Audibility of Group-Delay Equalization

Juho Liski, Aki Mäkivirta, and Vesa Välimäki, *Fellow, IEEE*

***Abstract*—This paper discusses the audibility of group-delay variations. Previous research has found limits of audibility as a function of frequency for different test signals, but extracting the tolerance for group delay to help audio reproduction system designers is hard. This study considers four critical test signals, three synthetic and one recorded, modified with digital allpass filters. The signals are filtered to produce a positive or negative group-delay peak covering the most sensitive frequency range from 500 Hz to 4 kHz, without changing the delay at other frequencies. ABX listening tests using headphones reveal the audibility thresholds for each signal. The perception is highly dependent on the signal, and the unit impulse and pink impulse are the most critical test signals. Negative group-delay variations are more easily audible than positive ones. The smallest mean threshold for the negative group delay was −0.56 ms and 0.64 ms for the positive group delay, obtained with a pink impulse. The thresholds are smaller than those obtained in previous studies. A synthetic hi-hat sound decaying 60 dB in 80 ms hides a positive group-delay variation. The variation is more difficult to hear in a recorded castanet sound than in the most critical synthetic signals. This work demonstrates how the group-delay response of headphones and loudspeakers can be perceptually tested, and leads to a better understanding of how audio systems should be equalized to avoid audible group-delay distortion.**

***Index Terms*—Acoustic signal processing, audio systems, delay systems, headphones, IIR filters, psychoacoustics.**

I. Introduction

GROUP-DELAY variation and its effect on sound reproduction have been studied widely since the 1970s [1]–[4]. This paper focuses on the audibility of a positive and negative group-delay change in short transient signals, such as clicks and percussive sounds, which have been found to be the most critical test signals for small audio impairments [5], [6].

The main motivation for investigating group-delay distortion, i.e., the variation of group delay from a constant value, is to understand its audibility and how much it can affect audio quality in loudspeakers and headphones [6]–[9] and in a complete system including loudspeakers and a listening room [10]. In well-designed loudspeakers, the group-delay variations can be designed to be small, but when loudspeakers are used in rooms, the group-delay variations can become substantial, up to several tens of milliseconds. The knowledge on the group-delay audibility is also useful for signal-processing techniques based on imperceptible allpass filtering, such as in data-hiding in an audio signal [11] and in linear audio compression, which reduces the peak signal value with phase processing [12].

The magnitude response of a loudspeaker is typically flattened with an equalizer [13]–[16]. In a multi-way loudspeaker, the equalizer can be applied separately after the crossover filtering for each transducer individually [15], [17], [18], applied before the crossover filter, or combined with the crossover filters [19]. The commonly used Linkwitz-Riley crossover filter is a minimum-phase filter [15], [20] introducing phase distortion, although it shows the same delay in all outputs. The phase distortion is often described in terms of the group delay, i.e., the negative first derivative of the phase response [21]. A digital allpass filter does not alter the magnitude response and is suitable for equalizing the phase or group-delay response independently of the magnitude response [22]–[25].

In addition to equalizing the phase around the crossover frequencies [26], [27], phase equalization is applied at low frequencies [14], [28]. Both Adam and Benz [26] and Herzog and Hilsamer [28] used time-reversed allpass filters to apply group-delay equalization. Similarly, the time-reversed version of the windowed loudspeaker impulse response can be utilized as a finite impulse response (FIR) filter to flatten the group delay [7]. This, however, doubles the ripple in the magnitude response and also affects the system latency.

The audibility thresholds for group-delay variation from several previous related studies are shown in Fig. 1. If not otherwise stated, these studies have been conducted using headphones. Green [3] applied Huffman sequences, or truncated impulse responses of second-order allpass filters, to study the audibility of phase distortion. He found a threshold value for the peak group delay of about 2 ms for center frequencies of 625 Hz, 1875 Hz, and 4062 Hz.

Later, also rectangular pulses were used in listening tests, not only impulse responses. Jensen and Møller compared plain 50-μs square pulses to ones filtered with second-order allpass filters with center frequencies of 250 Hz, 500 Hz, 800 Hz, 1200 Hz, and 2000 Hz [29]. This led to group-delay thresholds of 4.7 ms, 2.3 ms, 1.5 ms, 1.6 ms, and 1.6 ms, respectively. Blauert and Laws performed a similar listening test to [29], but used a 25-μs-wide rectangular pulse as the test signal and filtered them with allpass filters with the center frequencies of group-delay peaks at 0.5 kHz, 1 kHz, 2 kHz, 4 kHz, and 8 kHz [4]. The resulting

Manuscript received September 2, 2020; revised January 24, 2021; accepted May 24, 2021. Date of publication June 9, 2021; date of current version June 25, 2021. This work was supported in part by IMPRESS2 project funded by Genelec (Impulse-Response Shortening—Part 2, Aalto University project no. 400469) and in part by Nordic Sound and Music Computing Network—NordicSMC, NordForsk project no. 86892. The work of J. Liski was supported by Aalto ELEC Doctoral School. *(Corresponding author: Juho Liski.)*

Juho Liski and Vesa Välimäki are with the School of Electrical Engineering, Department of Signal Processing and Acoustics, Acoustics Lab, Aalto University, FI-02150 Espoo, Finland (e-mail: juho.liski@aalto.fi; vesa.valimaki@aalto.fi).

Aki Mäkivirta is with Genelec Oy, FI-74100 Iisalmi, Finland (e-mail: aki.makivirta@genelec.com).

Digital Object Identifier 10.1109/TASLP.2021.3087969

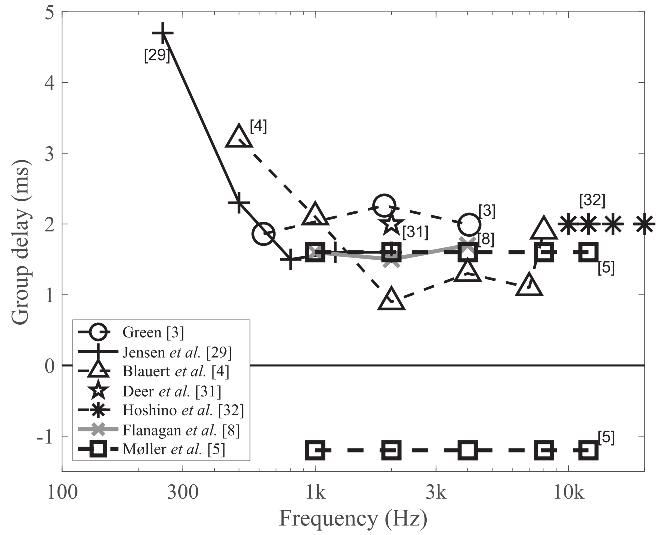

Fig. 1. Audibility thresholds for group delay from previous studies.

thresholds were 3.2 ms, 2.1 ms, 0.9 ms, 1.3 ms, and 1.9 ms, respectively [4]. Training was shown to lower these values [4].

The minimum integration time of the auditory system has been shown to be about 2 ms [30], which is of a magnitude similar to many reported group-delay audibility thresholds: Deer *et al.* reported a threshold value of 2 ms at 2 kHz using second-order allpass filters and clicks [31], whereas Hoshino and Takegahara were interested in the audible effects of high-frequency anti-aliasing filters and obtained a threshold value of 2 ms when comparing square-wave pulse trains with group-delay distortion at 10 kHz, 12 kHz, 15 kHz, and 20 kHz and the corresponding linear-phase version [32].

Minnaar *et al.* employed a listening test where the Q factor of a second-order allpass section was adjusted to find the threshold of audibility when a test signal and its filtered version were compared [5], [33], [34]. They applied both causal and non-causal filtering (using time-reversal) to impulses, and obtained constant group-delay thresholds for center frequencies of 1 kHz, 2 kHz, 4 kHz, 8 kHz, and 12 kHz: +1.5 ms (+1.6 ms was reported in [5]) and -1.2 ms for the causal and non-causal cases, respectively. The latter values are included in Fig. 1. They noted that the group-delay peak can predict the audibility, but actually the decay of the sinusoidal component in the impulse response determines the perceived difference [34].

Flanagan *et al.* performed an extensive test on group-delay audibility using both headphones and loudspeakers [8]. The results of the listening test with a lowpass filtered 20-μs pulse was compared to the results of the same signal filtered with a second-order allpass filter having the center frequency of 1 kHz, 2 kHz, or 4 kHz and a peak group delay of 0.5 ms, 1.0 ms, 2.0 ms, 4.0 ms, 8.0 ms, or 16.0 ms. The obtained threshold value was about +1.6 ms for all frequencies, also shown in Fig. 1, when using headphones, and only slightly higher thresholds were obtained with loudspeakers in a low-reverberant room.

Various researchers have published additional observations regarding the origin [7], [35] and audibility of group-delay distortion without discussing audibility thresholds [1], [36]. A big motivation for such test has been loudspeakers and their phase properties: the audibility of actual or simulated loudspeaker phase has been studied in [37]–[39] and loudspeaker equalization and its effects on the phase audibility have been studied in [14], [26], [28], [40]. Finally, even though group-delay distortion is often reported to be more easily audible with headphones than with loudspeakers [5], [8], [36], [37], contradicting results have also been reported [41], [42]. Bech's results showed no significant difference between the two reproduction methods [39].

The aim of this study is to determine audibility thresholds for group-delay variations. This is related to loudspeaker equalization in a range of frequencies previously known to be the most sensitive for hearing variations in the system group delay. Importantly, this also means that negative group-delay peaks must be studied. This case has been investigated in only one previous study using a single test signal [5]. The present study includes several test signals, mostly synthetic and also a recorded castanet signal.

As in several previous studies, group-delay variations are produced using allpass filters. As a novelty, this work avoids group-delay fluctuations at low frequencies by forming group-delay peaks with a combination of two allpass filters, one of which is applied backward in time. This is accomplished using time reversal, which is possible when test signals are processed offline. Extra care has been taken in setting up the listening test to ensure that the sample presentation method does not contribute to the test results.

Since negative group-delay peaks occur in phase-equalized loudspeakers, their audibility is considered a relevant research problem here. Positive group-delay peaks appear in loudspeaker responses due to their low- and high-frequency roll-off characteristics and the crossover filters [27], but negative peaks can arise when applying group-delay equalization. For instance, when processing the loudspeaker impulse response in order to obtain an approximately linear-phase, symmetrical impulse response (within the boundaries of measurement accuracy and filter order), some of the time-domain energy can be moved ahead of the main peak of the impulse response. This effect can be modeled by applying a time-reversed allpass filter to a signal, causing the energy of a narrow frequency band to occur earlier in time.

In this work, the audibility of the group-delay variations are studied in headphone listening using an ABX test [43] (two-alternatives forced-choice test), which is suitable for comparing audio signals with very small differences. This study has been designed carefully to minimize the effect of headphones or other external factors. The headphones used in this study have been verified to show sufficiently small group-delay variations to not be a factor in the threshold evaluation. The results obtained in the listening test are different for positive and negative group-delay peaks and for each test signal, and are believed to be relevant for loudspeaker design and testing.

This paper is organized in the following way. Sec. II introduces the test signals used in this study and explains how they are processed using a pair of allpass filters. Sec. III presents the design of the listening test, including an analysis of the

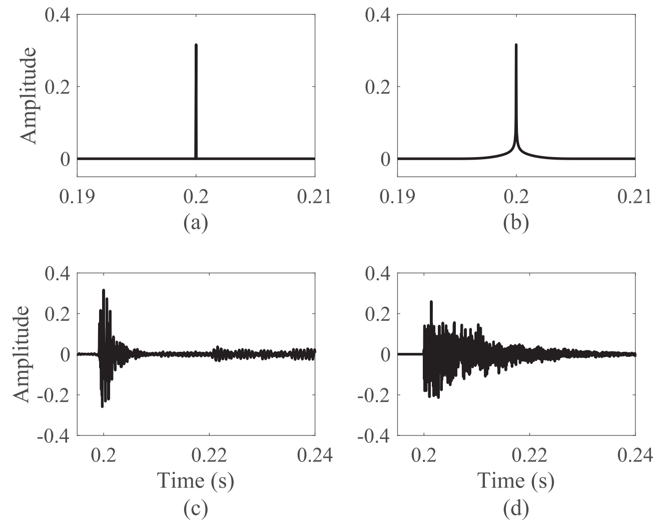

Fig. 2. Waveforms of test signals: (a) unit impulse, (b) pink impulse, (c) castanet (first strike), and (d) synthetic hi-hat.

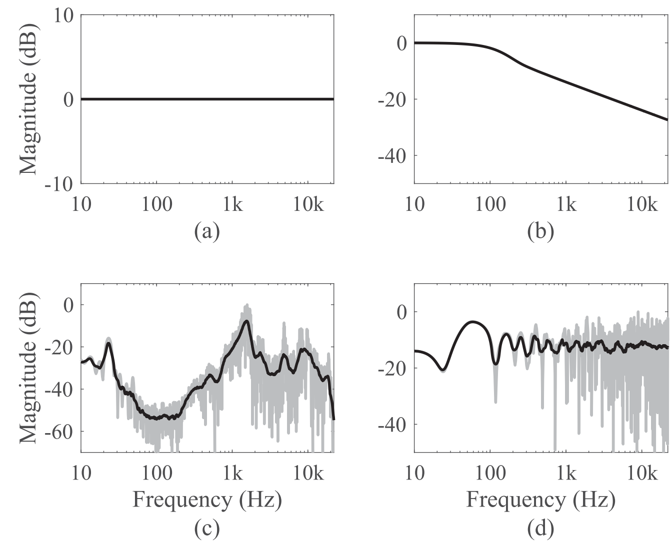

Fig. 3. Magnitude spectra of the test signals of Fig. 2: (a) unit impulse, (b) pink impulse, (c) castanet (first strike), and (d) synthetic hi-hat. In (c) and (d), the gray line is the spectrum and the black line is its third-octave smoothed version.

headphones used. Sec. IV analyzes the data obtained in the listening test and derives the main results of this work. Sec. V discusses our results and compares them with those of previous studies, and Sec. VI concludes the paper. The appendix derives the impulse responses of first-order and second-order allpass filters.

II. Test Signals

This section discusses the choice of audio test signals used in this study and proposes signal processing techniques to modify their group delay.

A. *Selection and Synthesis of Test Signals*

Both synthetic and recorded signals were used in order to achieve the highest sensitivity and to connect the test to real-world signals. All test signals were produced, stored, and played back at the sample rate of $f_s = 44\,100$ Hz. No sample-rate conversions were involved. The complete listening-test apparatus used a linear-phase reconstruction filtering. The contribution of the system impulse response was studied, and we concluded that the apparatus did not affect the test results.

One of the synthetic signals is the unit impulse, which is shown in Fig. 2(a). It was chosen since it has been used in previous studies (see, e.g., [5]), and due to its minimal temporal length it is sometimes thought to be the most critical test signal in audio. When the unit impulse is filtered with the allpass filter, the result is purely the impulse response of that filter, which is the shortest possible signal having the desired group-delay characteristics. The filtered ideal unit impulse has a flat magnitude response across the audible frequency range, as shown in Fig. 3(a), which suits this listening test, in which various center frequencies are of interest. The drawback of the unit impulse is its low total energy, which may result in a soft sound. However, this did not limit its use in this listening test.

The second synthetic sound was a pink impulse [6]. Due to the flat spectrum of the unit impulse, it sounds bright. To emphasize the low and middle frequencies, which dominate in music audio, a pink impulse has been specified to have the same spectrum as pink noise [6]. It was observed to improve the audibility of group-delay differences over the unit impulse in our previous study [6]. The spectrum of the pink impulse is $H(\omega) = 1/\sqrt{\omega}$, where $\omega = 2\pi f$ is the angular frequency (rad) and f is the frequency (Hz). Since the spectrum $H(\omega)$ is real-valued, it corresponds to a zero-phase symmetric pink impulse.

The pink impulse $H(\omega)$ was synthesized using the frequency sampling method at 16 384 frequency points containing both positive and negative frequencies [6]. Next, this $H(\omega)$ was inverse discrete Fourier transformed, and the two halves of the result were swapped to have the peak in the middle of the signal buffer. Finally, 440 samples (10 ms at the sample rate of 44 100 Hz) from the middle of the buffer were truncated with the Blackman window. This resulted in the short symmetric pulse shown in Fig. 2(b). Its spectrum, shown in Fig. 3(b), decays about -3 dB per octave everywhere except at frequencies below 100 Hz, where it has been flattened by truncation, so as not to excessively stress the headphones.

The other two test signals, one recorded and the other synthesized, were similar to percussive sounds commonly appearing in music, and thus make the results applicable to typical audio program material. The recorded signal was a castanet rhythm from the SQAM compact disc [44]. The waveform of the first strike and its magnitude spectrum are shown in Figs. 2(c) and 3(c), respectively. The recording contains room reverberation, and its energy is concentrated in a narrow frequency range in the vicinity of 2 kHz. However, since the castanets are a real sound source, the test subjects are familiar with it, and this links the test better to their everyday life than the synthetic signals.

Furthermore, due to the drawbacks of the castanet recording, such as high reverberation, one more test signal was synthesized. In order to obtain another signal resembling real-world musical sounds, a closed hi-hat cymbal sound was produced.

The synthesized hi-hat signal, shown in Figs. 2(d) and 3(d), was obtained by creating a 0.5-s burst of white Gaussian noise and multiplying it with an exponentially decaying envelope. Thus, a relatively flat magnitude spectrum is obtained. The duration of the hi-hat signal is longer than the other signals. The pole of the leaky integrator in the envelope generator is real, located at 0.998, which corresponds to a time constant of 11 ms (i.e., 500 samples). This results in a 60-dB decay in approximately 80 ms.

B. Generating Positive and Negative Group Delays

The test signals were filtered to introduce group-delay peaks of different magnitude at various center frequencies. The chosen nominal center frequencies were 500 Hz, 1 kHz, 2 kHz, 3 kHz, and 4 kHz, which cover the most sensitive frequency region of hearing. Concurrently, we did not want to cause changes in the magnitude spectrum of the signals. So, a second-order allpass filter was chosen for this purpose, due to its bell-like group-delay curve and flat magnitude response [25], [31], [45].

A second-order allpass filter section has the following transfer function:

$$A_2(z) = \frac{a_2 + a_1 z^{-1} + z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{1}$$

where the filter coefficients a_1 and a_2 are

$$a_1 = -2R\cos\phi \text{ and } a_2 = R^2, \tag{2}$$

respectively, R is the magnitude of complex conjugate poles, and $\pm\phi$ are the angles of the poles. These parameters are linked to the two main design parameters, center frequency f_c in Hz and peak group-delay value T in ms, by relations [8]

$$\phi = 2\pi f_\mathrm{c}/f_\mathrm{s} \tag{3}$$

and

$$R = \frac{\tau - 1}{\tau + 1}, \tag{4}$$

with

$$\tau = \frac{T f_\mathrm{s}}{1000}, \tag{5}$$

where τ is the peak group delay in samples at the sample rate of $f_\mathrm{s} = 44100$ Hz.

The bandwidth of a group-delay peak is defined as the frequency range between points where the group delay is 50% of the peak value [8]. This definition is used for all group-delay peaks in this work. The analytical bandwidth of the group-delay peak of a second-order section in Hz is [8]

$$B_2 = 2\cos^{-1}\left(\frac{R^2 - 4R + 1}{-2R}\right) = 2\cos^{-1}\left(\frac{\tau^2 - 3}{\tau^2 - 1}\right), \tag{6}$$

where (4) has been applied to reach this final form.

However, in addition to the group-delay peak, the second-order allpass filter causes some group delay to occur at low frequencies. An example is shown in Fig. 4(a), where a group-delay peak of 0.62 ms is produced at 980 Hz, but the group delay is also increased at lower frequencies so that it is approximately 0.27 ms at 10 Hz. In order to focus the listening test only on the narrow group-delay peak, we suggest using a first-order allpass filter backwards to cancel the delay at low frequencies, as shown in Fig. 4(b). This is possible since the test signals are prepared offline. This principle is shown in Fig. 5(a). The idea of reversing the signal for digital filtering has been proposed in several studies [5], [26], [28], but here it is applied for a different purpose: to shape the group delay at low frequencies.

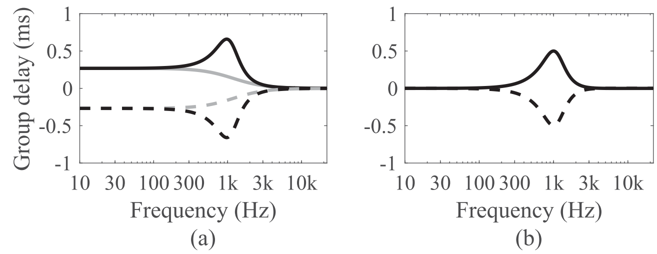

Fig. 4. (a) Second-order (black lines) and first-order (gray) allpass component filters used to create the (b) positive (solid lines) and negative (dashed) group-delay peak.

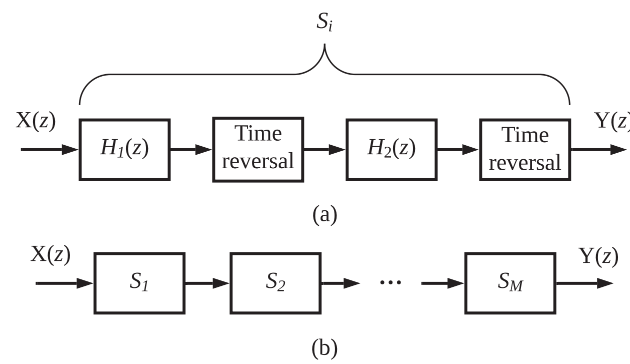

Fig. 5. (a) Block diagram showing allpass filtering forward and backward, where filter $H_1(z)$ is applied forward and filter $H_2(z)$ processes the input signal backward in time, and (b) the structure for creating group-delay peaks greater than 0.5 ms by cascading M blocks of S_i from (a).

The total group delay is generated by applying the first-order allpass filter and the second-order filter having the same group delay at low frequencies. A positive group-delay peak is obtained by running the second-order allpass filter forward and the first-order filter backward, whereas a negative peak is obtained with the second-order allpass filter backward and the first-order filter forward, as shown in Fig. 4.

The first-order filter must have the same group-delay behavior at low frequencies as the second-order one. Thus, we want to find a formula for the single parameter d in a first-order allpass filter of the form

$$A_1(z) = \frac{-d + z^{-1}}{1 - dz^{-1}}, \tag{7}$$

such that the pole d is determined by the design parameters of the second-order allpass filter R and ϕ. The formula for the first-order allpass filter coefficient d can be solved using analytic group-delay equations.

The group-delay function $\tau_2(\omega)$ of the second-order allpass filter, in samples, can be written as [21]

$$\tau_2(\omega) = 2 + \frac{2R\cos(\omega-\phi) - 2R^2}{1+R^2-2R\cos(\omega-\phi)} + \frac{2R\cos(\omega+\phi) - 2R^2}{1+R^2-2R\cos(\omega+\phi)}. \tag{8}$$

The group delay $\tau_1(\omega)$ of a first-order filter, on the other hand, is given by [21]

$$\tau_1(\omega) = \frac{1-d^2}{1+d^2-2\,d\cos(\omega-\phi_1)}, \tag{9}$$

where $\phi_1 = 0$, since it is the angle of the pole. In order to match the two group delays at low frequencies, we set them equal at 0 Hz:

$$\tau_2(\omega)|_{\omega=0} = \tau_1(\omega)|_{\omega=0}. \tag{10}$$

A small amount of algebra and the help of the quadratic formula give two possible solutions for d:

$$d = \frac{\nu+1}{\nu+1} = 1 \text{ or } d = \frac{\nu-1}{\nu+1}, \tag{11}$$

where ν is a temporary variable

$$\nu = \frac{2-2R^2}{1+R^2-2R\cos(\phi)}. \tag{12}$$

The former solution in (11) is a constant and is therefore discarded, leaving us with a single solution that can be expanded to obtain the equation for d:

$$d = \frac{4(R^2-1)}{R^2+2R\cos(\phi)-3} - 1. \tag{13}$$

Note that, when (7) and (13) are used to design the first-order allpass filters, the resulting magnitude responses are perfectly flat without any ripple, similarly to the second-order allpass filters used in this work.

Equations (7) and (13) are used to design the first-order allpass filters in Fig. 4(a) corresponding to the second-order filters in the same subfigure (both positive and negative peaks are shown); when $f_c = 980$ Hz and $\tau = \pm 27.3$ samples (or $T = \pm 0.62$ ms), we obtain $d = 0.8439$ and $d = 1.1850$ for the positive and negative peaks, respectively, and the match is perfect apart from the peak in both cases. Finally, determining the difference of the group delays of the two filters results in a single 0.5-ms group-delay peak, either a positive or a negative one, shown in Fig. 4(b), as desired. The obtained group-delay peaks are not exactly symmetric on the logarithmic frequency axis, but they satisfy the needs of this study, as the asymmetry is small. The same procedure is utilized for the other center frequencies using the f_c, T, and d values shown in Table I.

The first-order allpass filter affects the center frequency of the group-delay peak, which results in the actual second-order filter center frequencies differing from the nominal center frequencies of 500 Hz, 1 kHz, 2 kHz, 3 kHz, and 4 kHz. Iterating a correct, slightly modified center frequency for the second-order allpass section is quite easy, when a certain final center frequency is targeted. The modified center frequencies f_c of the second-order sections are shown in Table I. Also the bandwidth of the group-delay peak is affected by the second filtering operation. The original bandwidths of the second-order filters B_2 and the bandwidths of the twice-filtered final allpass filters B are given in Table I. The difference in the bandwidths is also shown in Fig. 4, as the second-order filter in (a) has a group-delay bandwidth of 1030 Hz whereas the group-delay peak in (b) has a bandwidth of 856 Hz.

TABLE I
ALLPASS FILTER AND GROUP-DELAY PEAK PARAMETERS FOR 0.5 MS. COPIES OF EACH ALLPASS FILTER PAIR ARE CASCADED TO PRODUCE LARGER GROUP-DELAY PEAKS

$A_2(z)$			$A_1(z)$	Filter pair	
f_c (Hz)	T (ms)	B_2 (Hz)	d	f_c (Hz)	B (Hz)
421.6	0.70	915	0.9415	500	702
980.1	0.62	1030	0.8439	1000	856
1997.5	0.56	1147	0.5792	2000	1040
2999.8	0.53	1198	0.2921	3000	1129
4000.2	0.52	1224	0.0376	4000	1177

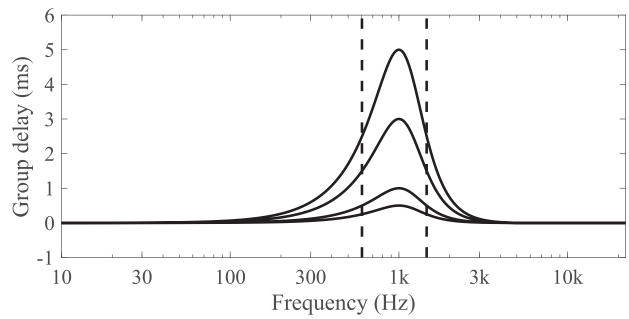

Fig. 6. A set of group-delay functions produced using the allpass filtering techniques of Fig. 5 to have the peak values of (bottom to top) $T = 0.5$ ms, 1.0 ms, 3.0 ms, and 5.0 ms at $f_c = 1$ kHz. The bandwidth of these group-delay peaks is the same (856 Hz), since they are all scaled from the 0.5-ms curve. The dashed lines indicate the 50-% bandwidth limits.

The proposed allpass filtering technique shown in Fig. 5(a) can be applied several times to produce group-delay peaks of multiples of ± 0.5 ms. This is shown in Fig. 5(b). For example, a group-delay peak of 1 ms is obtained by cascading the operations of Fig. 5(a) two times ($M = 2$). Similarly, a group-delay peak of 5.0 ms is obtained by cascading $M = 10$ such processes, since each of them produces a 0.5-ms peak. By cascading multiple structures of Fig. 5(a) instead of designing a new allpass-filter pair for each peak value, the bandwidth remains the same for each filter structure with the same center frequency. This is shown in Fig. 6.

C. Allpass Filtering to Produce Small Group Delays

Unfortunately, for smaller group-delay values, the peak produced with a second- and a first-order allpass filter becomes very wide. In order to produce signals where the processing is very hard or impossible to hear, we wanted to create group-delay peaks of ± 0.25 ms.

TABLE II
ALLPASS FILTER PARAMETERS FOR GROUP-DELAY PEAKS EQUAL TO 0.25 MS

Forward $A_2(z)$		Reverse $A_2(z)$		Filter pair	
f_c (Hz)	T (ms)	f_c (Hz)	T (ms)	f_c (Hz)	B (Hz)
519.6	1.25	536.6	1.03	500	264
966.4	0.79	923.5	0.60	1000	433
1947.1	0.50	1772.7	0.32	2000	748
2952.5	0.41	2642.4	0.22	3000	982
3959.4	0.37	3510.9	0.17	4000	1156

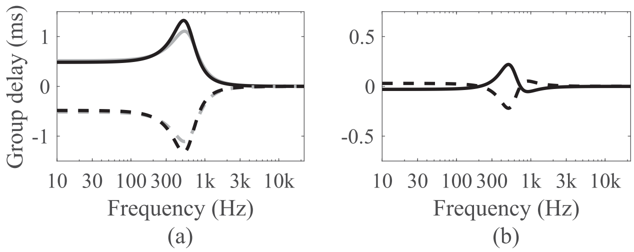

Fig. 7. (a) Two second-order allpass component filters (black and gray lines) used to create the (b) small positive (solid lines) and negative (dashed) group-delay peak of ± 0.25 ms.

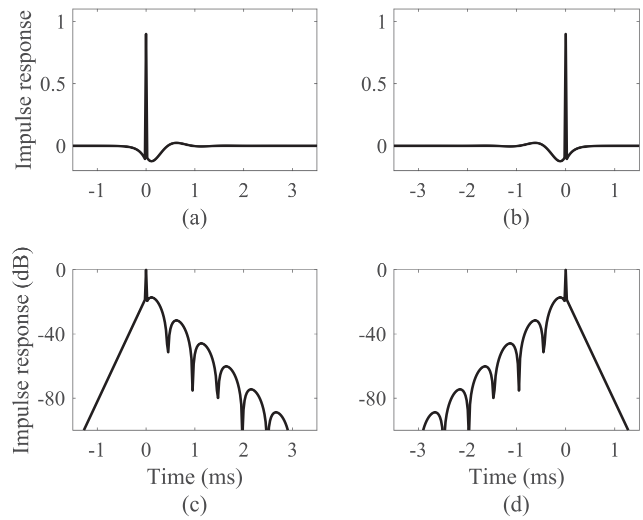

Fig. 8. Impulse responses of the allpass filters in Fig. 4(b) producing (a) a 0.5-ms (b) a -0.5-ms group-delay peak at 1 kHz and (c), (d) the same impulse responses on the dB scale.

Different filters are needed to produce a smaller group-delay peak. The forward filter is still a second-order allpass section, and the values of f_c and T are shown in Table II for the different center frequencies. However, another second-order allpass section is also used for the time-reversed filtering instead of the first-order one, as shown in Fig. 7(a). The center frequencies f_c and the values of the parameter T of the additional second-order filters are given in Table II. This leads to a single group-delay peak with the desired maximum value, as shown in Fig. 7(b). The additional second-order filters also affect the center frequency and bandwidth of the resulting peaks, which are presented in Table II.

D. Allpass Filtered Test Signals

All the desired group-delay peaks were produced with one of the two allpass filtering methods. The maximum values of the peaks are ± 0.25 ms, ± 0.5 ms, ± 1.0 ms, and ± 2.0 ms for the two impulse-like signals (unit and pink impulses) and -5.0 ms, -2.0 ms, -1.0 ms, 0.5 ms, 1.0 ms, 3.0 ms, and 5.0 ms for the castanet and synthetic hi-hat sounds. All the test signals used in this work are available online [46].

Finally, we consider the impulse responses of the allpass filters. Appendix A shows how these are derived from the transfer function coefficients, and here we discuss the responses of the filters in Fig. 4(b) as an example for both the positive and negative peak case. Figure 8 shows the impulse responses both on a linear and on a logarithmic scale: Figs. 8(a) and (c) correspond to the positive group-delay peak and Figs. 8(b) and (d) to the negative one. The impulse responses (23) and (24), which correspond to the second- and first-order allpass sections, respectively (see Appendix A), are combined using convolution.

The impulse response of Fig. 8(a) corresponding to the positive group-delay peak is obtained by convolving (23) with the time-reversed version of (24). However, the impulse response in Fig. 8(b), corresponding to the negative group-delay peak, is obtained by time-reversing the impulse response of the second-order section (23) and convolving it with (24), the impulse response of the first-order section. As is seen in Figs. 8(c) and (d), the example impulse response decays approximately 100 dB within approximately 5 ms in both directions in time. The actual test signals contain 8820 zeros before the signal begins and 13 230 zeros after the signal, or 200 ms and 300 ms, respectively, to ensure that the allpass filter responses are not cut during the processing.

As is seen in Fig. 8, the impulse responses for the positive and negative peak are mirror images of each other on the time axis. This is natural, since the same filters are used, but in opposite direction in time. One also sees that both impulse responses in Figs. 8(a) and (c) contain energy before the main impulse. Likewise, the impulse responses in Figs. 8(b) and (d) contain energy after the main impulse. This is due to one of the filters being applied backwards. Thus, in order to only obtain a group-delay peak without the excess group delay at low frequencies, a small compromise is required regarding the temporal behavior of the allpass filter.

III. LISTENING TEST

The unprocessed and processed test signals described in Sec. II were compared in a formal blind listening test. The test comprised four different test signals, five center frequencies, and eight or seven group-delay peak-values for the impulse-like signals and the real-life signals, respectively, resulting in 150 different signals. In addition, two extra repetitions for an anchor signal and four training signals were used for the test subjects to familiarize themselves with the distinct test-signal types and the user interface of the test. All in all, the test included 156 signals for each test subject.

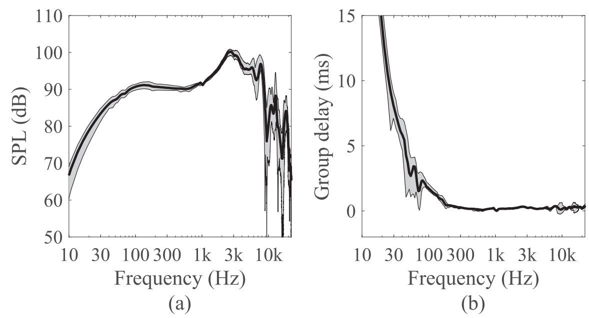

Fig. 9. Variability of (a) the magnitude response and (b) the group delay of the left and right earpieces of three pairs of Sennheiser HD-650 headphones used in the listening tests. The thick black curves are the average responses whereas the thin gray lines are the minimum and maximum responses showing the deviations between headphones.

The test subjects were all employees of Aalto University with normal hearing. All except one had prior experience of formal listening tests. The age of the test subjects ranged from 24 to 39 years, with the mean of 29.9 years. The test was conducted in sound-proof listening booths at the Aalto Acoustics Lab, Espoo, Finland. The test software was run on Mac Minis manufactured in 2014 and running MacOS 10.14.6, and the headphones were attached to the Objective2 + ODAC headphone amplifier/digital-to-analog converter combo.

The test was performed using Sennheiser HD-650 headphones, which are widely used in subjective testing. Three pairs of headphones were used. The magnitude responses and group delays of both the left and right earpiece of all headphones were measured beforehand using a dummy head with ear canal simulators to verify that their responses did not differ much from each other. The headphones were connected to the complete test setup during the measurements. Figure 9 shows the average and variability of all the earpiece responses. To account for the slight variation in the earpiece sensitivity, the magnitude responses in Fig. 9(a) are normalized at 1 kHz.

Figure 9 indicates that the magnitude response and the group delay of the headphones were in good agreement in the frequencies between 500 Hz and 4 kHz, the range of interest in the present study. The magnitude responses of the three headphone pairs varied by approximately ± 0.75 dB between 500 Hz and 3 kHz and by less than ± 1.5 dB between 3 kHz and 4 kHz. The group delay varied by ± 0.025 ms in the range from 1.5 kHz to 4 kHz and by less than ± 0.1 ms below 1.5 kHz. In addition, the left-right channel balance was tested and was found to be good: within the frequency range under test, the level difference was within 1.27 dB, 2.23 dB, and 2.30 dB for the three pairs of headphones. Thus, the headphones suited the test well and are not likely to have produced additional audible effects to the test signals, making the listening tests performed with the different pairs comparable.

Previous studies indicate that the audibility of phase distortion is largely unaffected by the reproduction level [5] as long as a certain minimum level is exceeded [3], [8]. Generally, the studies whose results are shown in Fig. 1 report a sound pressure level (SPL) of 80–90 dB SPL [3], [8], [32] or then talk about a "comfortable listening level" without further details [5], [31]. These listening levels can be assumed to be close to the other studies since Flanagan *et al.* [8] report that a peak level of 90 dB SPL results in a comfortable listening level for impulsive sounds.

We reproduced all test signals at a level of 80 dB SPL except the unprocessed and processed unit impulses, where the level was 85 dB SPL. This was due to the small amount of energy contained in the unit impulses. A subjective pilot test confirmed that the processing did not produce perceivable loudness differences between the reference and the processed sounds that could serve as an unwanted clue to detect a signal. No further loudness compensation was needed.

The listening test was conducted in accordance with the ABX method and was implemented using webMUSHRA [47]. On each trial, the subjects could listen to three versions of a given audio material. These were labelled "Reference," "A," and "B". The reference was always the original test signal against which both "A" and "B" were compared. Either item "A" or "B" was a processed signal, and the other was the hidden reference, which was identical to the reference signal. The processed signal differed from the original signal only in its group delay. The magnitude spectrum and the interval of test signals were identical. The task of the subject was to recognize which item was the hidden reference by selecting "A" or "B". The subjects could play all three test sounds as many times as they wanted and could freely switch between them.

The motivation behind the listening test was to obtain data to fit a psychometric function corresponding to the audibility threshold of the group-delay peaks at different center frequencies. The psychometric function relates a subject's answers to a test variable describing the probability of correct answers in a detection or discrimination task as a function of the stimulus intensity [48]–[50]. It can typically be estimated with a sigmoid function [48], [49], [51], and the detection threshold is then taken as the variable value at which a set probability is reached. In this work, the 75-% threshold point is used, just like in Flanagan *et al.* [8]. The 75-% point refers to half of the answers being correct; 50% refers to a guessing probability.

Thirteen subjects participated in the listening test, but one had to be excluded. The test included a particularly easy-to-hear low anchor signal three times (pink impulse with a -2-ms group-delay peak at 1 kHz), which all subjects had to detect correctly at least two times out of three to show that they can hear the differences and are concentrating on the task. The discarded subject responded wrongly two out of three times. Eight subjects performed the test twice with a three-week interval between tests. The long period between the repeated test, and the fact that all the test deliveries were randomized, ensured that the second test was independent of the first. There is no evidence of a training effect, apart from a faster execution of the test on the second time. A total of 20 responses per test condition were gathered when all responses across subjects were pooled.

IV. Results

This section presents the analysis of the listening test data and the results from the listening test.

A. Data Analysis and Confidence Intervals

The ABX listening test is a Bernoulli trial [52], i.e., an experiment with exactly two outcomes: "success" or "failure". The binomial proportion confidence interval [53] can be used to estimate the interval of a success probability, and only the number of experiments and number of successes are required for this. The Wilson score interval [54], which performs well with a small number of trials [53], is used in this work. The Wilson score interval gives the success probability interval p_{CI}, which is formed of two parts p_{A} and p_{B}, as [53]

$$p_{\mathrm{CI}} = \frac{n_{\mathrm{S}} + \frac{Z^2}{2}}{n + Z^2} \pm \frac{Z}{n + Z^2}\sqrt{\frac{n_{\mathrm{S}} n_{\mathrm{F}}}{n} + \frac{Z^2}{4}} = p_{\mathrm{A}} \pm p_{\mathrm{B}}, \quad (14)$$

where n is the total number of trials, n_{S} and n_{F} are the number of successes and failures, respectively, and Z is the Z-value of a standard normal distribution. In this work, we use the 68.3% confidence level (one standard deviation, $Z = 1$).

After having estimated the success probability and the corresponding confidence interval, these are grouped into p:

$$p = \{p_{\mathrm{A}} + p_{\mathrm{B}}, p_{\mathrm{A}}, p_{\mathrm{A}} - p_{\mathrm{B}}\}. \quad (15)$$

Next, the sigmoid functions were fitted to the data using the Matlab Curve Fitting Toolbox and a custom equation. The sigmoid function is expressed as

$$\sigma(x) = \min(p) + \frac{\max(p) - \min(p)}{1 + e^{-b(x-a)}}, \quad (16)$$

where the min() and max() functions set the range for the sigmoid, and parameters a and b set the half-range point and the slope of the sigmoid, respectively, using nonlinear least-squares fitting [55]. The min() and max() functions in (16) are obtained from (15), and thus we have three values for each function corresponding to the two confidence interval limits and the perceived responses from the test subjects. The largest $\max(p) \leq 1$ is obtained for the upper confidence interval limit, and the smallest $\min(p)$, which is obtained for the lower confidence interval limit, can be less than 0.5.

The following examples show how (14), (15), and (16) are used to derive the audibility thresholds and the corresponding confidence intervals. Two extra points are first added to augment the data for sigmoid fitting. One point is placed at a very low group-delay value and at the 50% level. Another point is placed at a very high group-delay value and at the 100% level. This is done so as to make the extremes of the sigmoid converge to 50% (no detection, random response) and to 100% (confident detection), respectively. This helps primarily the cases of the castanet and synthetic hi-hat sounds, particularly the negative group-delay cases. Without data augmentation, the mathematical sigmoid fit would not be constrained in a sensible way even if the results indicated that the listeners are responding to audible effects. Because of this need, the augmentation was added to all cases. We use the values ±0.01 ms and ±10 ms for the impulse-like sounds and ±0.01 ms and ±15 ms for the other two test sounds, for the certainly non-detectable and detectable cases, respectively.

After this pre-processing step, (14) is used to obtain three sets of data for each test case: one set to obtain the best fit to the observed answers, the other two to obtain the lower and upper confidence limits for the best fit. Here, a test case is defined as the percentage of correct answers for a single test signal and for either positive or negative group-delay peaks.

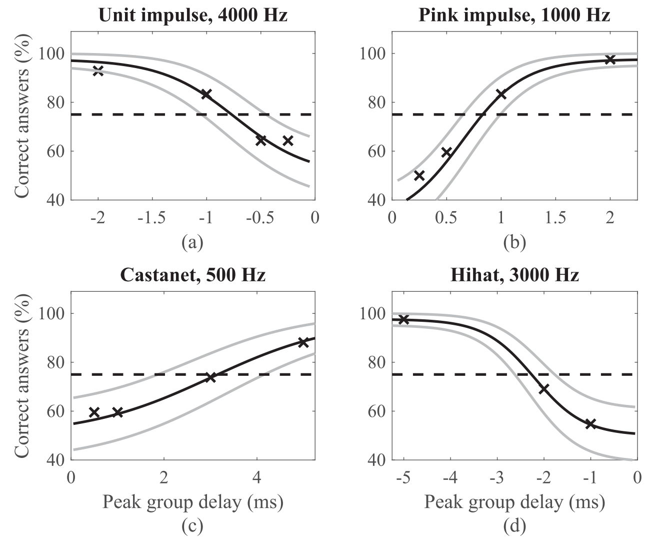

Fig. 10. Four examples of the sigmoid fitting with both positive and negative group-delay peaks. Crosses are the observed data points, black curves are the sigmoids fitted to these, and the gray curves are the upper and lower confidence intervals, respectively. The threshold limit of 75% is marked with the dashed line.

The sigmoids are then fitted to each data set (best fit, upper confidence limit, lower confidence limit) for each test case using (16). Some curve fits required manual weighting in order to converge to sigmoid shapes due to the distribution of the responses. Examples of the resulting sigmoid curve fits are shown in Fig. 10. They illustrate the range of the sigmoid shapes obtained using the data. The three sigmoid curves (the best fit and the two confidence limits) per case are presented in Fig. 10. The audibility thresholds and the confidence limits associated with the threshold are obtained at the 75% level in the sigmoid fits. The equation for the threshold group-delay value T_{th} (in ms) is derived from (16):

$$T_{\mathrm{th}} = -\frac{\ln\left(\frac{\max(p)-\min(p)}{0.75-\min(p)} - 1\right)}{b} + a, \quad (17)$$

where $\ln()$ is the natural logarithm. In this way we obtain the best fit detection threshold values with relevant estimates of the confidence interval for each value for each test signal and at each center frequency for both positive and negative group-delay peaks.

B. Audibility Thresholds of the Group Delay

The audibility thresholds obtained as described in the previous section are shown in Fig. 11 and are given as numerical values in Table III together with their confidence intervals for all signals and for positive and negative group-delay peaks. The impulse-like sounds (unit impulse and pink impulse) result in the lowest detection threshold values, whereas the two musical signals (castanet and synthesized hi-hat) require larger group

TABLE III
GROUP-DELAY AUDIBILITY THRESHOLDS AND CONFIDENCE-INTERVAL LIMITS IN MILLISECONDS

f_c Signal	500 Hz	1 kHz	2 kHz	3 kHz	4 kHz	Mean
Unit impulse	−0.84 [−0.44, −1.13] 0.96 [0.83, 1.08]	−0.51 [−0.44, −0.64] 0.77 [0.57, 0.98]	−0.68 [−0.50, −0.85] 0.69 [0.38, 0.94]	−0.65 [−0.46, −0.78] 0.64 [0.51, 0.76]	−0.77 [−0.45, −1.05] 0.75 [0.42, 1.00]	−0.69 0.76
Pink impulse	−0.45 [−0.41, −0.50] 1.09 [0.77, 1.47]	−0.49 [−0.44, −0.52] 0.83 [0.65, 0.99]	−0.46 [−0.39, −0.52] 0.48 [0.27, 0.81]	−0.64 [−0.44, −0.89] 0.33 [0.22, 0.42]	−0.76 [−0.64, −0.85] 0.46 [0.25, 0.69]	−0.56 0.64
Castanet	−1.61 [−1.46, −1.75] 3.15 [1.87, 4.16]	−0.94 [−0.89, −1.00] 2.10 [1.48, 2.59]	−2.18 [−1.26, −2.82] 1.49 [0.45, 2.39]	−1.19 [−0.88, −1.47] 1.65 [1.42, 1.81]	−1.41 [−1.14, −1.62] 1.61 [1.28, 2.14]	−1.47 2.00
Hi-hat	−1.41 [−1.14, −1.62] 2.77 [2.01, 3.43]	−0.99 [−0.70, −1.42] 4.55 [2.90, 5.89]	−2.05 [−1.62, −2.39] 3.00 [1.99, 3.75]	−2.27 [−1.77, −2.63] 1.41 [1.21, 1.57]	−2.28 [−1.46, −2.84] 2.13 [1.71, 2.45]	−1.80 2.77

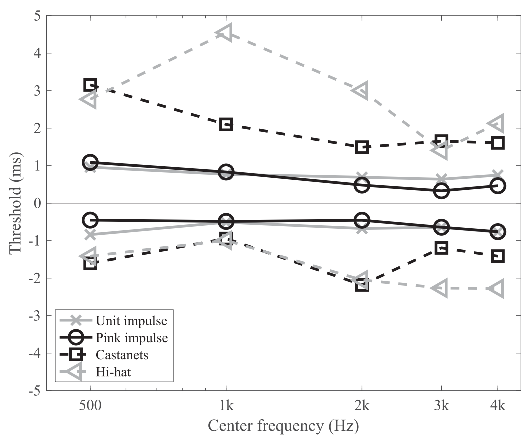

Fig. 11. Group-delay audibility thresholds.

delay in order to create an audible change in audio. This is likely due to the temporal length of the signal, i.e., how much energy appears after, or before in case of the negative peaks, the main part of the sound. In the two impulse-like signals, the majority of sample values are either zero or very small, and the impulse itself is either one sample (unit impulse) or few samples long (pink impulse).

For the castanet recording and the synthetic hi-hat, the signal decay requires time after the initial attack, as seen in Fig. 2. Positive group-delay peaks can move some of the signal energy to occur later in time, and this may be masked by the decay tail of the signals whereas no such masking happens for the negative group-delay effect moving some energy to an earlier time, before the initial attack. This seems to be especially clear in the results for the hi-hat signal at positive group-delay values. The hi-hat signal produces the largest threshold values at three center frequencies while also being the longest of the test signals.

For negative group delays, the differences in threshold values in Fig. 11 are smaller between the impulse-like signals and the other signals, particularly at lower frequencies. This can be explained using Fig. 2. Whereas the castanet and hi-hat signals require time to decay to zero, the sample value before the initial start of signal attack is either zero or really small. Thus, there is less energy to mask the effects of the allpass filters inflicting negative group delay, which results in smaller threshold values. However, the effective signal length does not explain the differences between the threshold values for the impulse-like signals with positive and negative group delay that occur especially at 500 Hz and 1 kHz. A similar difference was also reported in [5], and thus we can say that, below 1 kHz, the audibility of group-delay peaks is greater with negative peaks than positive peaks.

An interesting question about group-delay audibility is frequency dependency. Figure 11 shows that the frequency dependency depends on the test signal. Group-delay peaks (both positive and negative) applied to impulse-like signals are seen to result in horizontal straights indicating that there either is no frequency dependency or that its effect is minor. The two other signals, however, result in non-flat curves. When analyzing the possible reasons for the frequency dependency, no clear causes were found for the hi-hat test signal that produced the largest frequency dependency, especially for positive group-delay peaks. Before the listening test, we hypothesized that the non-flat magnitude of the castanet signal or the reverberation it contained (cf. Figs. 3 and 2, respectively) can affect the audibility of the group delay. For positive group delays, the threshold is seen to be rather constant, apart from the 500-Hz point. This might be due to the lack of energy compared to, e.g., the 2-kHz region. However, for the negative group delays, the 2-kHz data point differs from the overall trend of the castanet curve. This might be because the bulk of the signal energy occurs around the 2-kHz region. Thus, the tail of each successive click can mask part of the following click, and especially the sound energy before the actual click due to an allpass filter with a negative group-delay peak. Alternatively, backward masking can occur with the main click masking the sound that the allpass filter has moved before it.

Finally, we consider some of the verbal comments given by the test subjects. The test was generally considered hard, since there were many test cases where no differences could be heard. At the same time, most test subjects said that they learned to distinguish the audible cues after a certain number of sounds, which helped in the test. They commented that the most noticeable clue was the chirp-like property of some sounds. Many test subjects commented that the test felt binary: one either heard the difference immediately or did not hear it at all. They noticed that listening to a test sound multiple times did not help them. There was no consensus about the hardest test signal, since each of the four test signals was named as the hardest one by

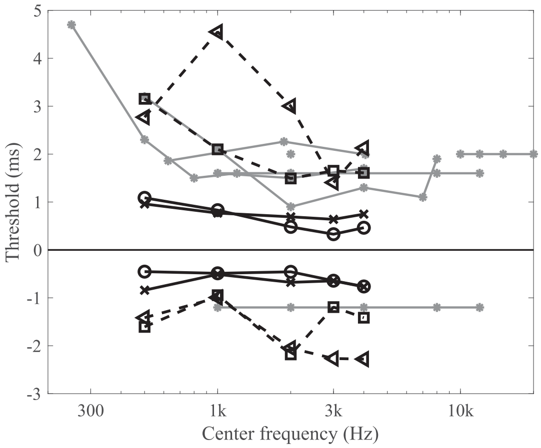

Fig. 12. Comparison of the previously reported group-delay thresholds (gray lines, cf. Fig. 1) and the results of this study (black curves, cf. Fig. 11).

different subjects. The comments from the few subjects who took the test twice did not change between the rounds. The test was still considered hard, but due to the familiar test sounds and the knowledge about the binary nature of the test, the time required for the second test decreased considerably, which was noted in the comments.

V. Discussion

When the results in Fig. 11 are compared to the earlier studies considered in Fig. 1, we see that our impulse-like signals produced lower threshold values than the ones found in the other studies for positive group-delay values. This is shown in Fig. 12. The castanet sound, however, matches the previously reported threshold values well, and the hi-hat sound produced mostly higher values than the ones shown in Fig. 1. Our threshold for the hi-hat signal also varies in a more unpredictable way than any other reported threshold as a function of frequency. For negative group-delay peaks, we can only compare the results from Møller *et al.* [5], [33], [34]. Our test resulted in lower threshold values for the impulse-like signals and in similar threshold values for the two real-life signals, aside from the 2-kHz point and for the hi-hat also above that.

The reason for the lower thresholds for the impulse-like signals found in our study may be the different processing from the other studies and, in the case of the negative group-delay peaks, a different test setup. Our allpass filtering comprising cascaded allpass-filter pairs produces only a single group-delay peak compared to the response of a single second-order allpass filter containing the peak and a constant group delay at low frequencies utilized in previous studies. The excess group delay at low frequencies might have affected the results of the other studies, or, on the other hand, the time-reversed allpass filter may increase the audibility of the group-delay processing by moving energy in front of the main impulse (or after it in the case of negative group-delay peaks). The test setup difference is between the ABX test of the present study and a variable-Q test used by Møller *et al.* [5], [33], [34], which might affect the threshold results for the negative group-delay peaks.

VI. Conclusion

This study comprised a novel signal processing method for creating group-delay peaks with a desired center frequency, and a listening test to determine the audibility thresholds for group delay. The group-delay peaks were generated with a second-order allpass filter and a first- or second-order allpass filter used backwards in time. The latter filter removed the low-frequency group delay caused by the first allpass filter so that only a single peak remained in the group-delay response.

The listening test contained four different signals: the unit impulse, the pink impulse, a castanet recording, and a synthetic hi-hat cymbal sound. The two former signals are known to be the most critical for group-delay audibility, whereas the latter two resemble real-life musical signals and thus help to generalize the results better. Group-delay audibility was tested at the frequencies 500 Hz, 1 kHz, 2 kHz, 3 kHz, and 4 kHz. The results indicate that the audibility thresholds for local group-delay variation are less than ± 1 ms for the most critical signals, and approximately 1.5 ms to 4.5 ms for a local positive group-delay peak and between -1.0 ms and -2.3 ms for a local negative group-delay peak for real-life signals.

In the present study, just like in multiple earlier studies, the impulse-like signals resulted in the lowest threshold values. However, the thresholds obtained in this work for the impulsive signals are considerably smaller than those reported earlier. The thresholds found in our study for the real-life signals were either the same or slightly higher than the ones reported in other studies. This study reports also the audibility of the negative group delay and finds this to be slightly more audible than a positive group delay. Earlier studies have primarily focused on the positive group delay alone.

To summarize, we found that earlier work has been able to establish the audibility threshold rather well. However, the slightly more stringent audibility limits reported in this paper indicate that, in highly critical applications, tighter assumptions about the potential audibility of the delay variation are necessary. The obtained group-delay audibility thresholds are useful when equalizing loudspeakers aiming for very high accuracy. The limits help to determine how much time-domain equalization is needed or, on the other hand, how much magnitude equalization can be applied such that unwanted audible changes to the time structure in audio signals are avoided.

Appendix A Impulse Responses of Second- and First-Order Allpass Filters

Here, we show how the filter coefficients a_1 and a_2 can be used to calculate the impulse response of the second-order allpass section having the transfer function shown in (1), and how the filter coefficient d can be used to calculate the impulse response of the first-order allpass section having the transfer function shown in (7).

In order to obtain the explicit formula for the impulse response of the second-order section, we need to find the partial fraction expansion (PFE) of the transfer function in (1). There are two ways to do this, resulting in two different structures (see, e.g., [56] or [57]). The resulting impulse responses are, however, identical. We have chosen to utilize the less common structure containing an IIR part delayed with respect to the FIR part, since it results in numerically better behavior during implementation [57]. Thus, the resulting transfer function after the PFE is

$$A_2(z) = F + z^{-1}\left(\frac{\tilde{r}_1}{1-p_1 z^{-1}} + \frac{\tilde{r}_2}{1-p_2 z^{-1}}\right), \qquad (18)$$

where F is the FIR part, $\tilde{r}_n$ are the residues, and p_n are the poles.

First, one must find the poles of $A_2(z)$ using the quadratic formula. This results in complex conjugate poles pairs

$$\begin{aligned} p_1 &= \frac{1}{2}\left(-a_1 + \sqrt{a_1^2 - 4a_2}\right), \\ p_2 &= p_1^*, \end{aligned} \qquad (19)$$

when $0 < \phi < \pi$. The next step is to apply the Heaviside cover-up method [58] to the non-causal expression of $A_2(z)$, i.e., the $A_2(z)$ where both the numerator and denominator are multiplied by z^2 to calculate the residues $\tilde{r}_n$:

$$\tilde{r}_n = (z - p_n)A_2(z)|_{z=p_n}. \qquad (20)$$

In the cover-up method, the term $(z - p_n)$ in the denominator of $A_2(z)$ is canceled out to determine the residue r_n of pole p_n. Since the allpass section has two complex conjugate poles, the residues will also be complex conjugates:

$$\begin{aligned} \tilde{r}_1 &= -\frac{(a_2-1)(a_1\sqrt{a_1^2-4a_2} - a_1^2 + 2a_2 + 2)}{2\sqrt{a_1^2-4a_2}}, \\ \tilde{r}_2 &= \tilde{r}_1^*. \end{aligned} \qquad (21)$$

Now we only need to calculate F. This is done by first setting (1) equal to (18) and substituting $z = \infty$. Thus, every component containing the term z^{-1} goes to zero and we are left with

$$F = a_2. \qquad (22)$$

This method works, since F is a constant across the entire frequency range.

Finally, in order to determine the impulse response, we need to calculate the inverse $\mathcal{Z}$-transform of (18). Since the terms of the PFE of $A_2(z)$ are simple, we can utilize inverse $\mathcal{Z}$-transform tables. Note that the delay term in front of the IIR part results in a time-shift:

$$h_2[n] = F\delta[n] + \tilde{r}_1 p_1^{n-1}u[n-1] + \tilde{r}_2 p_2^{n-1}u[n-1] \qquad (23)$$

for $n \geq 0$, where $\delta[n]$ is the unit impulse and $u[n]$ is the unit step function.

The first-order section is more straightforward. The recursive formula is obtained by simply inserting an impulse into the filter and observing the output. Thus, we obtain the following impulse response formula [59]:

$$h_1[n] = -d\delta[n] + (-d^{n+1} + d^{n-1})u[n-1]. \qquad (24)$$

Acknowledgment

The authors would like to thank their colleagues who participated in the listening test during the COVID-19 lockdown in May–June 2020. We are also grateful to Mr. Luis Costa for proofreading the manuscript.

References

[1] D. A. Ronken, "Monaural detection of a phase difference between clicks," *J. Acoust. Soc. Amer.*, vol. 47, no. 4B, pp. 1091–1099, Apr. 1970.

[2] J. H. Patterson and D. M. Green, "Discrimination of transient signals having identical energy spectra," *J. Acoust. Soc. Amer.*, vol. 48, no. 4B, pp. 894–905, Oct. 1970.

[3] D. M. Green, "Temporal acuity as a function of frequency," *J. Acoust. Soc. Amer.*, vol. 54, no. 2, pp. 373–379, Aug. 1973.

[4] J. Blauert and P. Laws, "Group delay distortions in electroacoustic systems," *J. Acoust. Soc. Amer.*, vol. 63, no. 5, pp. 1478–1483, May 1978.

[5] H. Møller, P. Minnaar, S. K. Olesen, F. Christensen, and J. Plogsties, "On the audibility of all-pass phase in electroacoustical transfer functions," *J. Audio Eng. Soc.*, vol. 55, no. 3, pp. 115–134, Mar. 2007.

[6] J. Liski, A. Mäkivirta, and V. Välimäki, "Audibility of loudspeaker group-delay characteristics," in *Proc. Audio Eng. Soc. 144th Conv.*, Milan, Italy, 2018, pp. 1–10.

[7] D. Preis, "Phase distortion and phase equalization in audio signal processing—A tutorial review," *J. Audio Eng. Soc.*, vol. 30, no. 11, pp. 774–794, Nov. 1982.

[8] S. Flanagan, B. C. J. Moore, and M. A. Stone, "Discrimination of group delay in clicklike signals presented via headphones and loudspeakers," *J. Audio Eng. Soc.*, vol. 53, no. 7/8, pp. 593–611, Jul./Aug. 2005.

[9] S. Choisel and G. Martin, "Audibility of phase response differences in a stereo playback system. Part 2: Narrow-band stimuli in headphones and loudspeakers," in *Proc. Audio Eng. Soc. 125th Conv.*, San Francisco, CA, USA, 2008, pp. 1–8.

[10] S. Cecchi, A. Carini, and S. Spors, "Room response equalization—A review," *Appl. Sci.*, vol. 8, no. 16/1, pp. 1–47, Dec. 2017.

[11] H. M. A. Malik, R. Ansari, and A. A. Khokhar, "Robust data hiding in audio using allpass filters," *IEEE Trans. Audio Speech Lang. Process.*, vol. 15, no. 4, pp. 1296–1304, May 2007.

[12] J. Parker and V. Välimäki, "Linear dynamic range reduction of musical audio using an allpass filter chain," *IEEE Signal Process. Lett.*, vol. 20, no. 7, pp. 669–672, May 2013.

[13] R. Greenfield and M. J. Hawksford, "Efficient filter design for loudspeaker equalization," *J. Audio Eng. Soc.*, vol. 39, no. 10, pp. 739–751, Oct. 1991.

[14] M. Karjalainen, E. Piirilä, A. Järvinen, and J. Huopaniemi, "Comparison of loudspeaker equalization methods based on DSP technique," *J. Audio Eng. Soc.*, vol. 47, no. 1/2, pp. 14–31, Jan./Feb. 1999.

[15] A. Mäkivirta, "Loudspeaker design and performance evaluation," in *Handbook of Signal Processing in Acoustics*. New York, NY, USA: Springer, 2008, pp. 649–667.

[16] V. Välimäki and J. Reiss, "All about audio equalization: Solutions and frontiers," *Appl. Sci.*, vol. 6, no. 5, pp. 1–46, May 2016.

[17] R. Wilson, G. Adams, and J. Scott, "Application of digital filters to loudspeaker crossover networks," *J. Audio Eng. Soc.*, vol. 37, no. 6, pp. 455–464, Jun. 1989.

[18] R. Wilson, "Equalization of loudspeaker drive units considering both on- and off-axis responses," *J. Audio Eng. Soc.*, vol. 39, no. 3, pp. 127–139, Mar. 1991.

[19] G. Ramos and J. J. López, "Filter design method for loudspeaker equalization based on IIR parametric filters," *J. Audio Eng. Soc.*, vol. 54, no. 12, pp. 1162–1178, Dec. 2006.

[20] S. H. Linkwitz, "Active crossover networks for noncoincident drivers," *J. Audio Eng. Soc.*, vol. 24, no. 1, pp. 2–8, Feb. 1976.

[21] A. V. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 3rd ed. Upper Saddle River, NJ, USA: Pearson Higher Education, pp. 274–341, 2010.

[22] N. Thiele, "Phase considerations in loudspeaker systems," in *Proc. Audio Eng. Soc. 110th Conv.*, Amsterdam, The Netherlands, May 2001, pp. 1–13.

[23] M. F. Quélhas, A. Petraglia, and M. R. Petraglia, "Efficient group delay equalization of discrete-time IIR filters," in *Proc. 12th Eur. Signal Process. Conf.*, Vienna, Austria, 2004, pp. 125–128.

[24] S. Bharitkar, C. Kyriakakis, and T. Holman, "Time-alignment of multi-way speakers with group delay equalization—I," in *Proc. Audio Eng. Soc. 124th Conv.*, Amsterdam, The Netherlands, May 2008, pp. 1–10.

[25] J. Rämö and V. Välimäki, "Graphic delay equalizer," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Brighton, U.K., 2019, pp. 8018–8022.

[26] V. Adam and S. Benz, "Correction of crossover phase distortion using reversed time all-pass IIR filter," in *Proc. Audio Eng. Soc. 122nd Conv.*, Vienna, Austria, 2007, pp. 1–6.

[27] A. Mäkivirta, J. Liski, and V. Välimäki, "Modeling and delay-equalizing loudspeaker responses," *J. Audio Eng. Soc.*, vol. 66, no. 11, pp. 922–934, Nov. 2018.

[28] S. Herzog and M. Hilsamer, "Low frequency group delay equalization of vented boxes using digital correction filters," in *Proc. Int. Conf. Digit. Audio Effects (DAFx)*, Erlangen, Germany, 2014, pp. 1–8.
[29] E. B. Jensen and H. Møller, "On the audibility of phase distortion in audio systems," in *Proc. Audio Eng. Soc. 47th Conv.*, Copenhagen, Denmark, Mar. 1974, pp. 1–7.
[30] F. A. Bilsen and I. Kievits, "The minimum integration time of the auditory system," in *Proc. Audio Eng. Soc. 86th Conv.*, Hamburg, Germany, 1989, pp. 1–8.
[31] J. A. Deer, P. J. Bloom, and D. Preis, "Perception of phase distortion in all-pass filters," *J. Acoust. Soc. Amer.*, vol. 33, no. 10, pp. 782–786, Oct. 1985.
[32] Y. Hoshino and T. Takegahara, "Permissible value of group delay distortion on tone quality due to low-pass filters," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Tokyo, Japan, 1986, pp. 25–28.
[33] P. Minnaar, F. Christensen, H. Møller, S. K. Olesen, and J. Plogsties, "Audibility of all-pass components in binaural synthesis," in *Proc. Audio Eng. Soc. 106th Conv.*, Munich, Germany, 1999, pp. 1–15.
[34] P. Minnaar, J. Plogsties, S. K. Olesen, F. Christensen, and H. Møller, "The perception of all-pass components in transfer functions," in *Proc. Forum Acust.*, Seville, Spain, Sep. 2002, pp. 1–6.
[35] D. Preis, "Linear distortion," *J. Audio Eng. Soc.*, vol. 24, no. 5, pp. 346–367, Jun. 1976.
[36] S. P. Lipshitz, M. Pocock, and J. Vanderkooy, "On the audibility of midrange phase distortion in audio systems," *J. Acoust. Soc. Amer.*, vol. 30, no. 9, pp. 580–595, Sep. 1982.
[37] H. Suzuki, S. Morita, and T. Shindo, "On the perception of phase distortion," *J. Acoust. Soc. Amer.*, vol. 28, no. 9, pp. 570–574, Sep. 1980.
[38] G. J. Krauss, "On the audibility of group distortion at low frequencies," in *Proc. Audio Eng. Soc. 88th Conv.*, Montreux, Switzerland, 1990, pp. 1–16.
[39] S. Bech, "Quantification of subwoofer requirements, part II: The influence of lower system cut-off frequency and slope and pass-band amplitude and group delay ripple," in *Proc. Audio Eng. Soc. 109th Conv.*, Los Angeles, CA, USA, 2000, pp. 1–38.
[40] R. Greenfield and M. Hawksford, "The audibility of loudspeaker phase distortion," in *Proc. Audio Eng. Soc. 88th Conv.*, Montreux, Switzerland, 1990, pp. 1–16.
[41] V. Hansen and E. R. Madsen, "On aural phase detection," *J. Audio Eng. Soc.*, vol. 22, no. 1, pp. 10–14, Jan./Feb. 1974.
[42] V. Hansen and E. R. Madsen, "On aural phase detection: Part II," *J. Audio Eng. Soc.*, vol. 22, no. 10, pp. 783–788, Dec. 1974.
[43] W. A. Munson and M. B. Gardner, "Loudness patterns—A new approach," *J. Acoust. Soc. Amer.*, vol. 22, no. 2, pp. 177–190, Mar. 1950.
[44] European Broadcasting Union, "Sound quality assessment material: Recordings for subjective tests; user's handbook for the EBU-SQAM compact disc,", Geneva, Switzerland, Sep. 2008, Accessed: May 5, 2020. [Online]. Available: https://tech.ebu.ch/docs/tech/tech3253.pdf
[45] J. S. Abel and J. O. Smith, "Robust design of very high-order allpass dispersion filters," in *Proc. Int. Conf. Digit. Audio Effects (DAFx)*, Montreal, QB, Canada, 2006, pp. 13–18.
[46] J. Liski, A. Mäkivirta, and V. Välimäki, "Companion Page to Audibility of Group-Delay Equalization," 2021, Accessed: Jan. 9, 2021. [Online]. Available: http://research.spa.aalto.fi/publications/papers/ieee-tasl-gdeq/
[47] M. Schoeffler *et al.*, "webMUSHRA—A comprehensive framework for web-based listening tests," *J. Open Res. Softw.*, vol. 6, no. 1, pp. 1–8, Feb. 2018.
[48] B. Treutwein and H. Strasburger, "Fitting the psychometric function," *Percept. Psychophys.*, vol. 61, no. 1, pp. 87–106, Jan. 1999.
[49] F. A. Wichmann and N. J. Hill, "The psychometric function: I. Fitting, sampling, and goodness of fit," *Percept. Psychophys.*, vol. 63, no. 8, pp. 1293–1313, Nov. 2001.
[50] S. A. Klein, "Measuring, estimating, and understanding the psychometric function: A commentary," *Percept. Psychophys.*, vol. 63, no. 8, pp. 1421–1455, Nov. 2001.
[51] M. A. García-Pérez, R. Alcalá-Quintana, R. L. Woods, and E. Peli, "Psychometric functions for detection and discrimination with and without flankers," *Atten. Percept. Psychophys.*, vol. 73, no. 3, pp. 829–853, Apr. 2011.
[52] J. Boley and M. Lester, "Statistical analysis of ABX results using signal detection theory," in *Proc. Audio Eng. Soc. 127th Conv.*, New York, NY, USA, 2009, pp. 1–7.
[53] L. D. Brown, T. T. Cai, and A. DasGupta, "Interval estimation for a binomial proportion," *Statist. Sci.*, vol. 16, no. 2, pp. 101–117, May 2001.
[54] E. B. Wilson, "Probable inference, the law of succession, and statistical inference," *J. Amer. Statist. Assoc.*, vol. 22, no. 158, pp. 209–212, Jun. 1927.
[55] MathWorks Inc., "Curve fitting toolbox," User's Guide. 2020. [Online]. Available: https://se.mathworks.com/help/pdf_doc/curvefit/curvefit.pdf
[56] J. O. Smith, *Introduction to Digital Filters With Audio Applications*. 2008. [Online]. Available: https://ccrma.stanford.edu/~jos/filters/
[57] J. Liski, B. Bank, J. O. Smith, and V. Välimäki, "Converting series biquad filters into delayed parallel form: Application to graphic equalizers," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3785–3795, Jul. 2019.
[58] G. B. Thomas, Jr. and R. L. Finney, *Calculus and Analytic Geometry*, 9th ed. Reading, MA, USA: Addison-Wesley, 1998.
[59] T. I. Laakso and V. Välimäki, "Energy-based effective length of the impulse response of a recursive filter," *IEEE Trans. Instrum. Meas.*, vol. 48, no. 1, pp. 7–17, Feb. 1999.

Juho Liski was born in Helsinki, Finland, in 1989. He received the M.Sc. and Doctor of Science in Technology degrees in electrical engineering from the Aalto University School of Electrical Engineering, Espoo, Finland, in 2016 and 2020, respectively.

Since 2015, he has been with the Acoustics Lab, Aalto University School of Electrical Engineering and is currently a Postdoctoral Researcher, focusing on audio signal processing and equalization.

Dr. Liski is a Member of the Acoustical Society of Finland.

Aki Mäkivirta received the M.Sc., Licentiate of Science, and Doctor of Science in Technology degrees in electrical engineering from the Tampere University of Technology, Tampere, Finland, in 1985, 1989, and 1992, respectively.

After positions with the Medical Engineering Laboratory, Research Centre of Finland and Nokia Corporation Research Center, he joined Genelec Oy, Iisalmi, Finland, in 1995, and since 2012 has been a Technology Director.

Dr. Mäkivirta is a Fellow of the Audio Engineering Society and a Life Member of the Acoustical Society of Finland. He was the recipient of the 2020 CTO of the Year Award in the small and medium-sized enterprise category in Finland.

Vesa Välimäki (Fellow, IEEE) received the M.Sc. and Doctor of Science in Technology degrees from the Helsinki University of Technology (TKK), Espoo, Finland, in 1992 and 1995, respectively.

In 1996, he was a Postdoctoral Research Fellow with the University of Westminster, London, U.K. During 2001–2002, he was a Professor of signal processing with the Pori unit of the Tampere University of Technology, Pori, Finland. During 2006–2007, he was the Head of the TKK Laboratory of Acoustics and Audio Signal Processing. During 2008–2009, he was a Visiting Scholar with Stanford University, Stanford, CA, USA. He is currently a Professor of audio signal processing and the Vice Dean for Research in electrical engineering with Aalto University, Espoo, Finland. His research interests include signal processing for headphones and loudspeakers, artificial reverberation, and sound synthesis.

Prof. Välimäki is a Fellow of the Audio Engineering Society and a Life Member of the Acoustical Society of Finland. During 2007–2013, he was a Member of the Audio and Acoustic Signal Processing Technical Committee of the IEEE Signal Processing Society. During 2005–2009, he was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS and during 2007–2011, of the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. During 2015–2020, he was the Senior Area Editor of the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. He is the Guest Editor of nine special issues published in journals. He was the Chair of the International Conference on Digital Audio Effects in 2008 and the Chair of the International

Conference on Sound and Music Computing in 2017. He is the Editor-in-Chief of the *Journal of the Audio Engineering Society*.