Forum Acusticum 2011 - Aalborg - Denmark

Perceptual Evaluation of Loudspeaker Nonlinearities

Pierre-Yohan Michaud, Philippe Herzog, Sabine Meunier
LMA - CNRS UPR 7051, 31 Chemin Joseph Aiguier, Marseille, France.

Antonín Novák
LAUM - CNRS UMR 6613, Avenue Olivier Messiaen, Le Mans, France.

Summary
This study deals with the evaluation of the perception of distortion in the restitution of musical excerpts played by loudspeakers. The main focus is to evaluate the perceived distortion induced by low frequency nonlinearities of the loudspeaker. A large panel is created in which only the distortion varies among the loudspeakers. The "virtual" loudspeakers are created from one single actual loudspeaker by associating its anechoic filtered tweeter recording and its woofer signal model in which the distortion can be modified. To model the woofer restitution, its identification is done using a swept sine technique which separates the linear and higher order impulse responses. These impulse responses are used as linear and nonlinear frequency filters characterizing the sound pressure of the system under test. The woofer synthesis consists in expanding the musical excerpts on the basis of Chebyshev polynomials combined with the identified frequency filters (Novák et al., DAFx 2010). Different "virtual" loudspeakers are obtained by amplifying or reducing the influence of specific nonlinearities. In order to evaluate the perceived distortion, in realistic listening conditions, the auralization technique was used to convolve the "virtual" loudspeakers with the impulse response of a suitable room. Upcoming listening tests are conducted with headphones to evaluate the perceived differences over the created panel of loudspeakers with a method allowing the evaluation of large sets of stimuli. The dissimilarity judgments can then be analyzed with a Multidimensional scaling (MDS) technique to reveal the underlying dimensions of the distortion space.

PACS no. 43.66.Lj, 43.38.Ja

1. Introduction

Loudspeakers can be approximated as linear systems within a limited dynamic range. However, for wider dynamic ranges, the loudspeaker is clearly a nonlinear system generating distortion components. The nonlinear distortion is characterized by the generation of frequency components that are not present in the input signal. For sinusoidal inputs, the other output frequencies may be referred as harmonics. The distortion of a loudspeaker is typically measured using sinusoidal input signals. The total harmonic distortion (THD) is measured when the input signal is one sinusoid, and the intermodulation distortion (IMD) is measured when the loudspeaker is excited by two sinusoids or a multitone signal [1]. Both criteria are usually expressed as the ratio of the distortion products to the total system output. However the perception of distortion is not clearly related to the percentage of THD or IMD. Moreover, these objective methods are not appropriated for evaluating the perception and audibility of distortion for more realistic input signals such as music or speech. Evaluating the perception of distortion depends on the characteristics of the musical or speech signal and the complexity of the reproduction system. The properties of the human auditory system and specially the masking effects have also a significant importance when evaluating distortion in loudspeakers restitution.

Most of the studies dealing with perception of distortion with musical signals refer to audibility thresholds of different types of distortions [2, 3, 4, 5]. The aim of our study is to evaluate the perceived difference between loudspeakers presenting different distortion characteristics. This idea is based on the work done by Lavandier *et al.* [6, 7] who studied the restitution of timbre by loudspeakers and elaborated a protocol. It consists in evaluating the perceived difference or dissimilarity rather than the perceived quality between loudspeakers further analyzed with a multidimensional scaling (MDS) technique. This technique leads to a perceptual space and allows to reveal the underlying dimensions that listeners used to make their dissimilarity judgments. The same procedure is used for physical measurements, made in the same environment, in order to elaborate an objective metric related to the subjective measures. Such distortion metrics for distortion audibility and preference are available in the literature. The DS metric and the R_{nonlin} provided by Tan *et al.* [4, 8] correlated the audibility of distortion in musical and speech signals and objective measures of nonlinearity. Geddes and Lee [9, 10] proposed the GedLee metric explaining the relation with perceive preference over different types of distortion. Our final goal is to provide a metric able to link perceived dimensions with objective ones by adapting the protocol of Lavandier *et al.* [6, 7] to the evaluation of perceived differences over a large panel of distorting loudspeakers.

Modeling the behavior of the restitution of a loudspeaker is widely used when studying the perception of distortion. Two main approaches for modeling are used. In the "physical" approach, the cause of each local nonlinear phenomenon is studied and leads to an approximated nonlinear law representing a nonlinear relation between two measured variables of the loudspeaker driver. For example, Klippel [11] displays the force factor, the voice coil inductance or the suspension stiffness as functions of the voice coil displacement. These nonlinear relations are used in a model of loudspeaker [12, 13, 14] based on a equivalent electrical representation including different nonlinear elements. The distortion is introduced by modifying in the model the amplitude of each known nonlinear parameters. In the "artificial" approach, the distortion model is not physically related to a loudspeaker but consists in adding distortion to an original signal. The distortion stems from an arbitrary Input/Output nonlinear relation allowing the creation of various types of nonlinearities. For example, Tan *et al.* [4] studied a variety of artificial distortion such as hard and soft clipping meant to resemble distortions generated in loudspeakers. Geddes and Lee [10] generated different distorting musical samples by modifying the I/O relation along different modifications such as Taylor series, a zero crossing discontinuity or Fourier series.

In the present study, a global approach was preferred. This third approach does not consist in modeling individual nonlinear phenomenon of a driver like the physical modeling, but allows to model the whole loudspeaker as a black box. It is based on two steps of identification and synthesis. The identification is done with the swept sine technique presented by Farina [15] and modified by Novák *et al.* [16] allowing the separation of the nonlinear higher order frequency responses characterizing the system behavior. The synthesis step involves the expansion of a musical signal over the basis of Chebyshev polynomials combined with the identified frequency responses for each higher order [17]. The modeling process used here offers a compromise between the physical and the artificial approach. First, it allows the identification of the loudspeaker nonlinearities leading to a controllable synthesis of its behavior. Secondly, the expansion of the signal on the base of Chebyshev polynomials could be applied to fit any artificial I/O transfer function. Using this model, the creation of various types of loudspeakers nonlinearities with different kinds and levels is possible.

The present paper is focused on the creation of virtual loudspeakers when studying the nonlinearities generated by the woofer restitution of a two ways loudspeaker. The first part concerns the separation of the tweeter and woofer sound fields. Then, the identification and synthesis of the woofer nonlinear model are detailed. Finally, a listening test is presented to validate our proposed synthesis of virtual loudspeakers. The creation of a large panel of different virtual loudspeakers and the dissimilarity listening tests are ongoing and the perceptual results will be presented later.

2. Creation of virtual loudspeakers

2.1. Separation woofer/tweeter and recording

In order to study the distortions induced by the woofer restitution, the first step consisted in separating the tweeter and the woofer restitution. Recording both ways simultaneously would have led to interferences between recordings. To measure independently the tweeter and the woofer sound fields, a pair of identical two ways loudspeakers were used inside and outside of an anechoic chamber. We call loudspeaker 1, the recorded loudspeaker placed inside the anechoic chamber, and loudspeaker 2 the one outside. Basically, we recorded one driver of loudspeaker 1 while the signal from the second crossover output was sent to the driver of loudspeaker 2 in order to respect the crossover electrical behavior. For example, when measuring the woofer sound pressure of loudspeaker 1, the low pass output was sent to woofer 1, and the high pass output was sent to tweeter 2.

Using this technique, the acoustic and the electric behaviors of the recorded loudspeaker could be considered as identical to the original loudspeaker. The recordings were conducted in an anechoic chamber following the principle described above. A Tannoy System 600 loudspeaker was chosen, as its woofer and tweeter drivers are concentric. The restitution and the recording were done with a Fostex VC-8 audio converter with a sampling frequency of 44100 Hz. A total of 11 different musical and speech excerpts were recorded using the following steps.

- Woofer recording : A 1/4" G.R.A.S 40BF microphone and a G.R.A.S. 26AC preamplifier were used on the axis of the driver at a distance of 2 mm.
- Tweeter recording : A Texen "Statex P48" static studio microphone was used at a distance of 20 cm, on the axis of the tweeter.

2.2. Nonlinear woofer model

The woofer model results from two complementary steps : the woofer identification and its synthesis.

2.2.1. Identification

The identification part consists in the analysis of the nonlinear behavior of the woofer restitution. The method is based on the nonlinear convolution method presented by Farina [15] and modified by Novák *et al.* [16]. This method uses a swept-sine as the input signal and allows the characterization of an arbitrary nonlinear system. A brief description is given in this section but the method is fully described in [16].

The exponential swept-sine signal $x_s(t)$ is generated and used as the input signal of the nonlinear system. The recorded output signal $y_s(t)$ is processed through the "nonlinear convolution" process. This convolution is performed between $y_s(t)$ and $\tilde{x}_s(t)$ which corresponds to the time-reversed $x_s(t)$ with an amplitude modulation such that $x_s(t)$ convolved with $\tilde{x}_s(t)$ leads to a Dirac function. The result of the convolution between this inverse filter $\tilde{x}_s(t)$ and the output signal $y_s(t)$ may be split as a set of higher-order impulse responses $h_n(t)$, Fourier transformed into $H_n(f)$. These frequency responses represent the frequency behavior of higher-order components. $H_1(f)$ is the frequency response of the linear part of the system and for higher orders ($n > 1$), the $H_n(f)$ correspond to nonlinear parts.

In our case, the identification process is performed by recording the swept sine played through the woofer. By applying the separation technique mentioned above, only the woofer driver is playing and recorded in the anechoic chamber allowing to identify its behavior independently from the tweeter. The input swept sine was arbitrarily set to cover the frequency range from 1 to 4400 Hz knowing that the crossover cut off frequency of the Tannoy 600 loudspeaker is about 1800 Hz. Playing and recording is performed synchronously using the Fostex VC-8 audio interface in order to allow the nonlinear deconvolution process.

The nonlinear identification technique leads to the higher order frequency responses characterizing the linear and nonlinear behavior of the woofer restitution which are introduced into a nonlinear synthesis process.

2.2.2. Synthesis

The synthesis first consists in expanding the input signal on a basis of Chebychev polynomials [17]. They

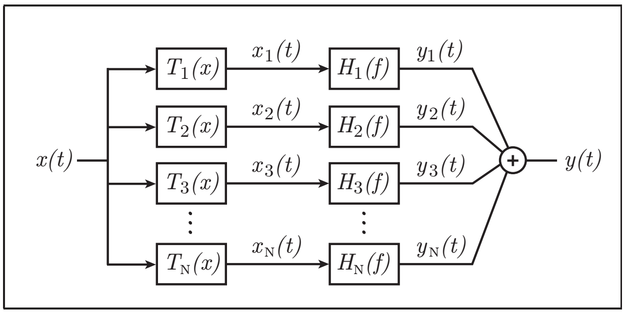

Figure 1. Nonlinear woofer model with Chebyshev polynomials $T_n(x)$ and higher order frequency responses $H_n(f)$, $x(t)$ is the input signal, $y(t)$ the output signal and n the order taking the values from 1 to N.

form a sequence of orthogonal polynomials linked by the relation $T_n(x) = 2xT_{n-1}(x) - T_{n-2}(x)$, with $n = 2, 3, ...$, and the initial conditions $T_0(x) = 1$ and $T_1(x) = x$. Their orthogonality allows an easy an unique expansion procedure.

The expansion of the musical signal on these polynomials is then combined with the identified frequency responses $H_n(f)$ characterizing the woofer behavior. Figure 1 shows the block diagram illustrating the non linear model used for the synthesis of the woofer. This nonlinear synthesis process allows to simulate the nonlinear behavior of the woofer excited by any input signal.

2.3. Recomposition and auralization

The virtual loudspeaker results from the combination of the tweeter recording and the woofer nonlinear model. In order to follow the protocol introduced by Lavandier *et al.* [6], the virtual loudspeaker should be evaluated in a listening room. Bose [18] said "We have no satisfactory correlation between any of these loudspeakers distortion measurements and their subjective effects on music reproduction and we realize the need for a method of detecting distortion in a way that is meaningful to the listener in the environment for which the speaker is intended." Cabot [19] explains that the room characteristics could have some masking effects on certain distortion products or could enhance some others. Being aware of the room influence, Boer [3] studied the audibility of nonlinear distortion of loudspeakers in a listening room and Kristoffersen *et al.* [13] evaluated the preference of distorting loudspeakers in a "typical Swedish sized living room". Even though anechoic recordings would appear more precise in terms of perception of distortion, the decision was taken to evaluate the perceived distortion in a "normal" and realistic listening environment.

To be able to compare the virtual loudspeakers as they would be in a listening room, the auralization technique was applied to the virtual loudspeaker signals. This method consists in convolving an anechoic

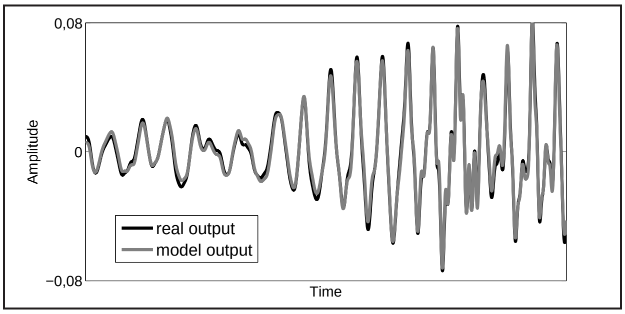

Figure 3. Comparison between actual loudspeaker output and virtual loudspeaker output for the same sample of musical input signal.

signal with a room impulse response (IR) in order to add the room reverberation characteristics, leading to realistic binaural content. Indeed, the virtual loudspeaker results from an anechoic tweeter recording and a woofer model for which the identification is done in an anechoic environment. By convolving the virtual loudspeaker signal with the IR of the room, the result is close to the signal as it would be heard in the room when played through the virtual loudspeaker. Figure 2 summarizes the whole process of creation of a virtual loudspeaker and the comparison with an actual loudspeaker.

3. Validation of protocol

3.1. Objective validation

To validate the synthesis process, woofer model outputs are compared to actual woofer outputs when both are excited by the same signals. Here, a preliminary comparison was made over four audio samples. The model order is set to seven for the woofer synthesis. The four musical materials retained are :

- a male voice recorded in an anechoic chamber, 2.4s
- a choir (Fransisco Guerrero, "Requiem"), 4s
- classical music (W.A. Mozart, "Flute Quartet"), 3.2s
- guitar with female voice (Rebecca Pidgeon, "Grandmother"), 3.8s.

An objective error criterion is the relative mean square error (MSE) between the real output $y_R[i]$ and the model output $y_S[i]$. This relative mean square error is averaged over all the samples. For the four musical excerpts, the average MSE is about 5%. The MSE value shows a good agreement between the woofer synthesis and the actual woofer for musical signals. As a visual example, Figure 3 displays both recorded and synthesized temporal signals.

3.2. Perceptual validation

As a preliminary step, a listening test consisted in comparing an actual loudspeaker recorded inside an actual room and the virtual loudspeaker convolved with the impulse response of the same room. The aim was to evaluate if the virtual loudspeaker was perceived similarly to the actual loudspeaker.

3.2.1. Procedure

To evaluate if a difference was perceived between actual and virtual stimuli an ABX test was conducted. For each trial, the listener heard three sounds. The sounds A and B were respectively an actual and a virtual stimulus and the sound X was A or B. The task consisted in determining if X was identical to A or to B. For each musical excerpt, all combinations of presentation were evaluated. The combinations ABA, ABB, BAA and BAB were presented to the listener in a random order. For each listener, the entire test consisted of 16 trials (4 excerpts and 4 combinations) presented in different random orders.

The validation test was performed using the same four musical excerpts as for the objective validation, played either by the actual loudspeaker or by the auralized virtual loudspeaker.

- 4 actual stimuli : recordings of the Tannoy 600 in a suitable listening room with the AB-ORTF binaural recording technique. The distances between the microphones and the loudspeaker was 2.20 m. These recordings were those made by Lavandier *et al.* [20].
- 4 virtual stimuli : musical excerpts passed through the virtual loudspeaker process. To apply the same room characteristics, the binaural IR was obtained from the actual loudspeaker recorded inside the actual room when playing a swept sine.

Twelve normal hearing listeners took part in the listening test. They were members of the laboratory or students.

Listening tests were conducted with a user interface made of 3 virtual buttons. During each trial, the three sounds were successively played to the listeners, who could then listen to any of the sounds as many times as they wanted, by clicking on the corresponding button. Listening tests were run in a soundproof room, and the audio stimuli were reproduced through headphones (Stax SR Lambda Professional). The headphone restitution is a commonly used technique which appears to be a good compromise between numerous parameters to be controlled during the loudspeaker evaluation. The distortion of the headphones was assumed low enough compared to the loudspeaker one.

3.2.2. Results

The ABX test result corresponds to the percentage of right answers given by listeners. Table I displays the score of each subject and its experience in the evaluation of audio stimuli. By averaging the scores over every listeners, the ABX test leads to 52% of right answers with a standard deviation of 16%. Listener 12 is an highly trained listener in audio evaluation. His score shows that our technique of creation of virtual

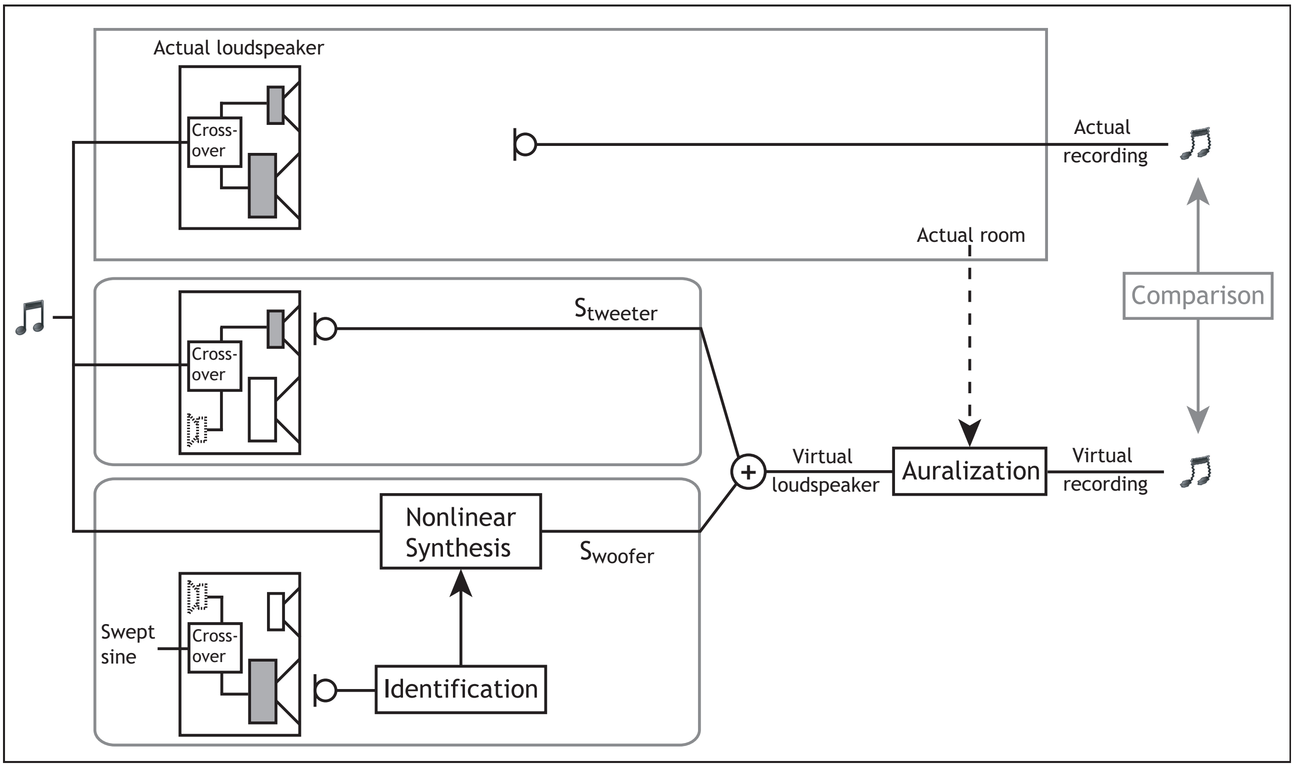

Figure 2. Protocol of creation of virtual loudspeakers : from a musical signal to an auralized virtual loudspeaker made of a tweeter recording and a woofer model. Comparison with the actual loudspeaker recording.

loudspeaker is not perfect and differences remain between a virtual and an actual stimulus. This listener experience could be very helpful in improving the process until he would not be able to detect the difference between an actual and a virtual stimulus. The purpose of our study consists in evaluating loudspeaker nonlinearities with a panel of listeners not specialized in loudspeaker evaluation. If the score made by listener 12 is removed, the ABX test result falls to 48% with a standard deviation of 9% which confirms that, for listeners with none and medium experience, there is no significant perceived difference between an actual and a virtual loudspeaker.

4. PERSPECTIVES

The presented protocol for creating a panel of "virtual loudspeakers" seems to be an efficient tool to evaluate the distortion of loudspeakers. For the evaluation of the nonlinearities generated by the woofer restitution, our protocol allows to separate the woofer and tweeter radiations, to model the woofer behavior using a nonlinear identification and synthesis, and finally to auralize the virtual loudspeaker signal.

The preliminary listening test indicates that the auralized virtual loudspeaker is very similar to the actual loudspeaker. The presented protocol allows us to simulate the restitution of an actual loudspeaker. The aim of our study is to evaluate the loudspeakers nonlinearities and the protocol also allows to deviate

Table I. Scores and experience in audio evaluation of each listener.

Listeners	Audio evaluation experience	Score (%)
1	None	37.50
2	Medium	37.50
3	None	43.75
4	Medium	43.75
5	None	43.75
6	Medium	43.75
7	None	43.75
8	None	50.00
9	None	56.25
10	None	62.50
11	Medium	62.50
12	Very High	93.50

from the actual loudspeaker and create several virtual loudspeakers with different nonlinearity types and levels. The nonlinear woofer model is made of several branches representing the nonlinearity orders, each of them combining one polynomial associated with its frequency response. Each branch can be modified by enhancing or reducing the influence of a specific order. Our project is now to use different I/O nonlinear relations offering different nonlinearity types. Using the expansion on Chebyshev polynomials, the polynomials may be identified and then increased or decreased in the woofer model. This results in a set of virtual loudspeakers made of a tweeter recording and a controlled distorting woofers.

To evaluate loudspeakers nonlinearities, a large panel will be created in order to study an auditory space containing many different distorting loudspeakers. Once the panel is created, listening test will be conducted to evaluate the perceived differences between loudspeakers. The paired comparison is the standard method allowing a measure of dissimilarity between each pair of stimuli. When evaluating large panels, this method is not suitable. The evaluation of the large panel of virtual loudspeakers will be conducted with the method of comparison to a reference [21, 22, 23]. This method consists in presenting on each trial a reference stimulus and three comparison stimuli. The listener is asked to choose which of the comparison stimuli is the most similar to the reference. For the entire test, every stimulus appears as the reference, which offers the possibility to conduct a test in several sessions. The resulting dissimilarity data will be then analyzed with a MDS technique to investigate the underlying dimensions of the distortion space.

Acknowledgement

The authors are grateful to the listeners who took part in the listening test.

References

[1] E. Czerwinski, A. Voishvillo, S. Alexandrov, A. Terekhov: Multitone testing of sound system somponents - some results and conclusions, Part 1: History and Theory. J. Audio Eng. Soc, **49** (2001) 1011-1048.

[2] M. Petri-Larmi, M. Otala, J. Lammasniemi: Psychoacoustic detection threshold of transient intermodulation distortion. J. Audio Eng. Soc, **28** (1980) 98-105.

[3] M.A. Boer, A.G.J. Nijmeijer, H. Schurer, W.F. Druyvesteyn, C.H. Slump, O.E. Hermann: Audibility of nonlinear distortion in loudspeakers. Proc. 104th AES Convention, Amsterdam, 1998.

[4] C.T. Tan, B.C.J. Moore, N. Zacharov: The effect of nonlinear distortion on the perceived quality of music and speech signals. J. Audio Eng. Soc, **51** (2003) 1012-1031.

[5] J. Rauhala, J. Ahonen, M. Tikander, M. Karjalainen: Perception and physical behavior of loudspeaker nonlinearities at bass frequencies in closed vs. reflex enclosures. Proc. 124th AES Convention, Amsterdam, 2008.

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

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

[8] C.T. Tan, B.C.J. Moore, N. Zacharov, V.V. Mattila: Predicting the perceived quality of nonlinearly distorted music and speech signals. J. Audio Eng. Soc, **52** (2004) 699-711.

[9] E.R. Geddes, L.W. Lee: Auditory perception of nonlinear distortion - Theory. Proc. 115th AES Convention, New York, 2003.

[10] L.W. Lee, E.R. Geddes: Auditory perception of nonlinear distortion. Proc. 115th AES Convention, New York, 2003.

[11] W. Klippel: Loudspeaker nonlinearities - symptoms, parameters, causes. Proc. 119th AES Convention, New York, 2005.

[12] R. Schmitt: Audibility of nonlinear loudspeaker distortions. Proc. 98th AES Convention, Paris, 1995.

[13] R. Kristoffersen, M. Kleiner, D. Västfjäll: Subjectively perceived sound quality in audio systems as a function of distribution and number of loudspeakers used in playback. Proc. 106th AES Convention, Paris, 1999.

[14] W. Klippel: Speaker Auralization - Subjective evaluation of nonlinear distortion. Proc. 110th AES Convention, Amsterdam, 2001.

[15] A. Farina: Simultaneous measurement of impulse response and distortion with a swept-sine technique. Proc. 108th AES Convention, Paris, 2000.

[16] A. Novák, L. Simon, F. Kadlec, P. Lotton: Nonlinear system identification using exponential swept-Sine signal. Proc. 2010 IEEE Int. Conf. on Instrumentation and Measurement, 2220-2229.

[17] A. Novák, L. Simon, P. Lotton, J. Gilbert: Chebyshev model and synchronized swept sine method in nonlinear audio effect modeling. Proc. 13th Int. Conf. on Digital Audio Effects (DAFx-10), Graz, 2010.

[18] A.G. Bose: On the design, measurement, and evaluation of loudspeakers. Proc. 35th AES Convention, New York, 1968.

[19] R.C. Cabot: Audible effects vs. objective measurements in the electrical signal path. Proc. 8th AES Conference, Washington, 1968.

[20] M. Lavandier, S. Meunier, P. Herzog: Perceptual and physical evaluation of differences among a large panel of loudspeakers. Proc Forum Acusticum, Budapest, 2005.

[21] P.Y. Michaud, S. Meunier, P. Herzog, G. Drouet d'Aubigny, M. Lavandier: M?ode de test adapt?à l'?luation perceptive d'un grand nombre de stimuli audio : Application aux enceintes acoustiques. Proc. 10ème Congrès Fran?s d'Acoustique, Lyon, 2010.

[22] P.Y. Michaud, S. Meunier, P. Herzog, M. Lavandier, G. Drouet d'Aubigny: A listening test method to evaluate the dissimilarity of large sets of stimuli: Application to loudspeakers. Abstract, J. Acoust. Soc. Am. **128** (2010), 2455-2455 .

[23] P.Y. Michaud, S. Meunier, P. Herzog, M. Lavandier, G. Drouet d'Aubigny: A listening test method to evaluate the dissimilarity of large sets of stimuli: Application to loudspeakers. J. Acoust. Soc. Am. (Submitted).