Liski, Juho; Mäkivirta, Aki; Välimäki, Vesa

Audibility of loudspeaker group-delay characteristics

Published in:
144th Audio Engineering Society International Convention 2018

Published: 01/01/2018

Document Version
Peer reviewed version

Please cite the original version:
Liski, J., Mäkivirta, A., & Välimäki, V. (2018). Audibility of loudspeaker group-delay characteristics. In *144th Audio Engineering Society International Convention 2018* (pp. 879-888). Audio Engineering Society. http://www.aes.org/e-lib/browse.cfm?elib=19404

Audio Engineering Society
Convention Paper
Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Audibility of Loudspeaker Group-Delay Characteristics

Juho Liski[1], Aki Mäkivirta[2], and Vesa Välimäki[1]

[1]*Aalto University, Acoustics Lab, Dept. of Signal Processing and Acoustics, P.O. Box 13000, FI-00076 AALTO, Espoo, Finland*
[2]*Genelec Oy, Olvitie 5, FI-74100 Iisalmi, Finland*

Correspondence should be addressed to Juho Liski (juho.liski@aalto.fi)

ABSTRACT

Loudspeaker impulse responses were studied using a paired-comparison listening test to learn about the audibility of the loudspeaker group-delay characteristics. Several modeled and six measured loudspeakers were included in this study. The impulse responses and their time-reversed versions were used in order to maximize the change in the temporal structure and group delay without affecting the magnitude spectrum, and the subjects were asked whether they could hear a difference. Additionally, the same impulse responses were compared after convolving them with a pink impulse, defined in this paper, which causes a low-frequency emphasis. The results give an idea of how much the group delay of a loudspeaker system can vary so that it is unlikely to cause audible effects in sound reproduction. Our results suggest that when the group delay in the frequency range from 300 Hz to 1 kHz is below 1.0 ms, it is inaudible. With low-frequency emphasis, the group delay variations can be heard more easily.

1 Introduction

In order not to change the nature of the electronic input waveforms in transforming them to output pressure variations of a loudspeaker, a constant gain across the audio frequency band as well as a constant input-to-output latency is required. While several methods exist for magnitude equalization [1, 2], delay equalization for creating constant-latency system responses use filters to create symmetrical system impulse responses at least within the most critical audible frequency range. An FIR filter or a collection of IIR allpass filters can be used to equalize in the time domain traditional minimum-phase systems.

A constant-latency system response has a symmetrical impulse response with the same of amount of energy before and after the largest peak in the impulse response. Increasing the order of such a system, for example with the aim to reduce the delay variation or in order to increase the crossover filter roll-off rate, tends to increase the length of the system impulse response. This has lead to discussions about the effect called "pre-echo" with speculation that parts of the impulse response, starting very early before the largest peak, might become audible as a separate auditory event apart from the main system response, and therefore constitute a change in the character of a loudspeaker.

The human auditory system presents some masking before and after an auditory event [3]. The premasking, i.e. the masking occurring before the auditory event, is particularly relevant when discussing the potential for a pre-echo. Unfortunately, while premasking is known to exist, data available on the level of premasking as a function of time is not very exact, and different sources do not agree on the level and extent of premasking [4, 3].

A constant-latency design or time-domain equalization can reduce the delay variation within a given frequency range down to a constant delay at any required accuracy. The relevant goal for the design of such an equalizer is to limit the delay variations so that they stay below the just-audible limit. Furthermore, the related increase in the time extent and level of the early parts of a system impulse response should be limited sufficiently such that they remain inaudible due to premasking in order to ensure absence of any pre-echo.

This paper investigates the audibility of loudspeaker group-delay characteristics, i.e. the length of the loudspeaker impulse response, by using original and time-reversed signals in a listening test. Previously, Lipshitz *et al.* studied the audibility of midrange phase distortions [5]. They used both headphones and loudspeakers to reproduce artificial and recorded sounds with and without allpass equalization. Similarly, Flanagan *et al.* performed listening tests with headphones and loudspeakers in order to discover the threshold for perceiving group-delay distortion in click-like signals [6]. Also, listening tests have been conducted to determine the audibility of phase equalization in loudspeakers by Greenfield *et al.* [7]. They used a minimum-phase equalizer to correct the magnitude response of the loudspeaker, after which the excess-phase component was equalized with a phase equalizer in order to achieve linear-phase. Test subjects listened to a stereo pair of loudspeakers with and without the equalization and gave informal comments on the sonic impression.

Krauss studied the audibility of group-delay distortion caused by crossover networks with headphones [8]. The group delay of the crossover networks was simulated with allpass filters, and tone bursts were compared with and without equalization. On the other hand, Karjalainen *et al.* studied the audibility of the loudspeaker group delay in an informal test [1]. They used digital allpass filters in order to produce group-delay distortions, and listening test subjects compared signals with and without the phase equalization with each other. Time-reversed signals have been used previously by Kulkarni *et al.* in listening tests related to head-related transfer functions (HRTF) [9]. They studied the sensitivity of human hearing on the HRTF phase spectra, and the time reversal produced the maximum phase difference while simultaneously keeping the magnitude response unchanged.

This paper is organized in the following way. Section 2 briefly discusses previous perceptual studies on short signals, and especially on temporal masking. Section 3 explains the listening test conducted in this study, and defines the pink impulse, which is used for emphasizing low and mid frequencies in the listening test. Section 4 presents the results of the listening test. These results are further discussed and analyzed in Section 5. Section 6 concludes this work.

2 Temporal Masking

Temporal masking refers to a masker sound partially or completely hiding other sounds occurring either before or after the masker [3]. The former phenomenon is called pre-masking (or backward masking) and the latter post-masking (forward masking). Temporal masking differs from frequency masking, where the masker and the masked sounds are present simultaneously.

Temporal masking has been studied with different types of sound signals. Typical test signals have been clicks, impulses, sinusoidal tones, tone bursts, and noises of varying length and bandwidth. Clicks and impulses try to minimize the signal length while creating a wide spectrum [3]. Tone bursts shaped with a Gaussian time window can have a short duration as well as a narrow-band spectrum [3]. White noise can be used in a situations in which a wide-band signal of a longer duration is required.

First, the temporal masking caused by clicks is discussed. Raab studied the masking caused by acoustic clicks in 1961 [10]. He found that both forward and backward masking occurred, and that the forward masking was more prominent and longer lasting: the duration for backward masking is approximately 15 ms whereas for forward masking it can last even for 100 ms. Both effects were also level dependent. Olive *et al.* also had clicks in their study of reflection perception [11]. They found that vertical and lateral reflections caused forward masking for at least 20 ms. However, at the same time they remarked that the masking effect is much greater when the two sounds arrive from the same direction [11]. Sporer *et al.* used a measurement system for perceptual audio coding to study forward and backward masking [12]. In their study, backward masking caused by clicks extends to approximately 20 ms and forward masking to approximately 110 ms. The system models the temporal behavior of human ears in an adequate way, and thus these masking results are also valid for humans.

Next, the masking properties of tones or tone bursts are reviewed. In these situations, masking is dependent on the frequencies in addition to the dependency on the masker level and the time delay [13]. Fastl studied both forward and backward masking of sine tones [14]. His results show that forward masking extends to over 100 ms, whereas backward masking is only relevant for approximately 10 ms. Jesteadt *et al.* studied forward masking caused by sine tones on sine tones [13]. They found that forward masking is greater at low frequencies, and that with all frequency combinations, notable masking extends to less than 100 ms.

Finally, the masking effects of noise and noise bursts are examined. Wilson *et al.* studied the masking caused by white noise on clicks [15]. They were interested in the additivity of masking. However, they also found the amount of backward and forward masking: in their test forward masking extends to over 150 ms whereas backward masking reaches low levels already after approximately 20 ms. Dolan *et al.* studied backward masking with noise masker and sinusoidal tones [4]. They found that the masking decays more rapidly with increasing tone frequency, and that the majority of the masking occurs 5 ms or less before the masker. Fastl *et al.* discovered the forward masking effect of white noise on Gaussian pulses and test tone bursts [3]. In both cases masking extends over 100 ms, but the amount of masking is higher in the former case. An important result related to the masking caused by white noise was that the amount of masking depends on the length of the noise masker [3].

3 Listening-Test Procedure

In this work, listening tests have been conducted to find whether hearing a difference between the loudspeaker response and its modification, which has the same magnitude response but a different group delay, is possible. The modifications included time-reversed and group-delay equalized versions of the responses. In addition to the impulse responses themselves, filtered versions of the responses were presented to the subjects. We discuss next the selection of the test signals, the proposed filtering process, and the listening test arrangements.

3.1 Test Signals

The test signals for the conducted listening test were loudspeaker impulse responses obtained either from loudspeaker models or measurements. The models were one-way or two-way models with varying parameters based on digital filters imitating the mechanical low (50 Hz) and high frequency (20 kHz) cutoffs as well as the possible crossover filters, as described in detail in a companion paper [16]. In addition, an extended model for a three-way loudspeaker was used based on the same principles as the other two models. The measured loudspeaker responses contained six different high-quality two- and three-way models from six different manufactures. The impulse response measurements were performed using the MLSSA software, which yielded the signal-to-noise ratio of approximately 50 dB.

Finally, the listening test also contained cases in which two sounds in a pair were identical and no difference was supposed to be heard. Such cases were included to test how reliable the subjects are. The specification of each test signal is given in Table 1.

The parameters that were varied in the loudspeaker models included the low cutoff frequency, the crossover frequency or frequencies, crossover filter type, and the group delay. The group delay was exaggerated in the same cases using FIR filters with an almost allpass response, i.e. the magnitude response was an allpass response at the passband of the loudspeaker model, but near the Nyquist limit the response was attenuated by a raised cosine function similarly as in [17]. The phase of the FIR filter was designed in a way that would result in a doubled or tripled group-delay response in comparison to the original group delay of the loudspeaker model.

3.2 Pink Impulse

The loudspeaker impulse responses themselves sound like clicks. Since their spectrum is approximately flat, apart from the highpass cutoff at low frequencies, they tend to sound very bright. As the phase differences appear mainly at low and mid frequencies, one idea is to filter the impulse response to attenuate the high-frequency content and to emphasize the lower frequencies. There is a similarity here to the relation between white and pink noise, as white noise sounds excessively bright whereas pink noise sounds approximately flat, since it contains a constant amount of sound energy per octave.

Table 1: Specification of the test signals used.

Signal	Description
Id	Hidden reference, a pair of identical impulse responses
1Wa	Impulse response of a 1-way loudspeaker model
1Wb	1-way model, higher low cutoff frequency (100 Hz)
2Wa	2-way model, crossover at 1 kHz
2Wb	2-way model, crossover at 2 kHz
2Wc	2-way model, crossover at 3 kHz
2Wd	2-way model, FIR crossover at 1 kHz
2We	2-way model, FIR crossover at 2 kHz
2Wf	2-way model, FIR crossover at 3 kHz
2Wg	2-way model, crossover at 1 kHz, doubled group delay
2Wh	2-way model, crossover at 3 kHz, doubled group delay
3Wa	3-way model, crossovers at 0.5 and 3 kHz
3Wb	3-way model, crossovers at 0.8 and 5 kHz
3Wc	3-way model, crossovers at 0.5 and 3 kHz, doubled group delay
3Wd	3-way model, crossovers at 0.5 and 3 kHz, tripled group delay
LS1–LS6	Measured loudspeaker impulse responses

Motivated by the relation between the white and pink noise, we propose to define a pink impulse and convolve all loudspeaker impulses with it before listening. The pink impulse was previously used to test artificial reverberation algorithms [18]. We define the pink impulse as a zero-phase (linear-phase) signal having the following spectrum:

$$H(\omega) = \frac{1}{\sqrt{\omega}}, \tag{1}$$

where $\omega = 2\pi f$ and f is the frequency (Hz). This spectrum decays by 3.1 dB per octave. The corresponding time-domain signal is infinitely long and symmetrical.

It is easy to synthesize a pink impulse of finite length using frequency sampling and windowing techniques. The spectrum of Eq. (1) is first sampled at sufficiently many frequency points at both positive and negative frequencies (remembering that the zero and Nyquist frequency points appear only once, i.e., they do not have mirror images). The spectrum at zero frequency cannot be computed using Eq. (1), however, because of the divide-by-zero error. To avoid this problem, we flatten the spectrum at the low end: we copy the spectrum value of the lowest non-zero frequency point and use this same value also at the zero frequency. The pink impulse response is then obtained using the

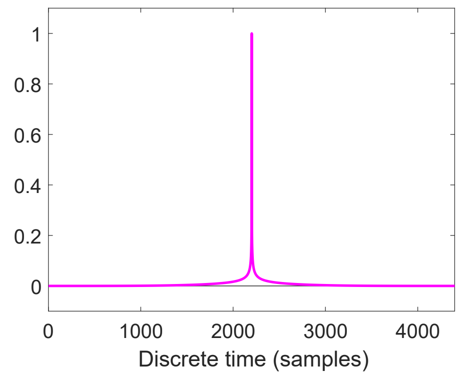

Fig. 1: Pink impulse obtained using the frequency sampling technique, truncated using a Blackman window, and time shifted.

inverse discrete Fourier transform. A window function can be used to symmetrically taper its beginning and end to smoothly truncate the pink impulse to a desired length. Finally, the pink impulse is shifted in time to make it causal.

Figure 1 shows a single pulse shape of the pink impulse obtained using the above method. The spectrum has been sampled at 16,384 frequency points, containing both the negative and the positives frequencies, and has been converted to the time domain using the

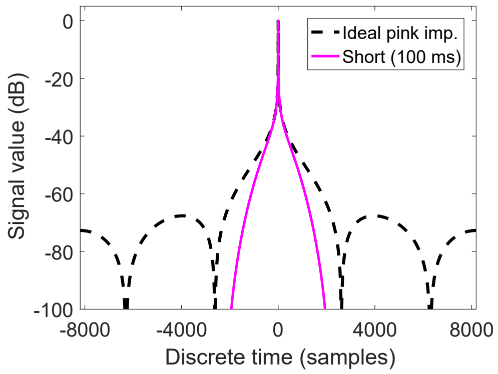

Fig. 2: Comparison of the ideal and truncated (4401 samples) pink impulses in the time domain on the logarithmic scale.

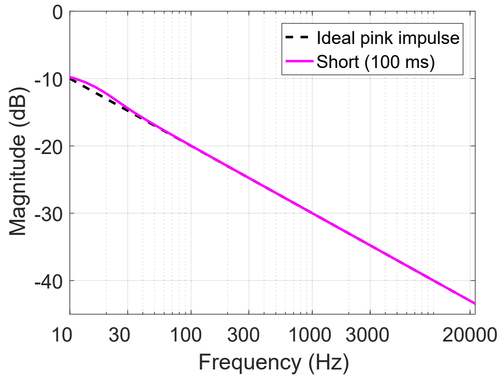

Fig. 3: Comparison of the power spectra of the ideal and truncated (4401 samples) pink impulses.

16,384-point inverse FFT (Fast Fourier Transform). A Blackman window has been used to smoothly truncate the pink impulse to a length of 4401 samples, which corresponds to 100 ms at the sample rate of 44.1 kHz. In Fig. 1, the windowed pink impulse has been shifted by 2200 samples so that it starts at zero.

Figure 2 compares the ideal (i.e., infinitely long) and the shortened pink impulse in the time domain on the logarithmic scale. The shortened pink impulse is seen to have neither sidelobes nor zero-crossings. Around the center peak, the ideal and short versions are identical.

Figure 3 compares the spectra of the ideal pink impulse and the shortened one. Other than the minor deviations that appear at low frequencies below about 100 Hz, they are identical. Since the deviations are smaller than 1.0 dB at frequencies above 10 Hz, the truncated pink impulse sound can be assumed to be the same as the ideal one. In this work, the pink impulse shortened to a length of 4401 samples are used to process the loudspeaker impulse responses for the listening test.

3.3 Listening-Test Operation

The listening test was implemented with the MATLAB software. A screenshot of the graphical user interface (GUI) is shown in Fig. 4. Since the target of the listening test was to find whether the subjects could hear differences between two sounds in a paired-comparison test, the operation of the GUI was simple. The two sounds of a pair were reproduced one after the other

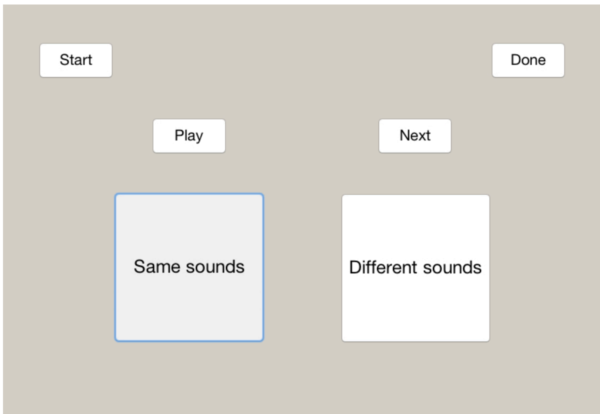

Fig. 4: Screenshot of the listening test GUI.

using Sennheiser HD-650 headphones in a quiet office room when the subject pressed "Play". Each sound pair could be repeated once, if the subject so desired. Next, the subjects had to report, whether they perceived the two sounds as the same or different from each other. This was done by pressing one of the large buttons in the GUI seen in Fig. 4.

The test contained 21 different sound signals, as specified in Table 1. Each signal was presented before and after convolving it with the pink impulse. Thus, the spectra of the signals were either flat in the passband (unconvolved) or they had a low-frequency emphasis due to the convolution with the pink impulse. The reproduced sound-signal pairs contained an impulse response and its time-reversed version. The length of each sound signal was 1.6 s, with the first impulse response located at 0.3 s and the second at 1.3 s (the maximum values of the impulses are centered to these time points). Each sound pair was presented twice, the second time in the opposite order. Finally, the presentation sequence of all the sound pairs was randomized individually for each test subject.

Five individuals (two females and three males), aged from 24 to 39 years, with normal hearing participated in the listening test. All had prior experience in different types of listening tests. In order to familiarize the subjects with the GUI as well as the nature of the test signals, a short training session similar to the actual test was performed with the test subjects before the actual test. The results of the training session were deleted.

The subjects were asked to report if the sounds in a pair were different, but at the same time, to not concentrate on level differences. Technically, there were no level differences due to the nature of the signals to be compared. Finally, they were also told that some test sounds are identical and that they should only respond "different" when they perceived a difference.

4 Analysis of Listening-Test Answers

The results of the listening tests are shown in Figs. 5 and 6. The bars indicate the proportion of the two samples judged to be the same in all responses. Thus, values close to unity mean that the test subjects perceived the sounds as the same, whereas values close to zero mean that most of the respondents judged the two sounds as being different. The bars have been sorted according to the mean of all results, i.e. when the results from the original impulse responses and the filtered ones are combined.

The bars are color coded: Similar types of impulse responses share the same color. The impulse responses arising from one-way, two-way, and three-way models each have their own color (orange, blue, and green, respectively), and the measured loudspeaker impulse responses are presented with another color (red). Black indicates the hidden reference or the pair in which the sounds are identical.

Most of the signals produced clear results where the two compared signals were either perceived as the same or as different by the majority of the test subjects. A few signals remained hard to classify, for example "3Wb", which corresponds to a modeled three-way speaker with high crossover frequencies (see Table 1).

When comparing the two cases (Figs. 5 and 6), listening directly to the impulse response or the impulse response convolved with the pink impulse, they generally resemble each other. However, for the impulse responses arising in the model, the overall level of judging the two signals to be the same is lower for the impulse responses convolved with the pink impulse than for the original impulse responses. This verifies the assumption that the group-delay differences are easier to perceive when the low and mid frequencies are emphasized.

In the analysis of the data, we use the proportion levels of 0.20 and 0.80 to divide the results into three groups. A value above 0.80 indicates a high likelihood of detecting the two signals in the pair to be the same. This suggests that the test signal is so short that even the maximal difference in the phase created by temporal inversion does not produce audible differences. In Fig. 5, this group involves the reference and the first eight test signals. These arise from the ideal loudspeaker models representing both of the one-way speakers and most of the two-way speakers. However, in Fig. 6, this group only comprises the reference and the first two test signals.

The second group spans the proportion levels between 0.80 and 0.20. In these cases, the test subjects were not confidently able to detect the impulse response and its time-reversed version. In Fig. 5, this group only involves three signals, whereas Fig. 6 contains six test signals.

The final group contains the test results with a proportion level below 0.20, implying that the test signal and its time-reversed version were regarded mostly (or fully in some cases) as different. In Fig. 5, this group contains most measured impulse responses, the impulse responses arising in the three-way model with lower crossover frequencies, as well as all the signals with exaggerated group-delay responses. In Fig. 6, two additional two-way models as well as the last measured impulse response ("LS3") also belong to this group.

5 Results and Discussion

The results for the hidden reference (the black bar in Figs. 5 and 6) indicate that the listeners could generally

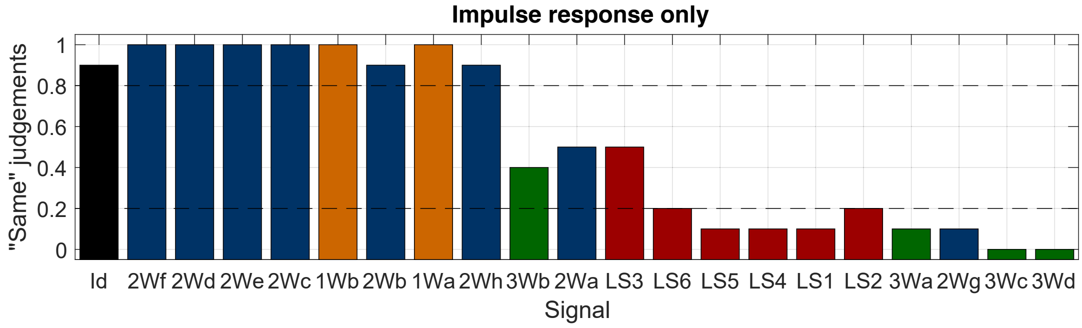

Fig. 5: Proportion of “same” judgments when using impulse responses themselves as the test signal.

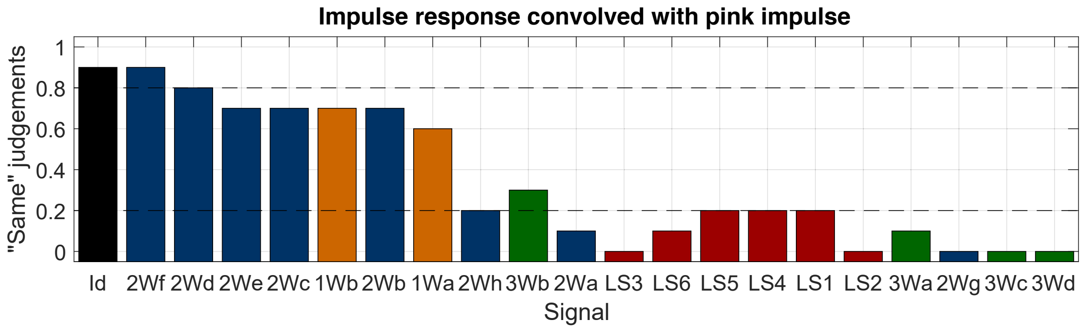

Fig. 6: Proportion of “same” judgments when using impulse responses convolved with pink impulse as the test signal.

correctly detect the two signals being identical. However, in both cases the hidden reference was evaluated slightly below unity (90%). The 20% and 80% limits appear to be suitable for evaluating the definitions of “same” and “different” due to the uncertainty caused by the small sample size.

The differences between Figs. 5 and 6 can be explained with the pink impulse. The pink impulse produces a low-frequency emphasis on the impulse responses due to its magnitude spectrum, thus accentuating possible audible differences in the test signals at low and mid frequencies. Since loudspeakers (both actual loudspeakers and the models used here) generally have the largest group-delay values at low frequencies, i.e. low frequencies need more time to occur in the impulse responses, the pink impulse produces an audible difference in some of the test cases. This, in turn, leads to a lower proportion of the “same” judgments.

Some of the results can be better understood by analyzing the loudspeaker impulse responses. An example of a short impulse response that did not sound different in reverse (the signal “2Wf”) is seen in Fig. 7 on the logarithmic scale. The main part of the impulse response is very short, and the low-frequency tail decays quickly, reaching the −60-dB level in about 12 ms.

In the second group, the impulse responses are definitely longer due to either a lower crossover frequency in the two-way model, an additional crossover filter in the cases of three-way model, or an actual measured loudspeaker. In the last case, however, also some of the two-way models with lower crossover filters as well as both one-way models fall to this group. An example of a synthetic impulse response with an ambiguous length (the signal “3Wb”) is shown in Fig. 8 on the logarithmic scale. The burst of this impulse response after the initial peak is longer than in Fig. 7, but the low-frequency tail decays approximately at the same rate.

The signals arising in the third group have noticeably the longest impulse responses. Therefore, it is not sur-

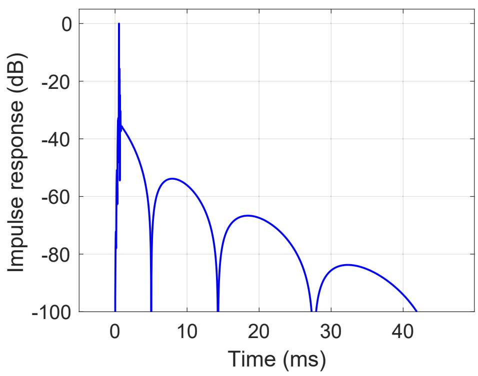

Fig. 7: Synthetic loudspeaker impulse response (signal “2Wf”) that is short and decays quickly.

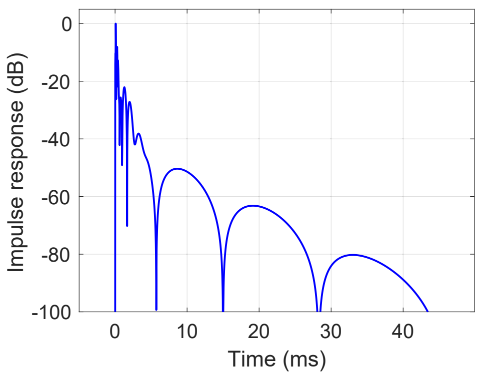

Fig. 8: Synthetic loudspeaker impulse response (signal “3Wb”) in which the difference was hard to perceive.

prising that the time reversal can cause audible differences. Figures 9 and 10 contain two example impulse responses from this group on the logarithmic scale. Especially in Fig. 10, the level of the measured impulse response decreases slower than the level of the loudspeaker models seen in the other figures.

It is interesting that almost all of the test signals containing measured impulse responses belong to the third group. One explanation for this is that the actual measurements of loudspeaker mechanics and electronics create longer impulse responses than the ideal models. Another explanation could be that the measurements contain other information, such as additional background noise or other disturbances, in addition to the loudspeaker response, and this cause audible differences. Analyzing this will require additional work.

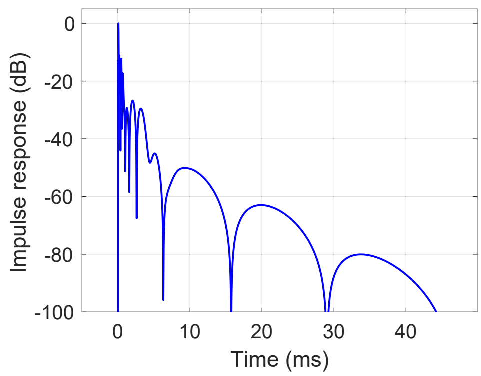

Fig. 9: Synthetic loudspeaker impulse response (signal “3Wa”) in which the difference was perceived almost always.

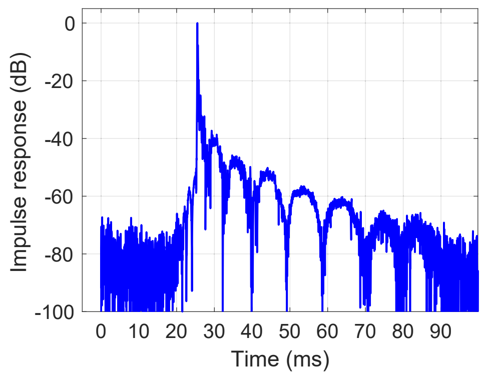

Fig. 10: Measured impulse response of a loudspeaker (test signal “LS2”) that contains noise and decays slowly. Note that a different time scale is used from the other impulse response figures.

Finally, an additional way to compare the test signals is to study their group-delay variations. Figure 11 shows the group delays of the loudspeaker model signals, i.e. all the signals from Table 1 except for “Id” and “LS1–LS6”. They have been normalized so that the group delay at 10 kHz equals zero, since we are interested about group-delay variations that cause the different frequencies of a signal to occur at different times. The group delays are divided into three groups according to the results shown in Fig. 5. Based on these figures, some audibility limits can be suggested.

Figure 11(a) contains the group-delay variations of the signals that were mostly judged “same”, when the signals were played back forwards and backwards.

It is seen that the group delay can exceed 10 ms below 200 Hz when compared to the group delay at high frequencies without the difference being audible. However, an interesting frequency range arises when Fig. 11(a) is compared to the two other subfigures, since the major differences between the three groups of signals are seen between 300–1000 Hz. In Fig. 11(a), the values are below 1.0 ms in this range and the differences are inaudible. In Fig. 11(b), the group delay values are between 1–2 ms, which resulted in the difference being sometimes audible. Finally, in Fig. 11(c), the values are above 2.0 ms and the differences were mostly audible.

6 Conclusion

In this work, loudspeaker impulse responses were studied using a paired comparison listening test to understand the effect of the their temporal structure. Several modeled and six measured loudspeaker impulse responses were included in the test. The impulse responses were compared with their time-reversed versions to maximize the differences in the temporal structure and differences in group delay. Thus, as the magnitude spectra of each test signal pair were identical, the audible differences could be assumed to be related to the temporal structure and group delay.

In addition to impulse responses, their low-frequency emphasized versions were used in the listening test. A pink impulse was defined for this purpose. This is a short, symmetric pulse having a spectral content similar to that of pink noise. Test signals were convolved with a 100-ms long pink impulse to create the low-frequency emphasized versions. Some of the test signals used in this work are available online [19].

The listening test results showed that hearing the differences was easier in the signals which were convolved with the pink impulse, as expected. The effects of the group delay were found to be easiest to perceive in the impulse responses that are long and decay slowly. This corresponds to signals with group delays amounting to above 2.0 ms between 300–1000 Hz. Similarly, perceiving differences was difficult in the impulse responses that are short and decay fast, i.e. where the group delay is less than 1.0 ms between 300–1000 Hz. This work gives an idea of the length of a short enough impulse response such that the temporal structure of the impulse response does not create audible cues.

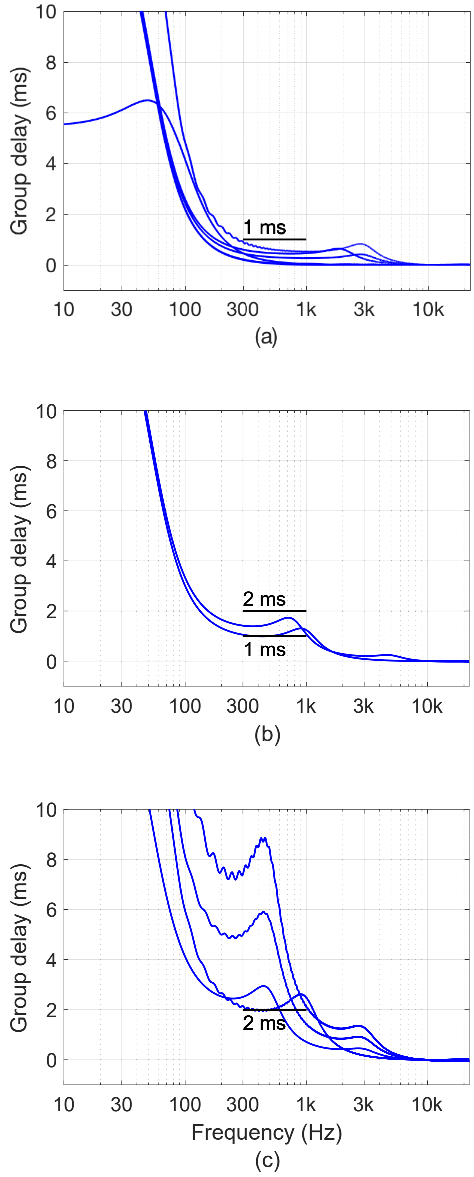

Fig. 11: (a) Zoomed-in group delay variations of the modeled loudspeakers that were mostly judged “same”, (b) that were hard to classify, and (c) that were mostly judged “different”. Note that the group delays have been normalized so that they equal to zero at 10 kHz.

The results of this study should not be interpreted as saying that the group-delay responses of the loudspeakers for which differences were heard are defective and would lead to poor sound quality. The results merely suggest that the group-delay deviations are large enough to create audible differences in an exaggerated test such as the pairwise comparison of a test signal auditioned in the forward and backward directions. However, the results could be interpreted to say that the loudspeaker impulse responses for which differences were not heard in this exaggerated arrangement are sufficiently short and that the group delay probably is not a factor in creating audible effects in sound reproduction.

7 Acknowledgment

The work of J. Liski and V. Välimäki has been supported in part by the IMPRESS project (Aalto University project no. 410682), which is funded by Genelec. The authors would like to thank Luis Costa for proofreading this paper.

References

[1] Karjalainen, M., Piirilä, E., Järvinen, A., and Huopaniemi, J., "Comparison of loudspeaker equalization methods based on DSP techniques," *J. Audio Eng. Soc.*, 47(1/2), pp. 15–31, Jan./Feb. 1999.

[2] Välimäki, V. and Reiss, J. D., "All about audio equalization: Solutions and frontiers," *Appl. Sci.*, 6(5), p. 129, May 2016.

[3] Fastl, H. and Zwicker, E., *Psychoacoustics: Facts and Models*, Springer, 3rd edition, 2006.

[4] Dolan, T. G. and Small, A. M., "Frequency effects in backward masking," *J. Acoust. Soc. Am.*, 75(3), pp. 932–936, Mar. 1984.

[5] Lipshitz, S. P., Pocock, M., and Vanderkooy, J., "On the audibility of midrange phase distortion in audio systems," *J. Audio Eng. Soc.*, 30(9), pp. 580–595, Sep. 1982.

[6] Flanagan, S., Moore, B. C. J., and Stone, M. A., "Discrimination of group delay in clicklike signals presented via headphones and loudspeakers," *J. Audio Eng. Soc.*, 53(7/8), pp. 593–611, Jul./Aug. 2005.

[7] Greenfield, R. and Hawksford, M. J., "The audibility of loudspeaker phase distortion," in *Proc. Audio Eng. Soc. 88th Conv.*, Montreux, Switzerland, Mar. 1990.

[8] Krauss, G. J., "On the audibility of group distortion at low frequencies," in *Proc. Audio Eng. Soc. 88th Conv.*, Montreux, Switzerland, Mar. 1990.

[9] Kulkarni, A., Isabelle, S. K., and Colburn, H. S., "Sensitivity of human subjects to head-related transfer-function phase spectra," *J. Acoust. Soc. Am.*, 105(5), pp. 2821–2840, May 1999.

[10] Raab, D., "Forward and backward masking between acoustic clicks," *J. Acoust. Soc. Am.*, 33(2), pp. 137–139, Feb. 1961.

[11] Olive, S. and Toole, F., "The detection of reflections in typical rooms," in *Proc. Audio Eng. Soc. 85th Conv.*, Los Angeles, CA, USA, Nov. 1988.

[12] Sporer, T., Gbur, U., Herre, J., and Kapust, R., "Evaluating a measurement system," *J. Audio Eng. Soc.*, 43(5), pp. 353–363, May 1995.

[13] Jesteadt, W., Bacon, S., and Lehman, J. R., "Forward masking as a function of frequency, masker level, and signal delay," *J. Acoust. Soc. Am.*, 71(4), pp. 950–962, Apr. 1982.

[14] Fastl, H., "Temporal masking effects: III. Pure tone masker," *Acta Acust. united Ac.*, 43(5), pp. 282–294, Dec. 1979.

[15] Wilson, R. H. and Carhart, R., "Forward and backward masking: Interactions and additivity," *J. Acoust. Soc. Am.*, 49(4), pp. 1254–1263, Apr. 1971.

[16] Mäkivirta, A., Liski, J., and Välimäki, V., "Effect of delay equalization on loudspeaker responses," in *Proc. Audio Eng. Soc. 144th Conv.*, Milan, Italy, May 2018.

[17] Lehtonen, H.-M., "Analysis of piano tones using an inharmonic inverse comb filter," in *Proc. Int. Conf. Digital Audio Effects (DAFX)*, pp. 333–340, Espoo, Finland, Sep. 2008.

[18] Vilkamo, J., "Personal communication with V. Välimäki," Espoo, Finland, Nov. 2012.

[19] Liski, J., Mäkivirta, A., and Välimäki, V., "Companion page to Audibility of loudspeaker group-delay characteristics [Online]," `http://research.spa.aalto.fi/publications/papers/aes144_audibility/`, 2018, [accessed Mar. 1, 2018].