differences that will be perceived by listeners. Various approaches have been proposed to make this link: visual comparisons between loudspeaker frequency responses (the interpretation of which involves many simultaneous criteria) and their evaluations along a given perceptual scale (Gabrielsson *et al.*, 1991; Staffeldt, 1974; Toole, 1986b) relying on "expert" eyes and thus differing from one author to the other (Olive, 2004a); "black box" models taking acoustical measurements as inputs and producing an output correlated with a perceptual rating (Bramsløw, 2004; Olive, 2004b), the overall link being then quantified but this approach not elucidating the link of each input measure with perception; correlations

between acoustical attributes derived from measurements and perceptual dimensions (Klippel, 1990; Lavandier *et al.*, 2008b; Michaud *et al.*, 2015; Olive, 2004a; Volk *et al.*, 2016). To bridge the gap between standard measurements and perception, researchers defined measures that consider the listening conditions to various extents: measures done directly in listening rooms (Gabrielsson *et al.*, 1991; Lavandier *et al.*, 2008b; Michaud *et al.*, 2015; Staffeldt, 1974), or re-simulating such environments more or less completely from anechoic measurements (Klippel, 1990; Olive, 2004b; Toole, 1986b). Some measures also consider the influence of the musical excerpt (Bramsløw, 2004; Gabrielsson *et al.*, 1991; Klippel, 1990; Lavandier *et al.*, 2008b; Michaud *et al.*, 2015; Volk *et al.*, 2016). Finally, the best correlations with perception were obtained for objective measures involving an auditory model (Bramsløw, 2004; Klippel, 1990; Lavandier *et al.*, 2008a,b; Michaud *et al.*, 2015; Staffeldt, 1974; Volk *et al.*, 2016). Even if such models are more and more available to manufacturers to develop their loudspeakers and headphones, it would be interesting to understand which signal processing steps in the auditory models are required to make the link with perception. To investigate this question, the aim of the present study was to compare different measurements and highlight the simplest measure that still correlates well with perceptual evaluations.

The basic audio quality (BAQ) of a system or signal-processing scheme is defined as a global attribute resulting from all dissimilarities between this system and a reference, which corresponds to an original undistorted sound relative to which the BAQ ratings are given (Schoeffler and Herre, 2016), where BAQ would result from a combination of perceptual attributes (Bech and Zacharov, 2006). However, there is no obvious ideal sound

reproduction that could be used as a reference to evaluate the BAQ of headphones or loudspeakers (Bramsløw, 2004). Instead, three main categories of characteristics have been highlighted concerning the perceptual evaluation of these sound-reproducing systems: the different perceptual dimensions underlying timbre-related accuracy (also called restitution of timbre, sound quality, or fidelity; Gabrielsson and Sjögren, 1979), the spatial qualities involving stereophony and multichannel reproductions, and the dynamics/distortion characteristics related to different levels of solicitation of the systems (AES20-1996, 1996; IEC Publication 60268-13, 1998). Several methods are available for the perceptual evaluation. It can be based on absolute ratings along particular scales, such as preference (Olive, 2003, 2004a; Olive *et al.*, 1994; Ravizza *et al.*, 2023), perceived quality or fidelity (Gabrielsson and Lindström, 1985; Toole, 1985), or specific attributes of reproduced sound (Bramsløw, 2004; Gabrielsson and Sjögren, 1979; Klippel, 1990; Staffeldt, 1974). One can also evaluate the relative dissimilarity between systems that are directly compared (Klippel, 1990; Lavandier *et al.*, 2008b; Volk *et al.*, 2016). Multidimensional scaling techniques can then be used to reveal the perceptual dimensions underlying these dissimilarity ratings, without listeners having to name or even be aware of what they were experiencing while listening (Eisler, 1966; Gabrielsson *et al.*, 1974; Klippel, 1990; Lavandier *et al.*, 2008b). Because of the data the present study is based on, it investigated timbre-related accuracy for untrained listeners who evaluated relative differences between single loudspeakers or headphones. Both overall dissimilarity and the underlying perceptual dimensions were considered, looking for simple objective measures able to describe them.

The present study is based on dissimilarity evaluations from two previous studies: one comparing 12 single loudspeakers in a listening room using three musical excerpts (Lavandier *et al.*, 2008a,b), one comparing 21 headphones using two musical excerpts (Volk *et al.*, 2016). Both studies highlighted two similar perceptual dimensions associated with the relative strengths of bass and midrange. Note that these dimensions seem characteristic of sound reproduction in general, as they were obtained for loudspeakers and headphones, as well as for a larger panel of 37 loudspeakers representative of a wide range of realistic sound systems (not considered here as it also involved a spatial dimension not related to timbre-related accuracy; Michaud *et al.*, 2015). The first dimension could be associated with the "brightness", "balance between bass and treble", "sharpness", "fullness" and "spectral balance" from other studies (Bramsløw, 2004; Gabrielsson *et al.*, 1974; Gabrielsson and Sjögren, 1979; Klippel, 1990; Olive, 2004a); while the second dimension could be associated with "clarity", "distinctness" and "clearness" (Bramsløw, 2004; Gabrielsson *et al.*, 1974; Gabrielsson and Sjögren, 1979; Klippel, 1990). To circumvent the experimental and psychological biases mentioned above, the sound reproductions of the loudspeakers and headphones were recorded and the recordings were compared using headphones, as done in other studies on timbre-related accuracy (Bech, 2002; Olive *et al.*, 1994; Pedersen and Mäkivirta, 2002; Toole, 1991).

Despite the general agreement that the frequency response is the most important factor related to timbre-related accuracy (Olive, 2004b), there is less agreement on the most relevant way to measure this response to link it with perception (Toole, 1986a). We chose to base our objective approach on signal analyses done directly on the recordings of the

sound reproductions rather than considering estimations of the loudspeakers/headphones responses, to remain as close as possible to the signals compared by the listeners. Lavandier *et al.* (2008b) and Volk *et al.* (2016) already described their two perceptual dimensions using an auditory model applied to these recordings, while Lavandier *et al.* (2008a) showed that this also allows to describe the overall dissimilarities within their loudspeakers. Here, we intended to replicate these results, extend them to describe the overall dissimilarities among the headphones, while highlighting the signal processing steps required to make the link with perception.

Instead of choosing one signal analysis a priori, three signal analyses were considered to evaluate the spectrum of reproduced sound, and for each analysis, three metrics allowing to compute an objective dissimilarity between two recording spectra were tested. The analysis proposed by Lavandier *et al.* (2008a) that uses an auditory model and its associated metric was also carried out as a reference. First, the overall dissimilarities between loudspeakers/headphones were considered, identifying the signal analyses and metrics leading to the best correlation with perceptual dissimilarities, for all musical excerpts. The spaces resulting from multidimensional scaling of the objectives and perceptual dissimilarities were then compared, thus investigating which objective analyses were able to account for the underlying perceptual dimensions. Finally, these analyses were used to define acoustical attributes describing each dimension.

II. PERCEPTUAL DATA USED TO TEST THE OBJECTIVE MEASURES

Lavandier *et al.* (2008a,b) made stereophonic recordings of 12 single loudspeakers in a listening room. The loudspeakers were chosen to represent the diversity of the audio market at the time. Three short musical excerpts were recorded on each loudspeaker, named respectively "Kan'Nida" (percussions, 1.7 s, maximum energy around 100 Hz), "McCoy Tyner" (jazz, 3.3 s, two spectral peaks at 100 Hz and 600 Hz) and "Vivaldi" (baroque orchestra, 4.7 s, broad spectrum from 200 to 2000 Hz). Such short excerpts were reported as suitable for the evaluation of perceived differences in timbre-related accuracy (Bech, 1995, 1996; Moore and Tan, 2003), and ensured that all untrained listeners base their judgment on the same part of the original excerpt (Volk *et al.*, 2016). The signals were sampled at 44.1 kHz. The third-octave spectra of the original excerpts are shown in the supplementary material [1] (Suppl. Fig. 1). Twenty-seven participants undertook three listening tests, one per musical excerpt, in a soundproof room using Stax SR Lambda Professional headphones, the overall loudness of the recordings being equalized to 70 phons. During each test, the 12 recordings were presented to the participant in pairs. For each pair, the overall dissimilarity was rated on a scale from 0 to 1, corresponding to "very similar" and "not similar at all", respectively. Dissimilarity ratings were averaged across listeners, after ensuring through cluster analysis that there were no subgroups with different rating strategies. The top panels of Fig. 1 show the two-dimensional spaces resulting from a multidimensional scaling (MDS[2]; Borg and Groenen, 1997) analysis applied to these perceptual dissimilarities. The proportion of variance accounted for by the MDS analysis was 93%, 86%, and 85% for

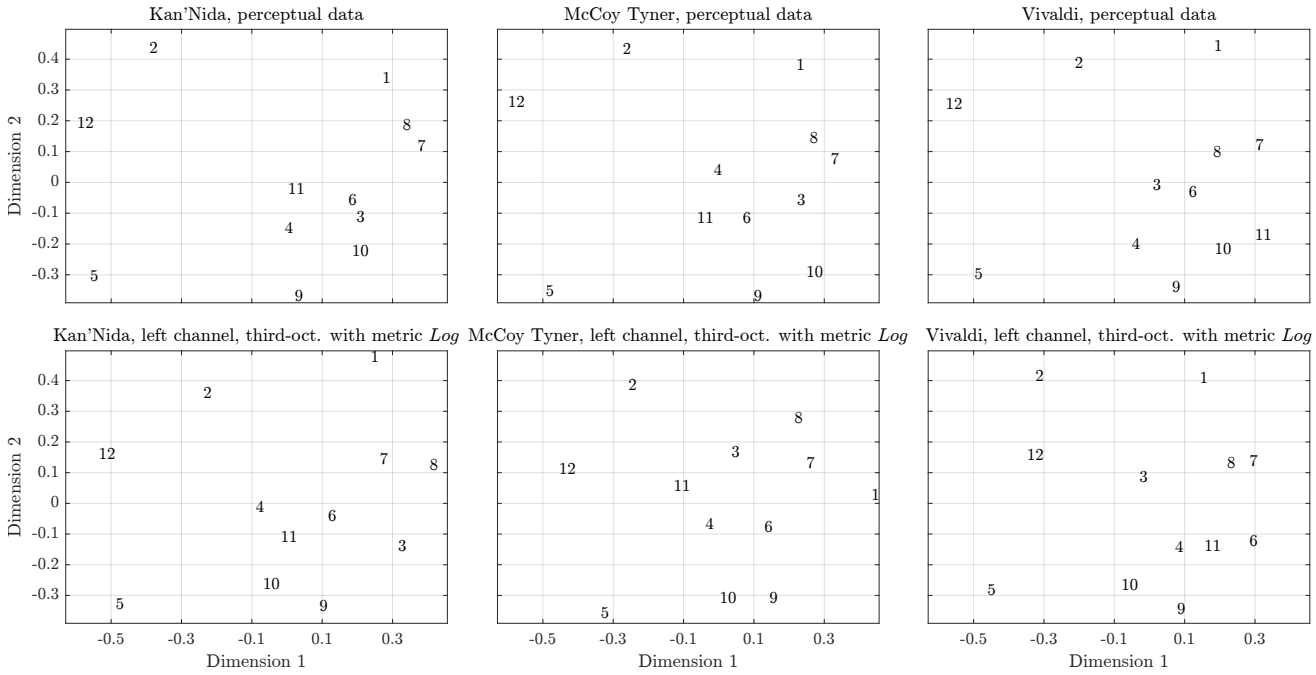

FIG. 1. Two-dimensional spaces resulting from the MDS analysis of the dissimilarities obtained for the three musical excerpts in the loudspeakers study (Lavandier *et al.*, 2008a,b). Top panels: perceptual dissimilarities. Bottom panels: objective dissimilarities computed with the third-octave analysis and the metric *Log* (defined in Table I). Each number represents a loudspeaker. The 95%-confidence ellipses corresponding to each loudspeaker are presented in Suppl. Fig. 2[1].

Kan'Nida, McCoy Tyner, and Vivaldi, respectively. When the individual dimensions are considered in isolation, dimension 1 accounts for 77%, 60%, and 55% of the variance for these excerpts, respectively; while dimension 2 accounts for 18%, 18%, and 20% of the variance.

Lavandier *et al.* (2008a,b) did not try to deliver the same signals to the ears of the listeners as they would have experienced in the room with the actual loudspeakers. They used stereophonic recordings rather than binaural recordings with a dummy head that was

not available to them at the time. Moreover, they did not try to compensate for the influence of the recording microphones and playback headphones. However, those remained the same for all loudspeaker comparisons that were focused on relative dissimilarity rather than an absolute evaluation of quality. They did not aim to highlight all the dissimilarities between the tested loudspeakers, as some of these dissimilarities were probably not captured. In particular, the spatial component of sound reproduction could not be reliably investigated, and their study was focused on timbre-related accuracy. The protocol offered the advantage that the remaining dissimilarities were associated with the loudspeakers under evaluation. The perceptual dimensions obtained even seem characteristic of loudspeaker reproduction in general: they were very similar for the three musical excerpts, and also remained unchanged using other recording techniques (Lavandier *et al.*, 2004), reproduction modes (Lavandier *et al.*, 2005), another listening room (Michaud *et al.*, 2015), and a much larger panel of loudspeakers (Michaud *et al.*, 2015).

Volk *et al.* (2016) recorded 21 pairs of electrodynamic headphones on a binaural dummy head, at a sampling rate of 48 kHz. They were a mix of open- and closed-back headphones, with circumaural and supra aural models spanning a large price range. Two musical excerpts were recorded: "Todd Terje" (electronic music, 1.9 s, most of the energy between 50 and 100 Hz, and above 7 kHz, see Suppl. Fig. 1[1]), and "Tina Dickow" (soft pop, 4.5 s, most of the energy between 200 and 1000 Hz, and above 7 kHz). The influences of the mannequin's ear canal and of the playback headphones used for the listening tests were compensated for. One reference signal was added to the 21 recordings for the listening tests: the original musical excerpt without processing that was directly reproduced with the playback headphones.

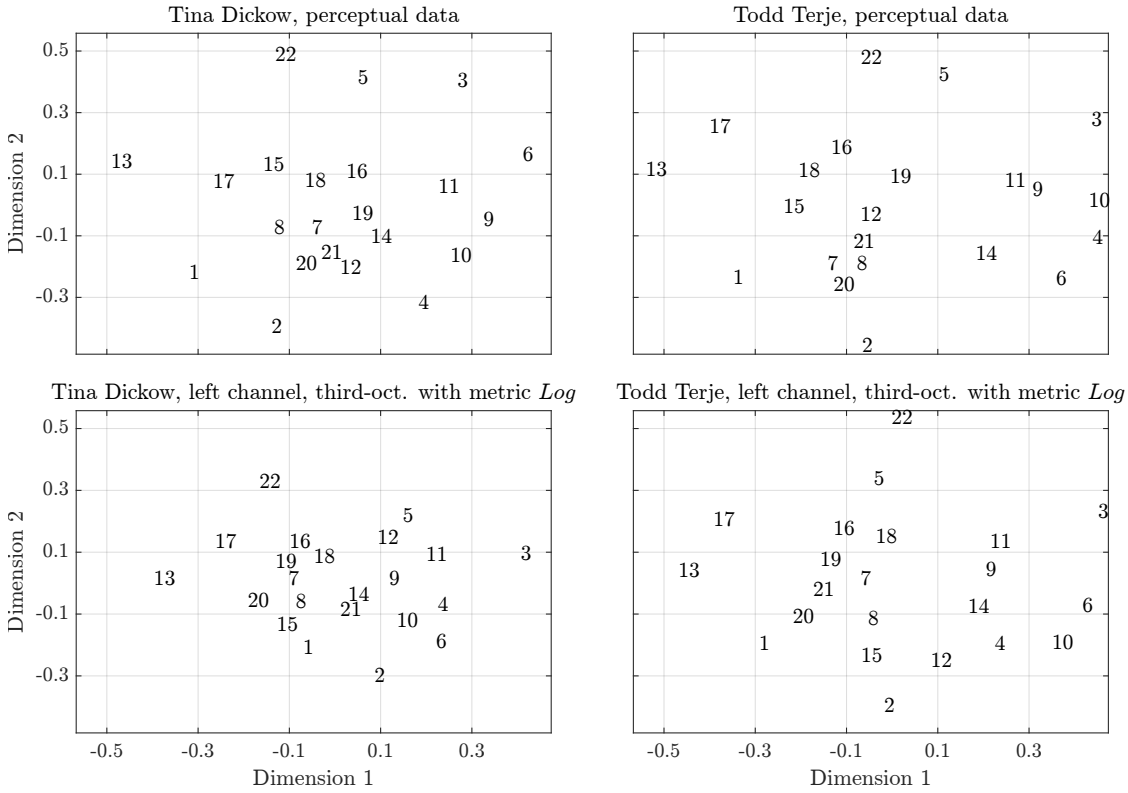

FIG. 2. Same as Fig. 1 but for the headphones study (Volk *et al.*, 2016) (for the 95%-confidence ellipses, see Suppl. Fig. 3[1])

This last signal was calibrated to 72 dB SPL, while the other stimuli were adjusted to the same loudness. For each listening test, one per musical excerpt, the 22 stimuli were reproduced by the playback (Sennheiser HD 650) headphones in a sound proof booth and compared by pairs to evaluate their dissimilarities. Fifteen listeners participated in each test. Dissimilarity ratings were averaged across listeners. The two-dimensional perceptual spaces resulting from the MDS analysis of these dissimilarities are presented on the top panels of Fig. 2. The proportion of variance accounted for by the MDS analysis was 78% and 77% for Todd Terje and Tina Dickow, respectively. When the individual dimensions are considered in isolation, dimension 1 accounts for 49% and 40% of the variance for these

excerpts, respectively; while dimension 2 accounts for 16% and 26% of the variance. The compensation for the influence of the recording process and playback headphones appears validated by the fact that the playback headphones reproducing the original excerpts stand very close to recorded headphones of the same model in the perceptual spaces (headphones 12 and 21 in Fig. 2).

In both studies, the stimuli were equalized in loudness to prevent overall loudness differences to dominate the dissimilarity evaluations and mask more subtle dissimilarities. Even if the perceptual verification of the equalization was only done informally by the experimenters, the results indicate that the equalization was successful, because overall loudness never came out of the MDS analysis as a criterion used by the listeners to discriminate the recordings.

III. OBJECTIVE ANALYSES, METRICS AND ATTRIBUTES

The recordings of the sound reproduction systems were compared by pairs in the frequency domain. Before comparing two spectra, the corresponding recordings were synchronized within one sample period by minimizing the quadratic distance of their derivatives in the time domain (Lavandier *et al.*, 2008a). This time alignment was realized for each pair of recordings to be compared, independently of the other pairs. Acoustical dissimilarities were then evaluated within each synchronized pair in the spectral domain. The right and left channels of the recordings were analyzed separately. Figure 3 presents an overview of the analyses used to compute these objective dissimilarities and compare them with perception, as detailed in the rest of this section.

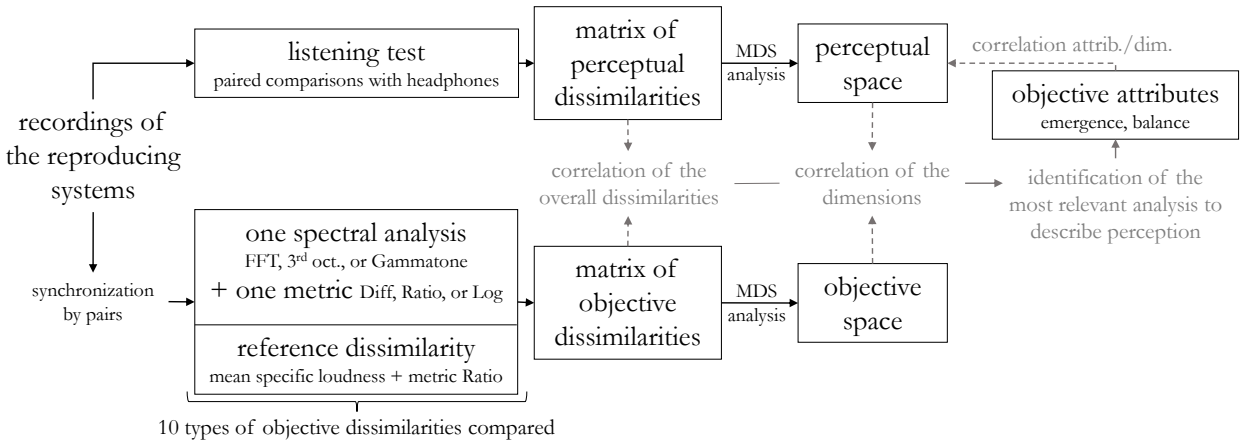

FIG. 3. Overview of the analyses applied for each panel of sound reproduction systems (loudspeakers, headphones), musical excerpt, and recording channel

1. Objective analyses

Three signal analyses were used to compare the spectral information in the recordings. Each spectral analysis is applied to a temporal waveform (the channel of a recording) $s(t)$, with a sampling frequency F_s. Its frequency spectrum is denoted $\hat{s}(f)$, expressed in Pa^2, where f is the frequency.

The first analysis is the FFT spectrum, resulting from a one-sided discrete fast Fourier transform. The frequency bands are linearly spaced and have all the same width of F_s/N, where N is the number of samples of the waveform. The maximum frequency is $F_s/2$, i.e. 22.05 kHz for the loudspeakers study and 24 kHz for the headphones study.

The second analysis is the third-octave spectrum computed in Pa^2. The frequency range of this analysis is $[22, 14031]$ Hz, decomposed into 28 bands of center frequency f_i and frequency width $[2^{-1/6} f_i, 2^{1/6} f_i]$.

The third analysis computes the signal energy at the outputs of a gammatone filter bank (Patterson *et al.*, 1987) with two filters per Equivalent Rectangular Bandwidth (Moore and Glasberg, 1983). The number of filters is determined by the sampling frequency. In the loudspeakers study ($F_s = 44.1$ kHz), 73 filters are used with center frequencies between 16 Hz and 21 kHz. For the headphones study ($F_s = 48$ kHz), 75 filters are used with center frequencies between 16 Hz and 27 kHz.

2. Metrics

For each spectral analysis, three metrics were defined to assess the overall dissimilarity between two spectra $\hat{a}(f)$ and $\hat{b}(f)$. The three metrics $Diff$, $Ratio$ and Log are presented in Table I, in which the spectra $\hat{a}$ and $\hat{b}$ are expressed in Pa^2 and result from one of the three spectral analyses (section III 1).

TABLE I. Metrics $Diff$, $Ratio$ and Log computing the overall dissimilarity between the spectra $\hat{a}$ and $\hat{b}$ defined on N_b frequency bands with center frequencies f_i (FFT bands, third-octave bands, or gammatone bands)

$Diff(\hat{a},\hat{b})$	$\frac{1}{N_b}\sum_{i=1}^{N_b}\lvert\hat{a}(f_i)-\hat{b}(f_i)\rvert$
$Ratio(\hat{a},\hat{b})$	$\frac{1}{N_b}\left[\sum_{i=1}^{N_b}\frac{\max(\hat{a}(f_i),\hat{b}(f_i))}{\min(\hat{a}(f_i),\hat{b}(f_i))}\right]-1$
$Log(\hat{a},\hat{b})$	$\frac{1}{N_b}\sum_{i=1}^{N_b}10\log_{10}\left[\frac{\max(\hat{a}(f_i),\hat{b}(f_i))}{\min(\hat{a}(f_i),\hat{b}(f_i))}\right]$

3. Reference dissimilarity

In addition to the nine dissimilarity estimates obtained by combining the three spectral analyses and metrics presented above, an estimator of dissimilarity based on an auditory model was computed. This reference corresponds to the dissimilarity that best correlated with perceived dissimilarity in the loudspeakers study (Lavandier *et al.*, 2008a). It is calculated on the temporal mean of the time-varying specific loudness of the recordings, computed using the model of Zwicker and Fastl (1983, 2013) to account for auditory masking. The specific loudness is calculated every 10 ms and averaged across time. The dissimilarity is computed using the metric *Ratio* applied on the average specific loudnesses (Lavandier *et al.*, 2008a).

4. Comparison of the overall dissimilarities

For each study (loudspeakers or headphones), musical excerpt and channel of the recordings (right or left), ten types of objective dissimilarity were computed (3 analyses $\times$ 3 metrics + 1 reference) and compared to the perceptual dissimilarities from the listening test. The first comparison was done by computing the Pearson correlation coefficient between the objective and perceptual dissimilarities. For the second comparison, MDS analysis was applied to both the objective and perceptual dissimilarities, and the objective space was rotated to best match the corresponding perceptual space using a generalized procrustes analysis procedure (Lavandier *et al.*, 2008b; Volk *et al.*, 2016). These spaces were then compared by computing the Pearson correlation coefficient of their dimensions.

5. *Objective attributes describing the perceptual dimensions*

After considering the overall dissimilarities between loudspeakers and headphones, the two dimensions underlying these dissimilarities were considered. The aim was to define objective attributes that could describe these dimensions, using the simplest spectral analysis that was found relevant to describe the overall dissimilarities in the first part of the study. Let $\hat{s}(f_i)$ be the spectrum $\hat{s}$ within the frequency band centered on f_i. Two types of attribute were investigated, based on previous attribute definitions using auditory models proposed by Lavandier *et al.* (2008b) and Volk *et al.* (2016) to describe the same perceptual dimensions. The first type of attribute, called *Emergence*: $E([f_i, f_j])$, is defined as the ratio of the energy in the frequency range $[f_i, f_j]$ to the energy of the full spectrum:

$$E([f_i, f_j]) = 10 \log_{10} \frac{\sum_{k=i}^{j} \hat{s}(f_k)}{\sum_{k=1}^{lastband} \hat{s}(f_k)}. \tag{1}$$

The second type of attribute, called *Balance*: $B\left(\frac{[f_i,f_j]}{[f_k,f_l]}\right)$, is the ratio of the energy in the frequency range $[f_i, f_j]$ to the energy in the range $[f_m, f_n]$:

$$B\left(\frac{[f_i, f_j]}{[f_m, f_n]}\right) = 10 \log_{10} \frac{\sum_{k=i}^{j} \hat{s}(f_k)}{\sum_{k=m}^{n} \hat{s}(f_k)}, \tag{2}$$

with non-overlapping frequency ranges, i.e. $i < j < m < n$.

These attributes were defined by testing all possible frequency ranges and keeping the definition leading to the best Pearson correlation with the coordinates of the stimuli along the perceptual dimensions (Volk *et al.*, 2016). The best value between the left and right

ear correlations was considered, assuming that listeners could discriminate the stimuli using their best ear.

IV. RESULTS

A. Overall dissimilarities

1. Correlation with the perceptual dissimilarities

Figures 4 and 5 present the correlations[3] between the objective and perceptual overall dissimilarities. Because 100 correlations are considered, their individual significance level was Bonferroni corrected to $0.05/100 = 0.0005$. For the loudspeakers study (Fig. 4), correlations follow similar trends across the two channels of the three musical excerpts. The metric $Diff$, regardless of the spectral analysis, leads to dissimilarities non-significantly correlated with perception for at least two musical excerpts. This is also the case for the FFT analysis with the metric *Ratio*. The metric *Log* leads to dissimilarities significantly correlated with perception for all musical excerpts and recording channels, for the three spectral analyses. This is also true for the reference analysis and the metric *Ratio* with the third-octave and gammatone analyses. For the headphones study (Fig. 5), the correlations seem on average weaker (not statistically tested here), but this could at least partly be explained by the fact that they are computed on more stimuli. These correlations follow the same trends than for the loudspeakers, except that the dissimilarities obtained with the metric $Diff$ are significantly correlated with perception, even if the corresponding correlations are sometimes low.

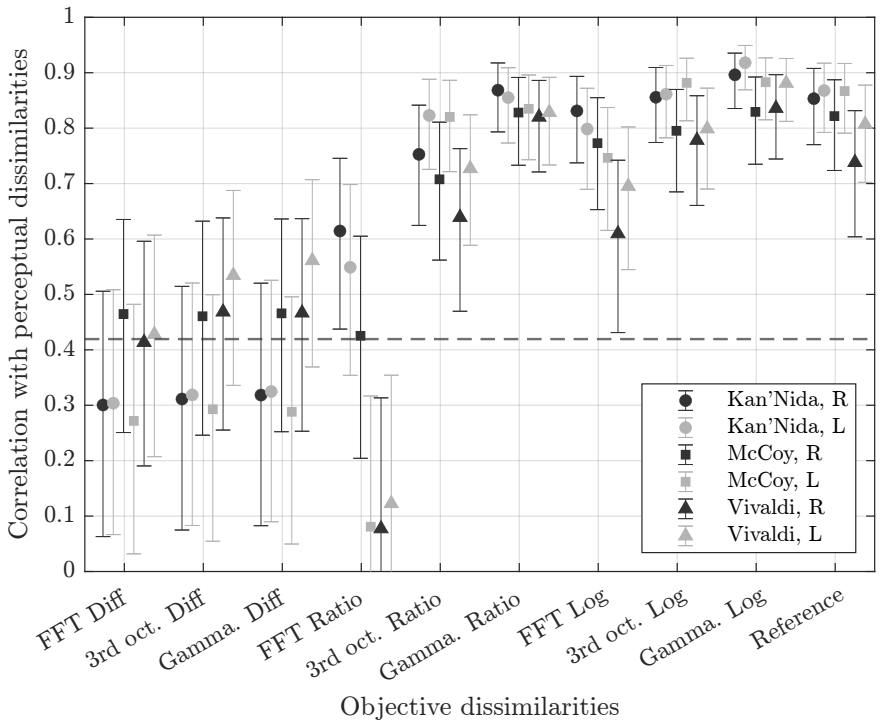

FIG. 4. Correlation between each of the 10 objective dissimilarities (3 spectral analyses × 3 metrics + 1 reference) and the perceptual dissimilarities from the listening tests, for the left (L) and right (R) channels of each musical excerpt (Kan'Nida, McCoy, Vivaldi) used in the loudspeakers study. The error bars indicate the lower and upper bounds for a 95%-confidence interval. Only the correlations above the horizontal dashed line are significant (significant differences among these correlations are presented in Suppl. Table 1[1]).

To assess which correlations are significantly different in Fig. 4 and 5, standard Williams t-tests were performed (Hittner *et al.*, 2003). Only the objective dissimilarities leading to significant correlations with perception for both studies and all excerpts and channels were compared, corresponding to 6 objective dissimilarities: the reference analysis, the metric *Log* with the three spectral analyses, and the metric *Ratio* with the third-octave and gammatone analyses. Because 150 comparisons of correlations are considered, their individual signifi-

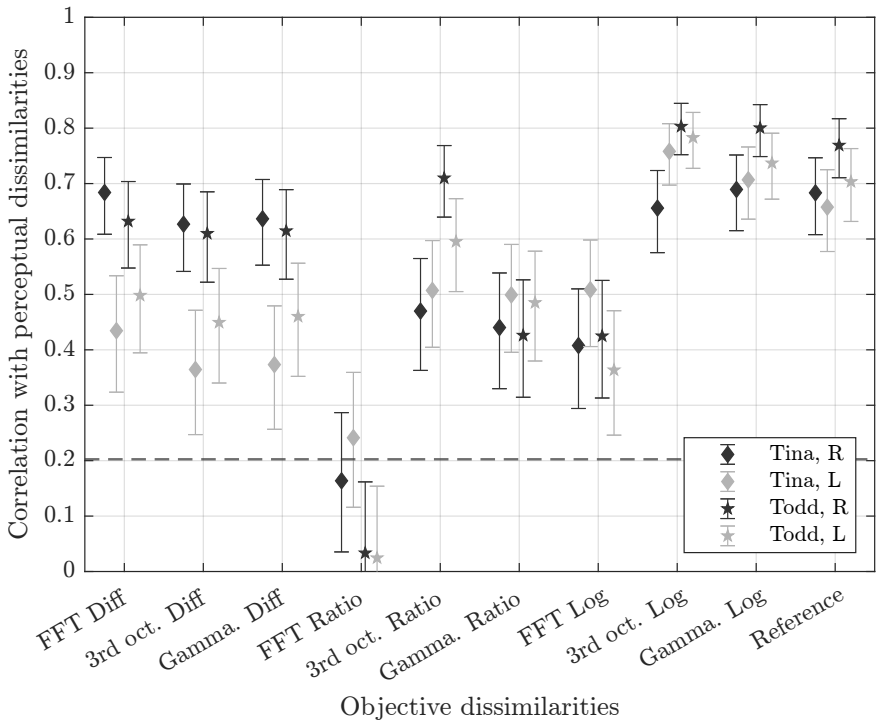

FIG. 5. Same as Fig. 4 but for the headphones study

cance level was Bonferroni corrected to $0.05/150 = 0.0003$. The detailed results of these tests can be found in the supplementary material (Suppl. Table 1[1]). They showed that the third-octave and gammatone analyses with the metric *Log* never lead to significantly different correlations with perception, these correlations being often significantly higher than those obtained with the other combinations of metric and spectral analysis, and almost never significantly different from those obtained with the reference analysis (only once across 10 comparisons). Overall, the third-octave and gammatone analyses with the metric *Log* describe the perceptual dissimilarities with the same good accuracy as the reference analysis. Thus, only the third-octave analysis, that is considered to be a simpler analysis method than the gammatone analysis, is considered in the rest of the study, and further compared with the reference analysis.

2. *MDS analysis of the overall dissimilarities*

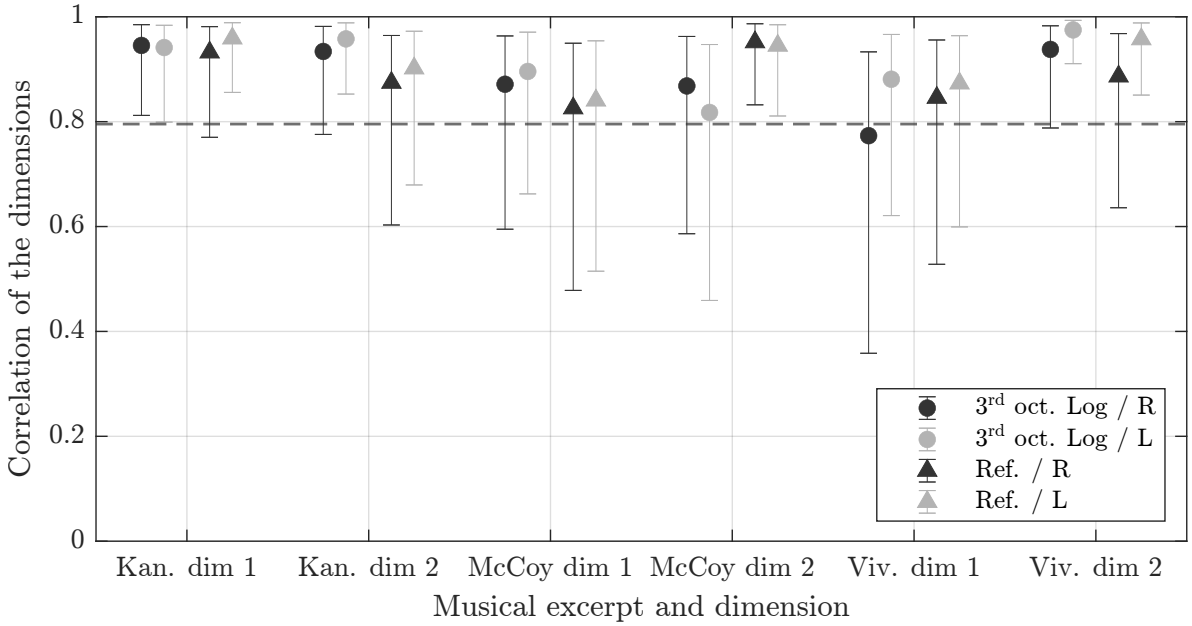

FIG. 6. Correlation between the perceptual dimensions and the objective dimensions resulting from the MDS analysis of the dissimilarities obtained with the third-octave analysis associated with the metric *Log* (3rd oct. Log) and the reference analysis (Ref.), computed on the left and right channels of the recordings, for each musical excerpt used in the loudspeakers study. The error bars indicate the lower and upper bounds for a 95%-confidence interval. Only the correlations above the horizontal dashed line are significant.

The MDS analysis of the objective dissimilarities obtained with the reference analysis and the third-octave analysis with the metric *Log* always led to two-dimensional spaces, like the perceptual spaces (Fig. 1 and 2). Figures 6 and 7 present the correlations between the dimensions of the perceptual and objective spaces. Because 40 correlations are considered, their individual significance level was Bonferroni corrected to $0.05/40 = 0.00125$. The comparison

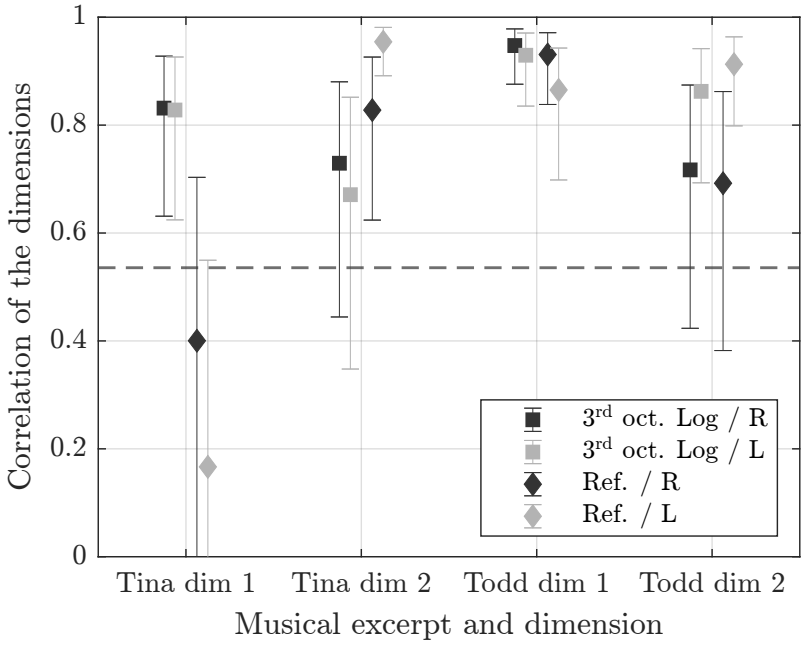

FIG. 7. Same as Fig. 6 but for the headphones study

of these correlations across the two analyses was done with Williams t-tests, only when both correlations are significant. Across studies, 17 comparisons of correlations are considered, and their individual significance level was Bonferroni corrected to $0.05/17 = 0.0029$. For the loudspeakers study (Fig. 6), dimensions 1 and 2 of the objective spaces show a high correlation (above 0.77) with perceptual dimensions 1 and 2, respectively. All correlations are significant except for one (Vivaldi, right channel, dimension 1, third-octave analysis). In all other cases, the third-octave and reference analyses never lead to significantly different correlations with perception. The similarities between the top and bottom panels of Fig. 1 further illustrate these high correlations between the objective and perceptual dimensions (only showed for the third-octave analysis here). In general, similar trends are observed for the headphones study (Fig. 7), except that the reference analysis completely fails to describe the perceptual dimension 1 for Tina Dickow. The two corresponding correlations are

not significant, while all other correlations are significant. The third-octave and reference analyses do not lead to significantly different correlations for the dimensions of Todd Terje, but the dimension 2 for the reference analysis of the left channel of Tina Dickow is significantly more correlated with the perceptual dimension than the corresponding dimension of the third-octave analysis (the two analyses of the right channel do not lead to significantly different correlations for this dimension).

B. Perceptual dimensions

Because the third-octave analysis was able to describe well the perceptual dissimilarities and the perceptual spaces, the third-octave spectra were used to define objective attributes describing each perceptual dimension. Definitions were sought across the loudspeakers and headphones studies, and independently within each study. The attributes giving the best correlations with the perceptual dimensions were sought as detailed in section III 5. Tables II, III and IV present the objective attributes that best describe the perceptual dimensions, leading to the best correlations on average across musical excerpts after taking the maximum correlation obtained with the left and right recording channels, for the loudspeakers study, the headphones study, and using common definitions across studies, respectively. Because 40 correlations are considered (only the maximum values between the left and right channels are presented here), their individual significance level was Bonferroni corrected to $0.05/40 = 0.00125$.

Table II highlights an accurate objective description of the two perceptual dimensions of the loudspeakers study. The first dimension can be described by the balance between

TABLE II. Correlation between the perceptual dimensions and the best objective attributes for the loudspeakers study. All correlations are significant.

Dimension	Attribute	Kan.	McCoy	Viv.
1	$B\left(\frac{[71,224]}{[1425,2806]}\right)$	0.96	0.93	0.92
2	$B\left(\frac{[281,449]}{[7127,14031]}\right)$	−0.97	−0.98	−0.96

TABLE III. Correlation between the perceptual dimensions and the best objective attributes for the headphones study. All correlations are significant.

Dimension	Attribute	Tina	Todd
1	$B\left(\frac{[56,224]}{[445,14031]}\right)$	0.93	0.96
2	$B\left(\frac{[445,1403]}{[1782,14031]}\right)$	0.85	0.93

bass (71 to 224 Hz) and high midrange (1425 to 2806 Hz), the second dimension by the balance between low midrange (281 to 449 Hz) and treble (7127 to 14031 Hz). For the headphones study (Table III), the first dimension can be described as the balance between bass (56 to 224 Hz) and midrange-treble (445 to 14031 Hz), or equivalently as the emergence of midrange and treble (281 to 14031 Hz), the correlations being then of opposite sign. For the second dimension, the best attribute is the balance between midrange (445 to 1403 Hz) and treble (1782 to 14031 Hz).

TABLE IV. Correlation between the perceptual dimensions and the best common objective attributes across the loudspeakers and headphones studies. All correlations are significant.

Dim.	Attribute	Kan.	McCoy	Viv.	Tina	Todd
1	$B\left(\frac{[45,224]}{[1425,14031]}\right)$	0.95	0.93	0.89	0.91	0.93
2	$B\left(\frac{[356,445]}{[3564,11136]}\right)$	−0.80	−0.91	−0.83	0.80	0.85

Common objective attributes across loudspeakers and headphones were investigated (Table IV). The first dimension is successfully described by the balance between bass (45 to 224 Hz) and treble (1425 to 14031 Hz), with a mean correlation across excerpts and studies of 0.92. For the second dimension, the search for a good common attribute was more difficult given that correlations are lower for the dimension 2 of Tina Dickow (headphones study; Table III). A compromise solution was reached by maximizing the mean correlation across the five excerpts and minimizing the corresponding standard deviation. The balance between low midrange (356 to 445 Hz) and treble (3564 to 11136 Hz) leads to a mean correlation of 0.84 with dimension 2. The correlations obtained with these common objective attributes were not significantly different from those obtained with the attributes defined independently for each study (Tables II and III), after setting the individual significance level of the Williams t-tests to $0.05/10 = 0.005$ for the 10 comparisons.

V. DISCUSSION

The present study shows that the third-octave analysis with the metric *Log* can describe the overall dissimilarity between a set of loudspeakers or headphones and the two underlying perceptual dimensions. It achieves this just as well as the reference analysis that uses an auditory model.

A. Objective evaluation of reproduced sound

The approach used here extends previous objective evaluations of reproduced sound. Instead of choosing one signal analysis a priori (Bramsløw, 2004; Klippel, 1990; Michaud *et al.*, 2015; Olive, 2004a; Volk *et al.*, 2016), several analyzes were compared (Lavandier *et al.*, 2008a,b). In addition, different metrics were tested for each signal analysis, so that their relative influence on the dissimilarity evaluation could be highlighted. To make the link with perception, the results show that it is crucial to consider the sound reproduction spectra in dB SPL rather than Pa^2 and with a frequency resolution decreasing with increasing frequency, but that it is not required to model auditory masking.

Figures 4 and 5 show that dissimilarities computed between spectra in Pa^2 with the metrics $Diff$ and *Ratio* are less correlated with perceptual dissimilarities than those computed with the metric *Log* taking the difference between spectra in dB SPL. This was observed for the three signal analyses (FFT, third-octave, and gammatone). Figures 4 and 5 also show that, using the metric *Log*, the dissimilarities computed on the FFT spectra, with linearly-spaced frequency bands of constant width, are less correlated with perception than those

computed on the third-octave and gammatone spectra, with log-spaced frequency bands of width increasing with frequency. The third-octave resolution, a crude approximation of human frequency resolution, is found here appropriate to describe the perceptual dissimilarities. Using a finer approximation with the gammatone analysis does not further improve the link with perception. This could be different when other perceptual dimensions are involved (see section V D).

The third-octave analysis is found as good as the reference analysis to highlight the perceptual dissimilarities. The reference analysis relies on an auditory model that computes the time-varying specific loudness of the stimuli that is then averaged across time. In comparison, the third-octave analysis computes the long-term spectrum across the whole duration of the stimuli without modeling any auditory masking phenomenon, which appears sufficient to reveal the perceptual dissimilarities. Note that Lavandier *et al.* (2008b) indicated that the long-term specific loudness computed on the loudspeaker reproductions was sometimes wrong at describing the perceptual dissimilarities. The third-octave and gammatone analyses, both long-term analyses, are shown to be accurate here. This indicates that one of the problems with the long-term specific loudness is probably that it simulates masking on the long-term spectrum. For non-stationary sounds, it thus simulates masking between frequency components that are not necessarily simultaneous (and masking each other), so that this simulation can be wrong.

B. Overall dissimilarities

Figures 4 and 5 indicate that the third-octave *Log* analysis and the reference analysis produce overall dissimilarities well correlated with the perceptual dissimilarities. The two analyses lead to the same level of correlation for the loudspeakers (mean correlation of 0.83), replicating the reference analysis correlations of Lavandier *et al.* (2008a). The two analyses also lead to the same level of correlation for the headphones, with mean correlations of 0.75 (third-octave) and 0.70 (reference). Volk *et al.* (2016) did not propose an objective evaluation of overall dissimilarity in their study. It was done here, extending the evaluation of overall dissimilarity to headphones. The gammatone *Log* analysis produces overall dissimilarities that lead to the same levels of correlation with perception. Because this analysis involves more filters, and more complicated filters, than the third-octave analysis, and because we were looking for the simplest analysis well correlated with perception, only the third-octave analysis was kept for the multidimensional comparisons with the perceptual spaces and the definition of objective attributes describing the perceptual dimensions.

C. Perceptual dimensions and objective attributes

For the loudspeakers, the two objective dimensions resulting from the MDS analysis of the third-octave dissimilarities are as correlated with the perceptual dimensions as the dimensions resulting from the reference analysis, with correlations varying between 0.77 and 0.97 depending on the musical excerpt, recording channel and dimension (Fig. 6). These correlations are varying between 0.67 and 0.96 for the headphones; apart for Tina Dickow for

412 which the reference analysis fails to highlight the first perceptual dimension, leading to non-
413 significant correlations below 0.4, while the third-octave analysis still provides significant
414 correlations above 0.8 (Fig. 7). Figures 1 and 2 show that the third-octave analysis results
415 in two-dimensional spaces generally very similar to the perceptual spaces. Compared to
416 the study of Lavandier *et al.* (2008b) which relied only on visual comparisons, this space
417 similarity was here quantified by correlations between dimensions.

418 To understand why the reference analysis fails to highlight the first dimension of Tina
419 Dickow, complementary analyses were carried out, varying the frequency range used to
420 compute the overall dissimilarity. The correlation with the first perceptual dimension drops
421 quickly when the frequencies above 3 kHz are included in the dissimilarity calculation. It
422 appears that the reference analysis is giving too much weight to these high frequencies,
423 so that the resulting overall dissimilarities do not describe well the dimension associated
424 with the balance between bass and midrange-treble for Tina Dickow, which corresponds to
425 the excerpt with the lowest bass/treble balance (Suppl. Fig. 1[1]). The weight given to a
426 frequency range in the dissimilarity computation corresponds to the number of frequency
427 bands contained in this range relative to the total number of bands used in the analysis.
428 While the frequencies above 3 kHz carries 32% of the weight in the dissimilarity computation
429 using the reference analysis, this weight is reduced to 18% for the third-octave analysis.
430 By giving less weight to these high frequencies, the third-octave analysis leads to overall
431 dissimilarities better describing the first dimension of Tina Dickow.

432 The objective attributes defined on the third-octave spectra are almost always as cor-
433 related with the perceptual dimensions as the attributes requiring an auditory model pro-

posed in previous studies, in particular the studies considering the same perceptual data. The correlations are above 0.90 for both dimensions and the three musical excerpts for the loudspeakers (Table II). The attributes defined by Lavandier *et al.* (2008b) using an auditory model were not directly compared to the perceptual dimensions but instead to the objective dimensions (as if the attributes were compared to the dimensions of the bottom panels of Fig. 1, whereas here they were compared directly to the perceptual dimensions of the top panels), so that their reported correlations did not account for the differences between the perceptual and objective spaces. Their correlation were above 0.90 for dimension 1 and for the dimension 2 of McCoy Tyner and Vivaldi. For the dimension 2 of Kan'Nida, their best correlation across left and right ears was 0.87, against 0.97 obtained here when comparing directly to the perceptual dimension. Michaud *et al.* (2015) partly re-analyzed this data for McCoy Tyner, computing the objective attributes proposed by Lavandier *et al.* (2008b) but comparing them directly to the perceptual dimensions as done here. The best correlation across ears was 0.85 for dimension 1 and 0.91 for dimension 2. For the headphones, the correlation between the third-octave attributes and the perceptual dimensions is 0.93 and 0.96 for dimension 1 for the two excerpts (Table III), and 0.85 and 0.93 for dimension 2. The attributes defined by Volk *et al.* (2016) using an auditory model led to similar correlations: around 0.95 for both dimensions, with a better correlation for the dimension 2 of Tina Dickow (0.95 vs. 0.85 here). To summarize, the attributes defined on the third-octave spectra tend to outperform those using an auditory model for dimension 2 of the loudspeakers, but it was the opposite for dimension 2 of one excerpt for the headphones. The two types of attributes were equivalent for dimension 1 of the loudspeakers and headphones.

Defining common attributes for loudspeakers and headphones tended to reduce the correlation with the perceptual dimensions (Table IV), essentially for dimension 2, but this trend was not statistically significant, indicating that it might be possible to define general attributes characteristics of sound reproduction. This remains to be confirmed, because here any difference associated with the type of sound reproducing systems (loudspeakers vs. headphones) could have also resulted from the differences in musical excerpts and recording methods (stereophonic vs. binaural), and from the influence of the room for the loudspeakers. This could be further investigated by comparing loudspeakers and headphones using the same musical excerpts and recording methods.

It is interesting to note that, even if the two perceptual dimensions considered here are associated with the relative strengths of bass and midrange, the definitions of the objective attributes proposed to describe them slightly differ from one study to the other. As just mentioned, this could be due to differences in the stimuli (musical excerpts, loudspeakers vs. headphones, room influence). It could also be due to the differences in the way the spectra of the sound reproductions were considered: third-octave analysis vs. different auditory models (Lavandier *et al.*, 2008b; Volk *et al.*, 2016). The procedure used to define the attributes could also produces differences: visual comparisons of the spectra (Lavandier *et al.*, 2008b; Michaud *et al.*, 2015) vs. automatic optimization procedure of the frequency limits used in the attribute definitions, as done here and by Volk *et al.* (2016). The optimization procedure returns precise frequency limits leading to the best correlation with perception, but the correlation remains often very good when these frequency limits are varied around their best values (Volk *et al.*, 2016).

D. Limitations of the study

The third-octave analysis proposed here can describe the two perceptual dimensions highlighted by Lavandier *et al.* (2008b) and Volk *et al.* (2016), along with the overall dissimilarity between loudspeakers/headphones when these two dimensions are involved; but it might not be able to do so when other perceptual dimensions are at play. Toole (1986a) has advocated that higher resolutions than third-octave produce better visual correlations between loudspeaker measurements and fidelity ratings, while Olive (2004b) showed that a model based on 1/20th-octave measurements was significantly more correlated with loudspeaker preference ratings than a third-octave model. Moreover, the analysis proposed here would not be able to describe spatial dimensions that would require binaural information to be incorporated in the analyses (e.g., Michaud *et al.*, 2015). Note that we did not try to combine the left and right ear spectra to define our objective attributes (e.g. as done by Volk *et al.*, 2016). Given the spectral nature of the perceptual dimensions involved and the already high correlations between attributes and dimensions, it seems relevant to consider attributes that can also be used with monaural measurements.

The proposed attributes and their correlation with the two perceptual dimensions confirmed that these dimensions are spectral, as already highlighted by Lavandier *et al.* (2008b) and Volk *et al.* (2016), and that the corresponding dissimilarity evaluations are related to timbre-related accuracy or sound quality/fidelity (AES20-1996, 1996; IEC Publication 60268-13, 1998). Our third-octave analysis does not make the link between the overall dissimilarities among loudspeakers/headphones or the underlying dimensions and the absolute

sound quality/fidelity of these systems or the listener preferences. It would be important to clarify this link in the future, in particular for the perceptual dimensions.

The third-octave analysis proposed here was tested on only five musical excerpts. Different musical excerpts might highlight different perceptual dimensions or involve different weightings of the perceptual dimensions in global judgments (dissimilarity here, quality or preference considered in other studies; Bech, 1994; Eisler, 1966). The limited number of loudspeakers on which the analysis was tested triggers the same issues. While the proposed analysis could prove useful to describe other spectral dimensions, it remains to be tested (and might require a higher spectral resolution). The analysis would not be able to describe any individual differences between listeners. Our MDS analysis only allows to investigate the main dimensions shared by the listeners (Volk *et al.*, 2016). Here, dissimilarities appear similar for musical excerpts having very different spectra, like in the listening tests of Gabrielsson *et al.* (1974), suggesting that it might even be possible to highlight these dissimilarities by considering directly the frequency response of the reproducing systems rather than recordings of their reproduction of musical excerpts. This last point remains to be investigated. This frequency response would probably need to account for the influence of the listening room in the case of loudspeakers (Klippel, 1990; Olive, 2004b; Toole, 1986b). This influence is currently accounted for in our analysis that is based on listening room recordings. The positions of the loudspeakers along each perceptual dimension result from the combined influences of the room and loudspeakers on the spectra at the ears (Bech, 1994; Olive *et al.*, 1994). The two perceptual dimensions involved here seem however characteristic of the reproducing systems rather than of the particular room used, because the

same dimensions were obtained when using another listening room (Michaud *et al.*, 2015) or when no room was involved in the headphones study (Volk *et al.*, 2016). The fact that the two perceptual dimensions are similar for different excerpts also suggests that it could be interesting to investigate whether conducting a listening test (and the objective analysis) using a test signal like pink noise (having attractive properties as a measurement signal) would generate the same results as music.

All the spectral analyses were performed over the entire available spectrum. However, some frequency bands may not be useful for differentiating the sound reproductions. These bands would then represent noise in the evaluation of dissimilarities (as discussed above for the reference analysis and the first dimension of Tina Dickow). A complementary investigation was carried out by varying the upper and lower frequency limits of each tested analysis (FFT, third-octave, and gammatone). For the third-octave analysis, the correlations with the overall perceptual dissimilarities are best for a frequency range between 56 Hz and 14 kHz, for both studies. For the gammatone analysis, the correlations are best for a frequency range between 16 Hz and 15.5 kHz. However, the improvements due to optimizing the frequency range remain minor. For the loudspeakers study, the mean correlation improves by 0.05 (third-octave) and 0.00 (gammatone). For the headphones study, the correlation increases by 0.03 and 0.04, respectively. For the FFT analysis, varying the frequency range did not improve sufficiently the low or non-significant correlations with perception.

When computing the overall dissimilarity between two spectra, one could weight the dissimilarities in each frequency band, e.g. by the signal level in this band. Again, given the already high correlations between objective and perceptual dissimilarities, the advantage

of this more elaborate computation would be very limited here. This option to revise the metric *Log* should be kept mind should it prove insufficient to describe perception, e.g. when dissimilarities are more dependent on the musical excerpt used.

VI. CONCLUSION

Considering the perceptual dissimilarity evaluations from a previous study comparing 12 loudspeakers with three musical excerpts and a study comparing 21 headphones with two musical excerpts, it was shown here that a third-octave analysis can describe both the overall dissimilarity between the set of loudspeakers or headphones, and the two perceptual dimensions underlying the dissimilarity judgments. The third-octave analysis achieves this just as well as a reference analysis that uses an auditory model. The two perceptual dimensions associated with the relative strengths of bass and midrange could be described by objective attributes defined on the third-octave spectra of the recorded sound reproductions. The present study highlights that it is crucial to consider the sound reproduction spectra in dB SPL rather than Pa^2, and with a frequency resolution decreasing with increasing frequency such as the third-octave resolution, but that it is not required to model auditory masking.

SUPPLEMENTARY MATERIAL

See supplementary material at [URL will be inserted by AIP] for the spectra of the musical excerpts (Suppl. Fig. 1), the perceptual spaces with 95%-confidence ellipses (Suppl. Fig. 2 and 3), and the results of the Williams t-tests used to compare correlations (Suppl. Table 1).

ACKNOWLEDGMENTS, AUTHOR DECLARATIONS AND DATA AVAILABILITY STATEMENT

This work was performed within the LabEx CeLyA (Grant No. ANR-10-LABX-0060). The authors have no conflicts to disclose, and thank E. Meziani and B. Larello for early work on this study, and C. P. Volk for advices on a previous version of this manuscript. The data involving human participants come from previous studies. The data that support the findings of the present study are available from the corresponding author upon reasonable request.

[1]See supplementary material at [URL will be inserted by AIP]

[2]The classical model MDSCAL with the algorithm SMACOF was used (Borg and Groenen, 1997).

[3]The three spectral analyses are defined over different frequency ranges. The correlations calculated over a common frequency range (the smallest of the three) differ on average by only 0.01 from those computed with the default ranges.

AES20-1996 (**1996**). "AES recommended practice for professional audio - Subjective evaluation of loudspeakers," J. Audio Eng. Soc. **44**(5), 382–400.

ANSI/CTA Standard 2034-A (**2015**). "Standard method of measurement for in-home loudspeakers," Consumer Technology Association, USA.

Bech, S. (**1994**). "Perception of timbre of reproduced sound in small rooms: Influence of room and loudspeaker position," J. Audio Eng. Soc. **42**(12), 999–1007.

583 Bech, S. (**1995**). "Timbral aspects of reproduced sound in small rooms. I," J. Acoust. Soc.
584 Am. **97**(3), 1717–1726.

585 Bech, S. (**1996**). "Timbral aspects of reproduced sound in small rooms. II," J. Acoust. Soc.
586 Am. **99**(6), 3539–3549.

587 Bech, S. (**2002**). "Requirements for low-frequency sound reproduction, Part 1: The audibil-
588 ity of changes in passband amplitude ripple and lower system cutoff frequency and slope,"
589 J. Audio Eng. Soc. **50**(7/8), 564–580.

590 Bech, S., and Zacharov, N. (**2006**). *Perceptual Audio Evaluation — Theory, Method and*
591 *Application* (John Wiley & Sons, Chichester).

592 Borg, I., and Groenen, P. (**1997**). *Modern multidimensional scaling. Theory and applications*
593 (Springer).

594 Bramsløw, L. (**2004**). "An objective estimate of the perceived quality of reproduced sound
595 in normal and impaired hearing," Acta Acustica united with Acustica **90**(6), 1007–1018.

596 Eisler, H. (**1966**). "Measurement of perceived acoustic quality of sound-reproducing systems
597 by means of factor analysis," J. Acoust. Soc. Am. **39**(3), 484–492.

598 Gabrielsson, A., and Lindström, B. (**1985**). "Perceived sound quality of high-fidelity loud-
599 speakers," J. Audio Eng. Soc. **33**(1/2), 33–53.

600 Gabrielsson, A., Lindström, B., and Till, O. (**1991**). "Loudspeaker frequency response and
601 perceived sound quality," J. Acoust. Soc. Am. **90**(2, Pt. 1), 707–719.

602 Gabrielsson, A., Rosenberg, U., and Sjögren, H. (**1974**). "Judgments and dimension analyses
603 of perceived sound quality of sound-reproducing systems," J. Acoust. Soc. Am. **55**(4), 854–
604 861.

605 Gabrielsson, A., and Sjögren, H. (**1979**). "Perceived sound quality of sound-reproducing
606 systems," J. Acoust. Soc. Am. **65**(4), 1019–1033.
607 Hittner, J. B., May, K., and Silver, N. C. (**2003**). "A Monte Carlo evaluation of tests for
608 comparing dependent correlations," The Journal of General Psychology **130**(2), 149–168.
609 IEC Publication 60268-13 (**1998**). "Sound system equipment - Part 13: Listening tests on
610 loudspeakers," International Electrotechnical Comission, Geneva, Swizerland.
611 IEC Publication 60268-5 (**1989**). "Sound system equipment - Part 5: Loudspeakers," Inter-
612 national Electrotechnical Comission, Geneva, Swizerland.
613 IEC Publication 60581-7 (**1986**). "High fidelity audio equipment and systems. Minimum per-
614 formance requirements - Part 7: Loudspeakers," International Electrotechnical Comission,
615 Geneva, Swizerland.
616 Illényi, A., and Korpássy, P. (**1981**). "Correlation between loudness and quality of stereo-
617 phonic loudspeakers," Acustica **49**(4), 334–336.
618 Klippel, W. (**1990**). "Multidimensional relationship between subjective listening impression
619 and objective loudspeaker parameters," Acustica **70**, 45–54.
620 Lavandier, M., Guyot, B., Meunier, S., and Herzog, P. (**2005**). "The influence of stereophony
621 on the restitution of timbre by loudspeakers," in *AES 119th Convention*, 6619.
622 Lavandier, M., Herzog, P., and Meunier, S. (**2004**). "Perceptual and physical evaluation of
623 loudspeakers," in *AES 117th Convention*, 6240.
624 Lavandier, M., Herzog, P., and Meunier, S. (**2008**a). "Comparative measurements of loud-
625 speakers in a listening situation," J. Acoust. Soc. Am. **123**(1), 77–87.

626 Lavandier, M., Meunier, S., and Herzog, P. (**2008**b). "Identification of some perceptual
627 dimensions underlying loudspeaker dissimilarities," J. Acoust. Soc. Am. **123**(6), 4186–
628 4198.

629 Michaud, P.-Y., Lavandier, M., Meunier, S., and Herzog, P. (**2015**). "Objective characteriza-
630 tion of perceptual dimensions underlying the sound reproduction of 37 single loudspeakers
631 in a room," Acta Acustica united with Acustica **101**(3), 603–615.

632 Moore, B. C., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-
633 filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**(3), 750–753.

634 Moore, B. C. J., and Tan, C. T. (**2003**). "Perceived naturalness of spectrally distorted
635 speech and music," J. Acoust. Soc. Am. **114**(1), 408–419.

636 Olive, S. E. (**2003**). "Differences in performance and preference of trained versus untrained
637 listeners in loudspeaker tests: a case study," J. Audio Eng. Soc. **51**(9), 806–825.

638 Olive, S. E. (**2004**a). "A multiple regression model for predicting loudspeaker preference
639 using objective measurements: Part 1 - Listening test results," in *AES 116th Convention*,
640 6113.

641 Olive, S. E. (**2004**b). "A multiple regression model for predicting loudspeaker preference
642 using objective measurements: Part 2 - Development of the model," in *AES 117th Con-*
643 *vention*, 6190.

644 Olive, S. E., Schuck, P. L., Sally, S. L., and Bonneville, M. E. (**1994**). "The effect of
645 loudspeaker placement on listener preference ratings," J. Audio Eng. Soc. **42**(9), 651–669.

646 Patterson, R. D., Nimmo-Smith, I., Holdsworth, J., and Rice, P. (**1987**). "An efficient
647 auditory filterbank based on the gammatone function," in *a meeting of the IOC Speech*

Group on Auditory Modelling at RSRE, Vol. 2.

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**(7/8), 581–593.

Ravizza, G., Villegas, J., Volk, C. P., and Stegenborg-Andersen, T. (**2023**). "An over-ear headphone target curve for Brüel & Kjær head and torso simulator type 5128 measurements," in *AES 155th Convention*, 127.

Schoeffler, M., and Herre, J. (**2016**). "The relationship between basic audio quality and overall listening experience," J. Acoust. Soc. Am. **140**(3), 2101–2112.

Staffeldt, H. (**1974**). "Correlation between subjective and objective data for quality loudspeakers," J. Audio Eng. Soc. **22**(6), 402–415.

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

Toole, F. E. (**1986**a). "Loudspeaker measurements and their relationship to listener preferences: Part 1," J. Audio Eng. Soc. **34**(4), 227–235.

Toole, F. E. (**1986**b). "Loudspeaker measurements and their relationship to listener preferences: Part 2," J. Audio Eng. Soc. **34**(5), 323–348.

Toole, F. E. (**1991**). "Binaural record/reproduction systems and their use in psychoacoustic investigations," in *AES 91st Convention*, 3179 (L-6).

Volk, C. P., Lavandier, M., Bech, S., and Christensen, F. (**2016**). "Identifying the dominating perceptual differences in headphone reproduction," J. Acoust. Soc. Am. **140**(5), 3664–3674.

Zwicker, E., and Fastl, H. (**1983**). "A portable loudness-meter based on ISO 532 B," in *Proc. 11th International Congress on Acoustics*, pp. 135–137.

Zwicker, E., and Fastl, H. (**2013**). *Psychoacoustics: Facts and models*, **22** (Springer Science & Business Media).