

Real-Time Adaptive Equalization for Headphone Listening

Juho Liski, Vesa Välimäki
Dept. Signal Processing and Acoustics
Aalto University, Espoo, Finland
Email: juho.liski@aalto.fi

Sampo Vesa, Riitta Väänänen
Nokia Technologies
Espoo, Finland

Abstract—The experienced sound quality produced by headphones varies between individuals. Especially with insert headphones a constant equalization may not work properly, when a specific perceived frequency response is desired. Instead, adaptive individualized equalization can be used. Previously this required multiple sensors in a headphone earpiece. This paper proposes a signal processing algorithm for continuous on-line equalization of a headset with a single microphone. The magnitude response of the headphones is estimated using arbitrary reproduced sounds. Then, the headphone response is equalized to a user-selected target response with a graphical equalizer. Measurements show that the proposed algorithm produces accurate estimates with different sound materials and the equalization produces results that closely match the target response. The algorithm can be implemented for multiple applications to obtain accurate and quick personalization, since the target response can be set arbitrarily.

I. INTRODUCTION

Headphones are now widely used to listen to reproduced sounds such as music or binaural sounds. In music listening, the source material is usually intended for stereo loudspeaker playback, and in binaural reproduction, the precise spectral balance at the eardrum of the listener is paramount in order to produce the desired perception. Thus, in both applications the frequency response of the headphones is important due to its effects on the listening experience. Furthermore, since there is no industry standard for the target frequency response, the headphones may need to be equalized to match the required frequency response needed for each application.

One popular headphone type is insert headphones that are placed partly inside the ear canal. They couple directly to the ear canal and produce a unique fit for each person every time they are inserted into the ear, since the ear canals vary between individuals. This varying fit results in varying experienced sound quality even when using the same pair of headphones. Therefore, a similar equalization may not work for everybody, but instead a personalized equalizer (EQ) is required. To design a personal EQ curve, one needs to know the sound pressure at the eardrum.

One way to estimate the sound pressure at the eardrum relies on solving the ear canal parameters, such as the impedance of the ear canal or the eardrum [1], [2], [3]. The parameters are used to construct a physics-based computational model of the ear canal [4], which in turn enables the pressure at the eardrum to be estimated based on the pressure at the ear canal entrance.

The downside is the arduous impedance measurement, which is not suitable for everyday use. Another method utilizes measurements of both the sound pressure and the velocity. Hiiipakka *et al.* proposed an accurate method, which estimates the pressure at the eardrum from measurements performed at the ear canal entrance [5], [6]. The method, however, requires an extra-small sound-velocity probe in addition to the normal sound pressure probe, which thus increases the complexity and cost of the headset. State-of-the-art adaptive headphones calibrate the frequency response only once [7].

This work presents an adaptive equalization system for headphone listening. The system includes a prototype headset, which contains microphones inside the ear canal when the headset is worn. These internal microphones are used during the sound reproduction, and the resulting recording is utilized to estimate the frequency response of each earpiece at the microphone location. The estimate is then mapped to the eardrum with an ear canal model. Finally, a third-octave graphic EQ is designed according to the user-defined target response to obtain a personalized frequency response.

There are many applications where individualized equalization is desired. In addition to music listening and binaural reproduction, specific equalization is required if one pair of headphones is used to simulate another pair. Rämö *et al.* proposed a method to simulate different headphones in a noisy environment, when the frequency response and the isolation properties of the simulated headphones are known [8]. Thus, new headphones could be tested before purchase in order to test their suitability for outdoor listening. Similarly, Olive *et al.* proposed a method to perform double-blind headphone listening tests with a single pair of headphones [9], which enables better comparison of sound quality, since the aesthetics and the feel of the headphones do not affect the listening.

The structure of the paper is as follows. The proposed algorithm is introduced and its building blocks are analyzed in Section II. Section III presents measurement results to illustrate the behavior of the proposed algorithm. Finally, Section IV concludes this paper.

II. ADAPTIVE-EQUALIZATION ALGORITHM

This section presents a novel signal processing algorithm to estimate the individual frequency responses of a pair of

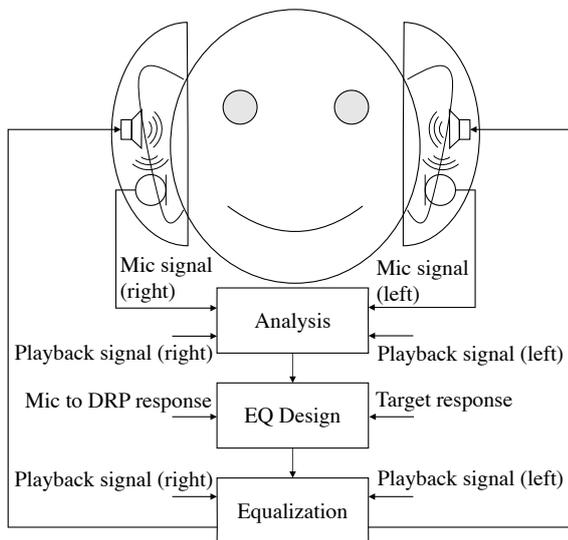


Fig. 1. Block diagram of the complete proposed system.

headphones and equalizing them to a desired target response in real-time.

A. Prototype Headset

The proposed algorithm is designed for a prototype insert headset with strong sound isolation. The prototype contains balanced-armature transducers for audio playback and microphones at the ear-canal sides of the earpieces. These internal microphones are used to estimate the frequency response of each earpiece independently. Although the internal microphones are not located at the eardrums, they can be used to estimate the perceived frequency response, since an ear canal model is utilized to map the response estimate from the microphone location to the eardrum.

B. Analysis Methods

The complete block diagram of the proposed system is shown in Fig. 1. It consists of the prototype headset described in Sec. II-A and three signal processing blocks, the first of which is the analysis block.

The structure of the analysis block is shown in Fig. 2. Here it is shown as a mono algorithm, with similar algorithms running for both ears independently. When the system is activated, a short logarithmic sine sweep is played as a starting sound. The sweep is recorded with the internal microphone in order to obtain the impulse response of the earpiece at the microphone location [10]. From this impulse response three initial values are derived.

The first initial value is the delay caused by the recording chain, which is used to synchronize the playback signal with the microphone recording. In addition, the impulse response acts as an initial guess for the least-mean-square (LMS) algorithm. In order to achieve synchrony, the impulse response is converted to minimum-phase and then truncated to the same length as the LMS algorithm. The third and last application of the impulse response is to estimate the effective ear-canal

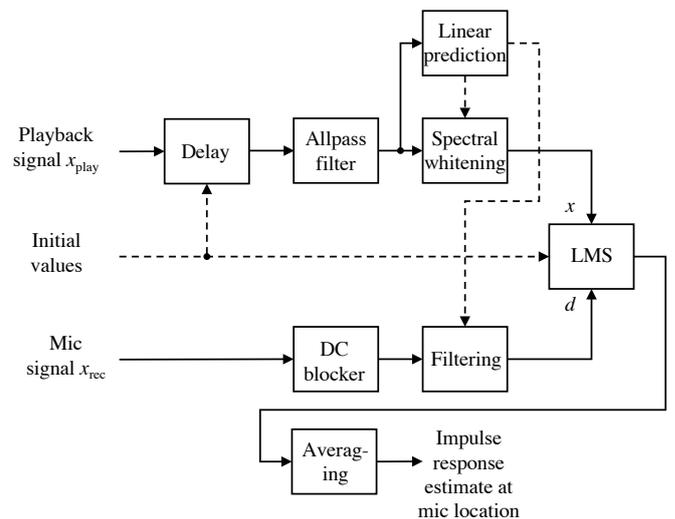


Fig. 2. Block diagram of the analysis part for one ear. The solid lines are signal paths and the dashed lines are control signals.

length needed for the ear canal model. This is described in Sec. II-C.

The proposed system utilizes two sound signals: the playback signal x_{play} and the microphone signal x_{rec} . As mentioned, the playback signal must first be delayed. If the two signals are not synchronized, the LMS algorithm further down the signal chain does not converge to the correct estimate.

Next, the two signals go through their respective filters, i.e., a DC blocker and a corresponding allpass filter. The signal x_{rec} is first filtered with a first-order DC blocker to remove low-frequency disturbances, such as the user's heart-beat. This DC blocker affects frequencies below 100 Hz. Due to the nonlinear group delay of the DC blocker [11], also x_{play} needs to be processed. An allpass filter is designed to compensate for the group delay of the DC blocker. This compensation is achieved by utilizing the same coefficient as with the DC blocker and cascading two identical first-order allpass filters to achieve the same group delay properties.

The next blocks in Fig. 2 are for the spectral whitening of x_{rec} and the corresponding filtering applied to x_{play} . These processes are performed to increase the convergence speed of the following LMS block, since adaptive algorithms converge the fastest with a white-noise input [12]. First, x_{play} goes through a first-order linear predictor, which extracts the spectral envelope of the signal. The obtained autocorrelation coefficients form an all-zero inverse filter $A_{\text{inv}}(z)$. When this filter is applied to x_{play} , its spectrum is approximately whitened. The signal x_{rec} is also filtered with the same filter to maintain the correlation.

Thus far, the blocks preprocess the signals. Next, a version of the LMS algorithm is utilized to estimate the desired impulse response. The LMS block in Fig. 2 contains a robust variable step size normalized LMS (RVSS-NLMS) algorithm, which takes x_{play} as the input signal x and x_{rec} as the desired signal d . The RVSS-NLMS algorithm was proposed by Vega *et al.* [13], and it is based on the minimization of the square

of the *a posteriori error*. It can be interpreted either as an algorithm that has two distinct modes, i.e., NLMS and the normalized sign algorithm (NSA), or as an NLMS algorithm with a variable step size. As a result, the algorithm has both the fast convergence rate of NLMS and the robustness of NSA against noise.

Two parameters need to be selected for the RVSS-NLMS. First, the adaptive filter length is set to $L = 128$, since preliminary tests showed that this offers a good compromise between the low-frequency estimate accuracy and the convergence speed. The other parameter is the memory factor κ , which depends on the color of the input signal [13]. This work uses a value of $\kappa = 6$ since a preliminary test found that this value produces the best results with multiple types of music signals.

Finally, the estimate produced by the LMS algorithm is averaged to reduce the effect of noise in the internal microphone signal and to stabilize the estimate. The averaging leads to slower changes in the estimated frequency response, which is beneficial, since the estimate is used to design the EQ. Rapid changes in the LMS estimate could lead to audible and distracting pumping of the EQ gains. The averaged estimate is the output of the analysis block, which corresponds to the impulse response at the internal microphone location. This result is passed back to the LMS algorithm to act as an initial guess for the next adaptation cycle and, in addition, it is passed forward to the EQ design block.

C. Equalization Methods

After the analysis block, there are two equalization blocks in the proposed system, as is seen in Fig. 1. The first, i.e., EQ design, takes as inputs the estimated impulse response from the analysis block, the microphone-to-drum-reference-point (DRP) response, and the target response. The microphone-to-DRP response is obtained from an ear-canal model.

The ear-canal model is obtained as the average of two different dummy heads: Head Acoustics HMS II and G.R.A.S. KE-MAR. First, the frequency response of the prototype headset was obtained with a sine sweep measurement for both dummy heads in two ways: the response from the transducer to the headset microphone and from the transducer to the dummy-head DRP microphone. When the dB magnitudes of these two frequency responses are subtracted, the remaining response is that of the ear canal. Finally, the two obtained responses were averaged and the default shape for the ear-canal response was extracted.

An example of the ear-canal response is seen in Fig. 3. It shows the default shape as well as one complete ear canal response, when the ear-canal length is 22 mm. The default shape contains a peak at 2 kHz, since that peak is part of the transducer response, but nonetheless does not show up in the internal microphone signal. The full response contains an additional resonance, which in Fig. 3 is seen at 8 kHz. It is due to the resonance inside a closed tube, and its location is derived from the initial values.

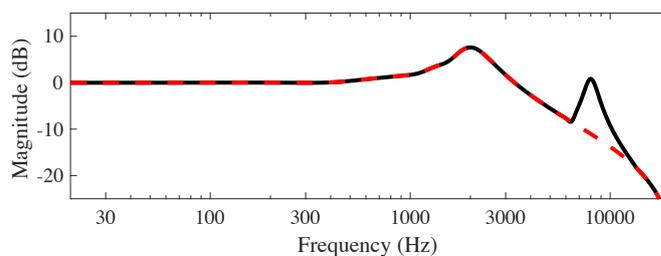


Fig. 3. Default shape for the ear-canal model (dashed red curve) and the modeled ear-canal response for an ear canal of length 22 mm (solid black curve).

When wearing the insert headset, the ear canal is seen as a tube with both ends blocked. Therefore, it acts as a half-wavelength resonator [14], and a pressure minimum is formed near the microphone due to standing waves. Its frequency is determined by the effective length of the ear canal from $f_{r,closed} = \frac{c}{2l}$, where c is the speed of sound and l is the length of the tube. This resonance is seen as a local minimum in the initial frequency response measured with the internal headset microphone. Thus, the resonance peak in the ear-canal response is placed around the same frequency as this minimum. The added peak has a Q value of 5, and its gain is determined so that the peak is 12 dB higher than the default shape at that frequency. Finally, the newly formed ear-canal response is converted to a minimum-phase impulse response.

The complete estimate of the headset frequency response is obtained by convolving the averaged LMS estimate with the ear-canal response and applying third-octave smoothing. Since this response should correspond to the perceived frequency response, in the next step it is compared to the target response to obtain the required gains for the EQ. First, however, the FFT of the complete estimate is obtained and the dB magnitude values calculated.

The target response can be selected arbitrarily depending on the desired application. In this work, good quality spatial reproduction was one of the objectives, and thus two targets were chosen: completely flat response to act as a starting point for further equalization (dashed red curve in Fig. 5(a)) and a listener-preferred headphone target proposed by Olive *et al.* [15] (dashed red curve in Fig. 5(b)). The latter was chosen, since the perceived quality of the frequency response as well as its smoothness affects the size of the perceived space [16].

In order to obtain the EQ gains, the target response and the estimated response are normalized to have the same energy at 1 kHz. Without the normalization the volume-level settings would change the overall level of the magnitude response of the EQ. After the normalization, the EQ response equals to 0 dB at 1 kHz. Now, the difference of the two responses is calculated at the selected EQ-filter center frequencies, and the result is the command gain vector g .

The proposed algorithm utilizes a cascade graphic EQ first proposed in [17]. However, due to the desired equalization accuracy, the octave EQ is unsuitable, and instead the third-

octave version must be used, which is presented in [18]. It has 31 bands with the center frequencies 19.5, 24.6, 31.0, 39.1, 49.2, 62.0, 78.1, 98.4, 124, 156, 197, 248, 313, 394, 496, 625, 787, 992, 1250, 1575, 1984, 2500, 3150, 3969, 5000, 6300, 7937, 10000, 12599, 15874, and 20000 Hz. The corresponding filter bandwidths are 9.1, 11.5, 14.5, 18.2, 22.9, 28.9, 36.4, 45.9, 57.8, 72.8, 91.8, 116, 146, 184, 231, 291, 367, 463, 583, 734, 925, 1166, 1469, 1850, 2331, 2824, 3475, 4219, 4999, 5645, and 5529 Hz.

The filtering is limited to frequencies below 10 kHz, since equalization of higher frequencies may result in audible and distracting artifacts [19]. Thus, the gain of the last three bands is fixed to 0 dB. Finally, the lower limit of the algorithm is set to 100 Hz due to the DC blocker and in order to avoid low-frequency distortion. Therefore, the gains for the lowest seven bands are not obtained directly from the response estimate. Instead, they are extrapolated from the gain of the eighth band so that the gain of the first band with the center frequency of 19.5 Hz equals to 0 dB. This results in a linearly decaying response when moving down the frequency axis starting from approximately 100 Hz.

The final step of the proposed algorithm is the equalization. The left and right EQs designed in the previous block are used together with the stereo playback signal to obtain the equalized signals for each earpiece. When these signals are reproduced with the headset transducers, the combined frequency response of the transducers and the EQs should match closely the user-defined target response in the selected frequency range of 100–10000 Hz.

III. RESULTS

This section discusses the behavior of the proposed algorithm, and validates it with measurements. The algorithm was implemented with MATLAB using the PlayRec utility [20] to simulate real-time usage.

A. Estimation Accuracy

The algorithm frequency-response estimate was compared with the frequency response of the headset. The latter was obtained with a sine-sweep measurement using the dummy head DRP microphone. The measured response is seen as the dashed line in Figs. 4(a) and 4(b). The behavior of the algorithm, on the other hand, was measured with two different music signals (rock with male vocals (a) and pop with female vocals (b)) using the headset internal microphone. The algorithm was run until a stable estimate was found, and the mapping to eardrum was used. The results are the solid curves in Figs. 4(a) and 4(b).

As can be seen in Fig. 4(a), the response estimate from the proposed algorithm closely resembles the ideal response measured with a sine sweep. The largest errors are found below 100 Hz, since a DC blocker is used in the algorithm, and at high frequencies around 10 kHz, where the resonance due to the ear-canal length lies. Between these two extremes the estimation error varies mainly between ± 1 dB, the largest estimation error being approximately 2 dB.

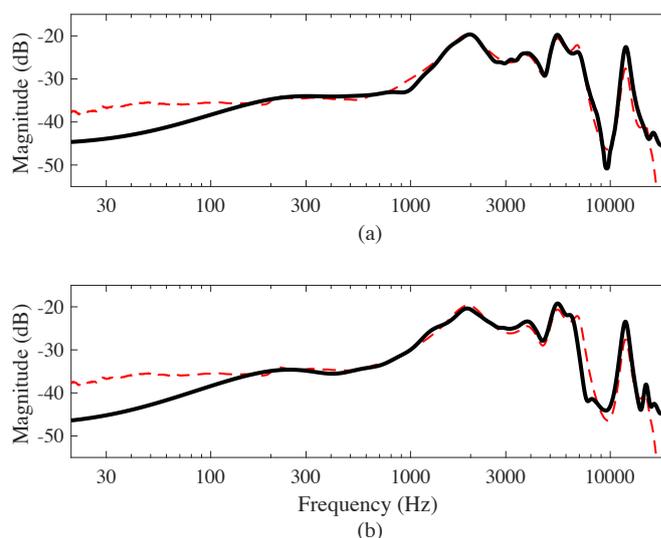


Fig. 4. Frequency response of the prototype headset measured with a sweep (dashed red curve), and an estimate of the same response using the proposed algorithm with (a) rock and (b) pop music (solid black curve).

In Fig. 4(b), a different, calmer music signal was used to demonstrate its effect on the response estimation. As can be seen, the overall shape resembles the one seen in Fig. 4(a), since, for example, there are similar errors below 100 Hz. The largest differences between the two response estimations are found around 10 kHz. The largest differences between the response estimate and the ideal sweep response are also found around the same frequencies. At other frequencies, however, the estimate and the sweep response match closely, and the estimation error mainly varies between ± 1 dB. Therefore, the estimates produced by the algorithm are accurate with multiple types of signals and can be used in the EQ gain derivation.

B. Equalization Accuracy

After the accurate response estimate was obtained with the analysis part of the algorithm, two different equalization filters were designed based on the target responses (a flat target and the Olive target mentioned in Sec. II-C). Next, sweep measurements were performed, and the results were recorded with the dummy head DRP microphone. When the EQ was utilized in the sweep measurement, the combined response of the headset and the EQ filter was obtained. The measured frequency responses are shown in Figs. 5(a) and 5(b). Note that the frequency band is limited to 100–10000 Hz.

Figure 5(a) shows the measured response as the solid black curve, when the flat target response was used. As can be seen, the measured response follows the target response well in the chosen frequency range. The largest deviations are found around 100 Hz, due to the small estimation error discussed in Sec. III-A, and at high frequencies (between 4–5 kHz and around 9 kHz), due to the rapid changes in the headset response, which the EQ is unable to flatten. However, when the result is compared to the starting point (dashed red curve

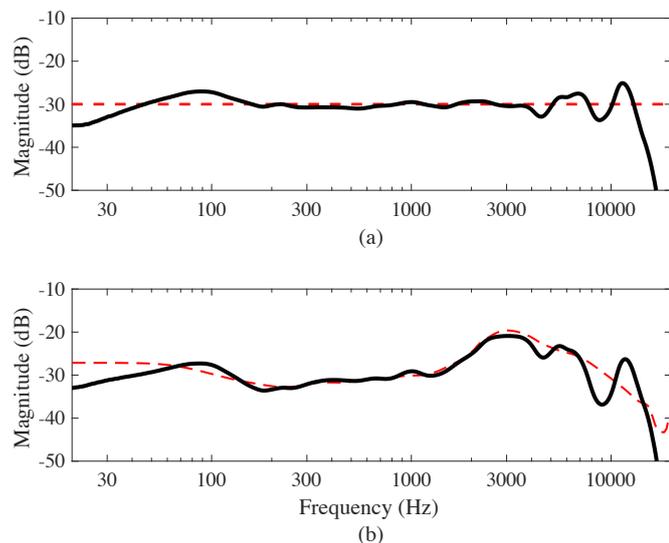


Fig. 5. Measured response with (a) a flat target and (b) the Olive target. The dashed red curves are the targets and the solid black curves are the measured responses.

in Fig. 4), the improvements towards the target response are clearly observed.

In Fig. 5(b), the Olive target response (dashed red line) was used. The solid black curve shows the measured combined effect of the earpiece transducer and the EQ, where similar characteristics as in Fig. 5(a) are seen. Below 100 Hz the responses differ due to the EQ design choices described in Sec. II-C. Additional differences are seen around 10 kHz similarly to the flat target case, which are due to the peakiness of the headset response at these frequencies. Otherwise, however, the measured response closely matches the target response, and thus the proposed algorithm provides the desired effect.

IV. CONCLUSIONS AND FURTHER WORK

This work presented a novel adaptive equalization algorithm for headphone listening. The algorithm estimates the individual frequency response of the prototype headset on-line using the reproduced sounds. The individual ear-canal length is first estimated in order to map the internal microphone response to the eardrum. The obtained response is compared to a target response, and the proposed method then designs a cascade graphical EQ to produce the desired response for the user.

Measurement results indicate that the algorithm accurately estimates the frequency response and that the equalized response follows the user-defined target response between the selected frequency range of interest (100–10000 Hz). Therefore, the proposed algorithm can be utilized, for example, in applications where the frequency response of the headphones needs to be similar for every person despite the differing ear-canal acoustics, or where the headphone could move during listening so that the equalization needs to adapt.

Further work on the topic includes a real-time implementation of the algorithm and a usability test. The usability test could provide information on the change rate of the

EQ gains in order to find the optimal values that result in the most pleasant listening experience. Another topic would be the expansion of the algorithm for different types of headphones. This would require altered ear-canal models due to the differing internal microphone locations.

ACKNOWLEDGMENT

This work was conducted as a part of the ADEQ project (Aalto project number 410642) funded by Nokia Technologies. The authors would like to thank Matti Hämäläinen for helpful discussions and Luis Costa for proofreading.

REFERENCES

- [1] H. Hudde, "Estimation of the area function of human ear canals by sound pressure measurements," *J. Acoust. Soc. Am.*, vol. 73, no. 1, pp. 24–31, Jan. 1983.
- [2] D. H. Keefe, R. Ling, and J. C. Bulen, "Method to measure acoustic impedance and reflection coefficient," *J. Acoust. Soc. Am.*, vol. 91, no. 1, pp. 470–485, Jan. 1992.
- [3] S. Schmidt and H. Hudde, "Accuracy of acoustic ear canal impedances: Finite element simulation of measurement methods using a coupling tube," *J. Acoust. Soc. Am.*, vol. 125, no. 6, pp. 3819–3827, Jun. 2009.
- [4] H. Hudde, A. Engel, and A. Ludwig, "Methods for estimating the sound pressure at the eardrum," *J. Acoust. Soc. Am.*, vol. 106, no. 4, pp. 1977–1992, Oct. 1999.
- [5] M. Hiipakka, M. Karjalainen, and V. Pulkki, "Estimating pressure at eardrum with pressure-velocity measurement from ear canal entrance," in *Proc. WASPAA'09*, New Paltz, NY, Oct. 2009, pp. 289–292.
- [6] M. Hiipakka, "Estimating pressure and volume velocity in the ear canal for insert headphones," in *Proc. ICASSP*, Prague, Czech Republic, May 2011, pp. 289–292.
- [7] J. Backman, T. Campbell, J. Kleimola, and M. Hiipakka, "A self-calibrating earphone," in *Proc. AES 142nd Conv.*, Berlin, Germany, May 2017.
- [8] J. Rämö and V. Välimäki, "Signal processing framework for virtual headphone listening tests in a noisy environment," in *Proc. AES 132nd Conv.*, Budapest, Hungary, Apr. 2012.
- [9] S. E. Olive, T. Welti, and E. McMullin, "A virtual headphone listening test methodology," in *Proc. AES 51st Int. Conf.*, Helsinki, Finland, Aug. 2013.
- [10] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. AES 108th Conv.*, Paris, France, Feb. 2000.
- [11] R. Yates and R. Lyons, "DC blocker algorithms," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 132–134, Mar. 2008.
- [12] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Prentice Hall, 1985.
- [13] L. R. Vega, H. Rey, J. Benesty, and S. Tressens, "A new robust variable step-size NLMS algorithm," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1878–1893, May 2008.
- [14] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sørensen, "Design criteria for headphones," *J. Audio Eng. Soc.*, vol. 43, no. 4, pp. 218–232, Apr. 1995.
- [15] S. Olive, T. Welti, and E. McMullin, "Listener preferences for in-room loudspeaker and headphone target responses," in *Proc. AES 135th Conv.*, New York, USA, Oct. 2013.
- [16] P. Gutierrez-Parera and J. J. Lopez, "Influence of the quality of consumer headphones in the perception of spatial audio," *Appl. Sci.*, vol. 6, no. 4, Apr. 2016.
- [17] V. Välimäki and J. Liski, "Accurate cascade graphic equalizer," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 176–180, Feb. 2017.
- [18] J. Liski and V. Välimäki, "The quest for the best graphic equalizer," in *Proc. Int. Conf. Digital Audio Effects (DAFx-17)*, Edinburgh, UK, Sep. 2017.
- [19] V. Riikonen, M. Tikander, and M. Karjalainen, "An augmented reality audio mixer and equalizer," in *Proc. AES 124th Conv.*, Amsterdam, The Netherlands, May 2008.
- [20] R. Humphrey, "Playrec, multi-channel MATLAB audio [Online]," <http://www.playrec.co.uk/>, 2008–2014, [accessed Feb. 14, 2017].