Available online at www.sciencedirect.com
ScienceDirect
C. R. Mecanique 334 (2006) 732–736
Comptes Rendus Mecanique
http://france.elsevier.com/direct/CRAS2B/

Physical and perceptual estimation of differences between loudspeakers

Mathieu Lavandier [1], Philippe Herzog *, Sabine Meunier

Laboratoire de Mécanique et d'Acoustique, 31, chemin Joseph-Aiguier, 13402 Marseille cedex 20, France

Received 25 November 2005; accepted after revision 4 October 2006

Available online 7 November 2006

Presented by Évariste Sanchez-Palencia

Abstract

Differentiating the restitution of timbre by several loudspeakers may result from standard measurements, or from listening tests. This work proposes a protocol keeping a close relationship between the objective and perceptual evaluations: the stimuli are musical excerpts, and the measuring environment is a standard listening room. The protocol involves recordings made at a listener position, and objective dissimilarities are computed using an auditory model simulating masking effects. The resulting data correlate very well with listening tests using the same recordings, and show similar dependencies on the major parameters identified from the dissimilarity matrices. ***To cite this article: M. Lavandier et al., C. R. Mecanique 334 (2006).***

Résumé

Comparaison objective et perceptive de la reproduction du timbre par des enceintes. Pour comparer la reproduction du timbre par différentes enceintes acoustiques, il est possible d'utiliser des mesures normalisées ou des tests d'écoute. Ce travail propose un protocole qui relie les approches objectives et perceptives : des extraits musicaux sont reproduits par les enceintes dans une pièce usuelle, et enregistrés à la position d'écoute. Un modèle auditif simulant le masquage permet de calculer des dissemblances entre les signaux enregistrés. Ceci conduit à une très bonne corrélation avec les résultats de tests d'écoute utilisant les mêmes enregistrements, et les principales influences perçues se retrouvent lors de l'analyse des matrices de dissemblances ainsi obtenues. ***Pour citer cet article : M. Lavandier et al., C. R. Mecanique 334 (2006).***

Keywords: Acoustics; Loudspeaker; Measurement

Mots-clés : Acoustique ; Enceinte ; Mesure

* Corresponding author.
E-mail addresses: lavandiermn@cardiff.ac.uk (M. Lavandier), herzog@lma.cnrs-mrs.fr (P. Herzog), meunier@lma.cnrs-mrs.fr (S. Meunier).
[1] Present address: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff, CF10 3AT, UK.

1631-0721/$ – see front matter
doi:10.1016/j.crme.2006.10.004

Version française abrégée

L'objectif principal de ce travail est de proposer une méthode d'évaluation objective de la différence de reproduction du timbre par des enceintes acoustiques, telle qu'elle donne des résultats aussi proches que possible de ceux de tests d'écoute. Il s'agit uniquement d'évaluer des différences, et non d'estimer la qualité intrinsèque des enceintes. Pour que les évaluations objectives et perceptives soient aussi proches que possible, elles utilisent exactement les mêmes signaux de mesure.

Nous avons ainsi choisi d'enregistrer des signaux musicaux, reproduits par les différentes enceintes placées à la même position dans une salle d'écoute. La prise de son fait appel à une paire de microphones en configuration stéréophonique « AB ORTF », située à la place de l'auditeur. Les différences entre ces enregistrements sont alors évaluées au casque, par paires. Les mêmes enregistrements sont soumis à des méthodes d'analyse objectives, pour lesquelles ont été défini des « métriques » qui permettent de remplir des matrices de dissemblances similaires à celles issues des tests d'écoute. Ce protocole favorise ainsi le lien entre les deux évaluations, au détriment des caractéristiques spatiales de la reproduction du fait de l'écoute au casque, ce qui nous restreint à l'étude de la reproduction du timbre.

Des mesures ont été conduites à partir d'un échantillon de douze enceintes, qui reproduisaient trois extraits musicaux variés, et dont les dissemblances ont été évaluées par vingt-sept auditeurs non entraînés. Les résultats des tests d'écoute de l'ensemble des auditeurs ont permis d'établir une matrice de dissemblance moyenne, à laquelle ont été corrélées les neuf matrices de dissemblances objectives issues des métriques que nous avons proposées. Cette matrice perceptive moyenne a également fait l'objet d'une analyse statistique multidimensionnelle, qui a révélé que les différences perçues peuvent être représentées selon deux dimensions. La même démarche a été appliquée à chaque matrice de dissemblances objectives, et les espaces résultants ont alors été comparés visuellement à l'espace perceptif.

Parmi les différentes techniques d'analyse mises en jeu pour définir nos métriques, celles qui conservent toutes les informations du signal ont conduit à une corrélation relativement faible avec les résultats perceptifs. L'utilisation de densités spectrales de puissance, qui éliminent l'information de phase, a sensiblement amélioré cette corrélation. Elle est cependant encore nettement meilleure en utilisant des modèles auditifs qui prennent en compte le masquage fréquentiel et temporel (voir Fig. 1).

Cette hiérarchie est conservée en considérant, à la place d'une corrélation, la comparaison entre les espaces multidimensionnels issus de l'analyse des matrices de dissemblances : seules les métriques utilisant des modèles perceptifs semblent donner des espaces objectifs proches des espaces perceptifs (Fig. 2).

1. Introduction

Physical measurements used nowadays to differentiate loudspeakers do not take into account the mechanisms of perception. Conversely, the perceived characteristics of loudspeakers are evaluated rigorously through standard listening tests [1]. The aim of the present work is to propose a method for measuring loudspeakers that combines physical and perceptual points of view, so that the two approaches can be linked.

To achieve our goal, physical as well as psychoacoustical measurements were done in parallel, on a large panel of loudspeakers. We focused on the differences between loudspeakers, instead of a direct evaluation of their quality. Our main objective was to find a physical method of analysis that allows the loudspeakers to be differentiated from objective measurements in the same way as they are differentiated perceptually. To avoid many cultural biases, we did not try to evaluate the perceived quality of the loudspeakers. Such a quality evaluation might be a further work, if a reliable and non biased link between perception and physics could be established, eventually leading to methods for predicting the specific qualities of individual loudspeakers.

In this preliminary approach of estimating only the dissimilarities between loudspeakers, we have kept the objective and perceptual measurement protocols as close as possible, so that they could share almost the same data. This builds several constraints, which prevent the evaluation of spatial components of the reproduction: as most of these have probably been filtered out by our proposed protocols, we focus on the reproduction of timbre.

2. Experimental protocol

The protocol consists in recording separately the acoustical field radiated by each loudspeaker, ensuring that all loudspeakers are at the same position in the same room, with the microphones located at a fixed listening position, as specified by tests standards [1]. For estimating perceptive dissimilarities, these recordings are played to listeners via headphones in a soundproof room. In parallel, objective dissimilarities are estimated by analysing the same recordings using different signal processing techniques. This protocol keeps both estimations close together, but it induces distortions into the sound transmitted to the listeners' ears. As we are looking for differences between loudspeakers, we assume that these distortions do not influence our dissimilarity results because they are almost identical for all loudspeakers. However, this protocol prevents a rigorous study of the spatial components of the sound reproduction, because of headphone listening. Spatial aspects may indeed influence such a listening, but they cannot be studied in this way.

The stimuli were three musical excerpts, selected to cover a wide range of spectra and dynamics: Kan'nida ('Konsyans', percussions, 1.7 s), Mc Coy Tyner ('Miss Bea', jazz, 3.3 s) and Vivaldi ('L'Europa Galante', symphonic orchestra, 4.7 s). Twelve loudspeakers were involved, each one facing the microphones, in a monophonic reproduction situation. The influence of stereophonic reproduction has been studied, and does not significantly change the dissimilarity results [2]. The field radiated by the loudspeakers has been recorded using a stereophonic AB ORTF technique. Two other recording techniques have also been tested: stereophonic MS and monophonic. As they do not lead to different conclusions, they are not presented here, but the corresponding results can be found in [3].

3. Perceptual and objective dissimilarities

Perceptual dissimilarities have been estimated from listening tests involving twenty-seven listeners, without training. Three tests were run, one for each musical excerpt. The loudness of the recordings has been equalized to avoid that listeners make their judgments on level differences, and pay less attention to other aspects. This loudness was set to a level of roughly 70 phons, as judged by the experimenters during informal listening sessions, and none of the listeners reported any level difference during the experiment. Paired comparisons were then used to evaluate dissimilarities between the different loudspeakers, using stereophonic headphones for reproducing the sound field recorded in the room for each single loudspeaker.

Pairs of recordings were thus presented in random order to the listeners. They were requested to quantify the dissimilarity within each pair by moving a cursor on a line displayed on a screen. The two end points of the line were labelled 'very similar' and 'very dissimilar'. The final dissimilarity matrix, from which the statistical analysis was done, was obtained by averaging the individual perceptual dissimilarities.

Our aim requires also to build dissimilarity matrices from the objective analyses, so that the objective and perceptual approach can be compared. To fill such matrices, we defined 'metrics' giving a single value from the differences between objective analyses of the recordings of the different loudspeakers. These objective analyses use exactly the same recordings as the listening tests. The right and left channels of the recordings have been analysed separately, but the figures below show the mean of both channels.

Nine metrics have been proposed (see [3] for details). They are differentiated by the letters (a) to (i) thereafter. Three of them involve the signals 'as is': temporal (a), spectral (b) and time/frequency (c) representations. Three metrics are calculated from weighted power spectral densities of the signals: linear weighting (d), A-weighting (e), and equal-loudness contour at 70 phons [4] (f). Weightings for (e) and (f) result from average properties of hearing. The last three metrics were calculated from the output of auditory models. These models have been proposed by Zwicker and Fastl [5] to calculate loudness, and try to model auditory masking. These models analyse the signals in terms of specific loudness, which is the loudness of the signal along the bark scale, i.e. a 'perceptual' frequency scale.

The model involved in the seventh metric takes into account only frequency masking, and evaluates loudness of the signal as a whole, from its overall mean spectrum (g). A more elaborated model takes into account both frequency and temporal masking, calculating the 'time-varying specific loudness' at short time intervals (every 10 ms). Dissimilarities between two specific loudnesses can thus be calculated every 10 ms, and a metric is defined as the temporal integration of this single time-varying dissimilarity (h). The last metric is based on the dissimilarity calculated between the two temporal means of the two time-varying specific loudnesses (i).

4. Comparison between perceptual and objective results

A comparison can then be made between the dissimilarity matrices resulting from both approaches, which rely on common signal recordings. Subsequent discussions deal with the mean of the results obtained using the three musical

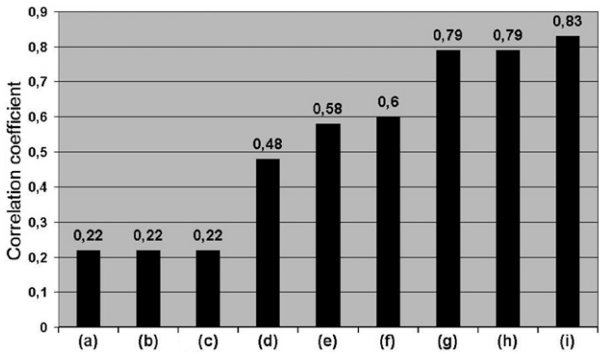

Fig. 1. Correlation between perceptual and objective dissimilarities averaged within the three musical excerpts: (a) temporal domain, (b) spectral domain, (c) time/frequency domain, (d) power spectral density, (e) A-weighting, (f) equal-loudness at 70 phons weighting, (g) overall specific loudness, (h) time-varying specific loudness, (i) temporal mean of time-varying specific loudness.

Fig. 1. Corrélations entre les différences perceptives et objectives, prises en moyenne sur les trois extraits musicaux : (a) domaine temporel, (b) spectre complexe, (c) plan temps/fréquence, (d) densité spectrale de puissance, (e) spectre pondéré A, (f) spectre pondéré par la courbe iso-sonique à 70 phones, (g) sonie spécifique globale, (h) moyenne de sonie spécifique instantanée, (i) moyenne temporelle de la sonie spécifique instantanée.

excerpts, but it should be noted that the same tendencies are observed on each excerpt. Another approach could thus have been to use physical measurements totally independent of the musical signal, as the frequency response of loudspeakers, but our main concern was to keep the perceptual and objective measurements very close to each other.

4.1. Direct comparison of dissimilarities

Correlations between perceptual and objective dissimilarities were calculated for the two channels of each musical excerpt. Fig. 1 shows the average of these correlations.

Dissimilarities based on temporal (a), frequency (b) and time/frequency (c) representations give equivalent results when they are compared to perceptual dissimilarities. This is not surprising, because all three contain the same information on the signal.

When the dissimilarities are evaluated on metrics based on the power spectral density ((d) to (f)), the correlations are higher, suggesting that the information carried by the phase, which is not taken into account in the power spectral density, is not pertinent in relation to our listening conditions. It even seems to mask somewhat the suitable information, as in dissimilarities based on a full representation ((a) to (c), taking into account the phase). The weightings applied to the power spectral density yield a small increase of the correlation for the comparison showed here (see (e) and (f)), but this result was not confirmed by other studies [2,6] and seems to be dependent on the signal [3].

On Fig. 1, the last three bars correspond to the correlation between perceptual dissimilarities and objective dissimilarities calculated using auditory models. One can observe that taking into account auditory model to evaluate objective dissimilarities greatly improves the correlation with perceptual dissimilarities.

4.2. Multidimensional comparison of dissimilarities

Multidimensional scaling analysis (MDS) has been applied to each dissimilarity matrix to build the corresponding spaces [7]. The MDS is used to represent each loudspeaker in a multidimensional space, so that all the distances between the loudspeakers in this space are as close as possible to the values from the initial dissimilarity matrix. Each dimension of this multidimensional space reveals a significant component of the dissimilarities, although MDS does not provide any interpretation of the underlying nature of these components.

Comparison between MDS analysis of the two matrices resulting from objective and perceptual measurements provides more information than correlation, as it shows if both approaches are using similar components for estimating dissimilarities. It abandons the criterion of linearity to take into account the multidimensional nature of our perception of sound reproduction. The main perceptual dimensions used by listeners to discriminate the recordings are revealed, and the comparison of objective and perceptual spaces tells us if an objective analysis seems appropriate to extract these dimensions from the recordings. On the other hand, such a comparison is done yet only by visual examination, resulting in less reliable conclusions.

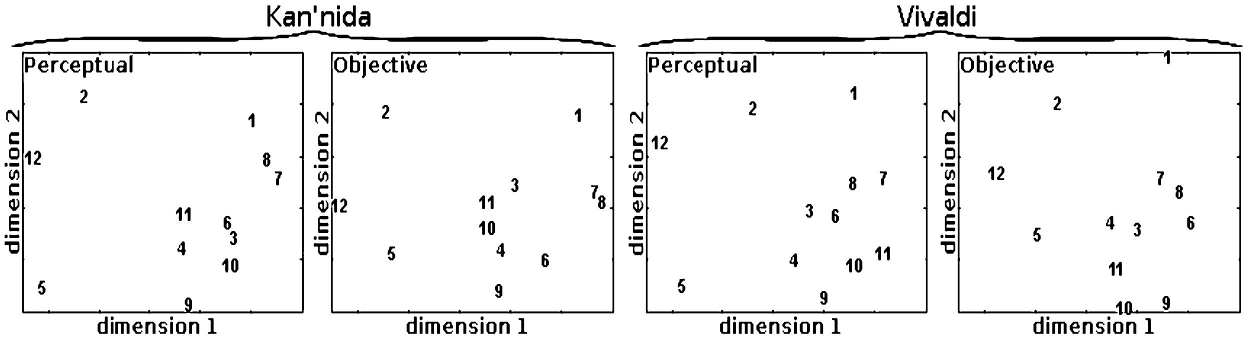

Fig. 2. Perceptual (left) and objective(right) spaces for musical excerpts Kan'nida and Vivaldi. Objective spaces are based on temporal mean of the time-varying specific loudness of the signals. The same scale is used for all space.

Fig. 2. Espaces perceptifs (gauche) et objectifs (droite) pour les extraits musicaux Kan'nida et Vivaldi. Les espaces objectifs sont basés sur la moyenne temporelle de la densité spécifique de sonie instantanée des signaux. Les échelles sont communes à tous les espaces.

Fig. 2 (left graphs) shows the 'perceptual spaces' obtained for two excerpts: Kan'nida and Vivaldi. It can be observed that the listeners have used at least two dimensions to differentiate the recordings. The dimension 1, which is predominant, seems to be linked to the bass/treble balance of recordings. The dimension 2 would be linked to the amount of medium frequencies, and could be associated with the notion of sound clarity [3].

Nine objective spaces (one per metric) have been obtained for each channel of each musical excerpt. Only the metrics based on auditory models that take into account both frequency and temporal masking led to objective spaces similar to the perceptual ones. As an example, Fig. 2 shows objective spaces obtained using the temporal mean of the time-varying specific loudness, for the right channel of Kan'nida and the left channel of Vivaldi. When overall specific loudness is used, the objective and perceptual spaces may differ for some signals. This last result implies that it is important to take into account the temporal variations of the signal during analysis, just as the listeners most probably do.

5. Conclusion

This research concerns objective and perceptual evaluation of the restitution of timbre by loudspeakers. It shows that auditory models can be used as a tool for signal analysis, in order to define objective metrics that behave similarly to an 'average' listener. The auditory models used here are based on auditory masking and were developed to measure loudness [5]. This study suggests a reason why standard measurements of frequency responses in anechoic environment, usually used to evaluate loudspeakers, cannot be straightforwardly linked to perceptual evaluation: these measurements cannot take into account the time-varying auditory mechanisms responsible for masking effects.

The study does not conclude on the specific contribution of the various aspects of the metrics proposed from auditory models (different ways of summing, time scales, ...). More experimental data are needed to clarify these points. Other definitions for metrics could be proposed, and these could lead to an ever tighter link between objective and perceptual measurements. This has also to be developed further.

References

[1] IEC Publication 60268-13, Sound system equipment – Part 13: Listening tests on loudspeakers, International Electrotechnical Commission, Geneva, Switzerland, 1998.
[2] M. Lavandier, B. Guyot, S. Meunier, P. Herzog, The influence of stereophony on the restitution of timbre by loudspeakers, AES 119th Convention, paper 6619, New York, 2005.
[3] M. Lavandier, P. Herzog, S. Meunier, The restitution of timbre by loudspeakers in a listening room: perceptual and physical measurements, AES 117th Convention, paper 6240, San Francisco, 2004.
[4] British Standard ISO 226: 2003, Acoustics – normal equal-loudness level contours, BSi (2003).
[5] E. Zwicker, H. Fastl, Psychoacoustics: Facts and Models, Springer-Verlag, Berlin/New York, 1999.
[6] M. Lavandier, S. Meunier, P. Herzog, Perceptual and physical evaluation of differences among a large panel of loudspeakers, Forum Acusticum 2005, Budapest, 2005, in press.
[7] I. Borg, P. Groenen, Modern Multidimensional Scaling. Theory and Applications, Springer-Verlag, Berlin/New York, 1997.