COLORATION METRICS FOR HEADPHONE EQUALIZATION

Braxton Boren[1*], *Michele Geronazzo*[2], *Fabian Brinkmann*[3], *Edgar Choueiri*[1]

[1]Princeton University
3D Audio and Applied Acoustics Lab
Princeton, NJ USA
*bboren@princeton.edu

[2]University of Padova
Dept. of Information Engineering
Padova, Italy

[3]TU Berlin
Audio Communication Group
Berlin, Germany

ABSTRACT

Headphone equalization is necessary for accurate binaural reproduction over headphones, but so far no metrics have been adopted for evaluating human perception of spectral coloration in post-equalization headphone transfer functions (HpTFs). A metric for peak error is proposed that represents the average HpTF error from narrow peaks per third octave band. In addition, a new metric for broadband error is defined by subtracting the average error from narrow peaks and notches from that of an auditory filter bank model. Used together, the peak error and broadband error terms are shown to represent the critical information necessary for transparent headphone reproduction.

1. INTRODUCTION

For non-binaural audio applications listeners do not usually prefer a completely flat headphone frequency response [1, 2], but because of the spectral sensitivity of spatial hearing a flat headphone response is important for accurate binaural rendering of a virtual auditory space. Previous work suggests that flatter frequency responses are important for sound externalization using non-individual head-related transfer functions (HRTFs) [3], that headphone equalization may improve consistency in auditory distance perception [4], and that incorrect headphone equalization may in some cases degrade auditory localization [5].

The headphone transfer function (HpTF) consists of lower-frequency resonance effects primarily due to the volume enclosed by the headphone cups and the higher-frequency resonances of the listener's pinnae [6, 7]. Early examinations of HpTFs employed a theoretical model wherein the headphone's response could be completely removed from the audio chain via inverse filtering [8, 9, 10]. However, in practice the refitting of headphones by the same listener leads to significant geometric changes at small wavelengths, leading to shifting of the pinna-dependent notches at high frequencies [11]. Equalization via simple inversion of the measured response at these frequencies leads to large peaks in the inverse filter which do not match up with the notches from the earlier fitting, and psychoacoustic studies show that such peaks are much more noticeable than the notches they are intended to equalize [12, 13]. Because of this, more recent headphone equalization approaches use some form of discrimination to reduce gain at high frequencies [14, 15, 16, 17], though some recent approaches still use full-spectrum inversion [18].

Pralong and Carlile found significant deviations between HpTFs for different subjects above 4 kHz, which led them to emphasize the importance of individualization in headphone equalization [19]. However, 4 kHz is also approximately the same frequency at which variations begin to be introduced by refittings of the same headphones for the same individual [14, 16]. Thus inter-subject and intra-subject variations are sufficiently intertwined that it is difficult to account for one and not the other. Since the frequency-dependent headphone equalization algorithms mentioned above reduce their gain at high frequencies, they may also reduce some of the benefits from individualized equalization. It may be the case that much of the advantage from headphone equalization may be obtained using a generic filter that only takes into account the resonance of the headphones themselves and not individual anthropometry [15, 17].

The various headphone equalization algorithms in use today use a variety of frequency discrimination methods to establish a robust filter for transparent reproduction. In particular, these include a frequency-domain peak compression algorithm [14], a statistical approach inverting the 95th percentile of each frequency bin's magnitude for a set of HpTF measurements [16], and a variety of frequency regularization methods [15], While these algorithms all have similar goals (i.e. the reduction of large high frequency peaks), they use very different approaches, and in general their parameters have been hand-tuned using informal listening tests over small databases of HpTFs.

Using the Spatially Oriented Format for Acoustics (SOFA) [20], we have consolidated many of the existing HpTF databases into a single publicly available dataset [21]. This allows the large-scale comparison of different algorithms and input parameters under a variety of different conditions. In addition, machine learning techniques could be used to design completely new filters for more transparent generic or user-specific HpTF equalization, depending on what data inputs are available. However, such approaches require numerical metrics for auditory coloration perception in order to be evaluated on a large scale data base.

In this paper we develop a 2-dimensional metric for the perceptual contributions of both narrow 'ringing' peaks and broadband coloration for equalized HpTFs. We first examine the existing psychoacoustic literature on coloration perception and discuss the benefits and shortcomings of the existing auditory filter bank model for spectral error. Next we develop an algorithm for detecting narrow peaks and notches, and use these to define numerical metrics of peak error and broadband error in post-EQ HpTFs.

These two error metrics allow a more precise analysis of the narrowband and broadband contributions to auditory coloration perception.

2. BACKGROUND

2.1. Coloration Perception

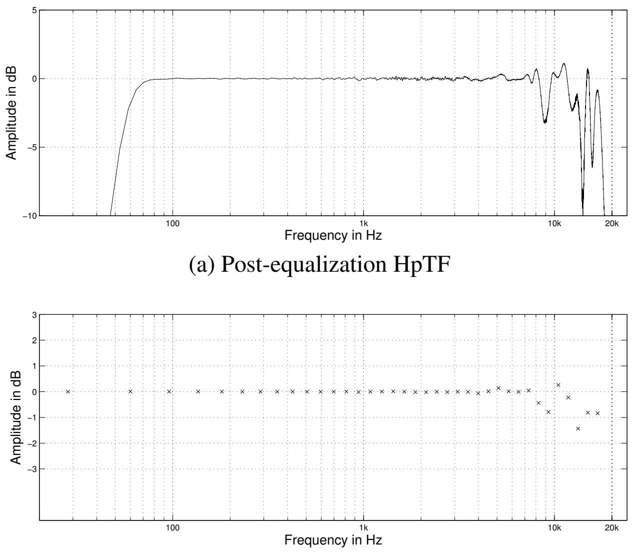
Previous studies on auditory coloration perception have investigated the detection thresholds for spectral peaks and notches of varying gains, bandwidths, and center frequencies. In general these have concluded that narrow peaks are much more perceptible than narrow notches and also that resonances of any kind become more noticeable as their bandwidth increases [22]. Bucklein [12] observed differences by frequency, but without any obvious pattern. Green [23] found a roughly U-shaped detection threshold with higher sensitivity at mid-frequencies, similar but not identical to an A-weighted equal loudness contour. Olive and his colleagues also found that the shapes of perceptual detection threshold contours may vary based on the type of signal being presented [24]. Moore [13] performed many experiments at two frequency bands, and concluded that sensitivity was generally higher at 1 kHz than at 8 kHz. In the absence of such detailed studies across the entire audible spectrum, numerical representations of spectral coloration have not thus far incorporated frequency-dependence into their calculations [15, 17].

2.2. Auditory Filter Models

An approximation of auditory coloration might simply use a logarithmic integration of peaks and notches relative to a flat target spectrum. But since the auditory system's frequency perception is not completely logarithmic but rather made up of critical bands varying in width, a more accurate approach is to sum the error contributions along equivalent rectangular bandwidth (ERB) filters. This was the approach used by Scharer and Lindau [17], who measured error in reference to a target spectrum along 40 ERB filters across the audible spectrum, as shown in figure 1.

Although a completely flat spectrum could be used as a target for error calculations, in practice this may cause unwanted high filter gains beyond the low- or high-frequency rolloff of the headphones. Because of this, the target function was a linear phase FIR bandpass filter with pass band from 50 to 21000 Hz [17]. The target function's low-frequency rolloff is the reason why figure 1(b) has such low error values for the lowest critical bands. This approach in general yielded small errors in the lower-frequency bands, increasing at the higher frequencies where the HpTF experiences the most variation. A single average ERB error metric, E_{ERB}, was calculated as the mean of the modulus of the error for each of the ERB filters.

This metric depicts the auditory system's overall perception of large-scale coloration well in most cases. However, in the context of HpTF equalization, the relatively wide-band ERB filter bank may underemphasize large narrow peaks that may be produced under certain equalization schemes. This may happen for several reasons: first, high-Q resonances contain relatively little energy in the context of a broadband auditory filter. Second, within the range of a single auditory filter, complementary peaks and notches could cancel each other out, yielding low total band error while still sounding audibly colored. Finally, it is possible for such peaks, on the cusp of two critical bands, to split their energy between two ERB filters.



(a) Post-equalization HpTF

(b) Calculated error for each band of an auditory filter bank

Figure 1: Example of ERB error calculated in [15, 17]. The post-equalization HpTF is compared to a target spectrum at 40 bands of an auditory filter bank.

While narrow notches still have a very high detection threshold, narrow peaks are much more likely to be perceptually salient [13]. Since the elimination of these high frequency 'ringing' effects is one of the primary goals of headphone equalization, a different metric is needed that detects such artifacts more accurately and consistently than the auditory filter model. Therefore we propose a new metric to account for narrow peak coloration.

3. PEAK ERROR

Motivated by the considerations above, we have developed an algorithm to detect the net contributions from relatively narrow peaks in a given post-equalization HpTF spectrum. Given an input power spectrum H, we create two smoothed versions of the spectrum: one coarsely smoothed version using full-octave smoothing, H_c, and another finely smoothed version using 1/48 octave smoothing, H_f. Using these we first calculate the difference spectrum $H_d = H_f - H_c$.

All peaks on H_d are located by finding changes in the sign of the gradient of H_d. From this longer list of peaks, smaller peaks below a threshold are removed to account for signal noise. We chose a threshold value of 1 dB to find perceptually significant peaks, similar to [17]. The peaks are sorted in order of decreasing height, and smaller peaks within ± 1/6 octave of a higher peak are removed. This prevents the over-counting of doublet or triplet peaks above the threshold value, which cause broadband coloration rather than the 'ringing' associated with narrow peaks. As shown in figure 2, this increases the numerical contribution of narrow peaks by emphasizing their difference with the spectrum immediately around them. Conversely, the contribution of very wide peaks is lessened, as it is already captured well by the ERB filter bank. This allows detection of peaks whose values are below 0 dB but may still be perceived as ringing within the context of a larger broadband notch. Additionally, here we are normalizing to the average level between 200 and 400 Hz, not the average level over the entire spectrum, since this frequency range provides a more constant magnitude response and is more invariant toward measurement errors [25].

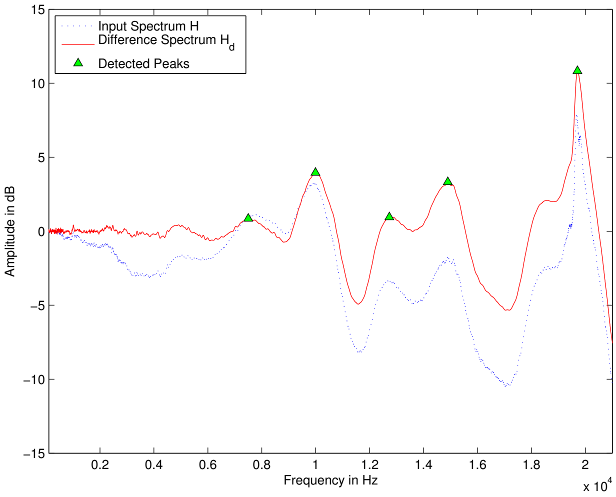


Figure 2: Example of peak detection on an HpTF.

Using the vector of remaining peak locations p, the peak error E_{pk} is then determined by

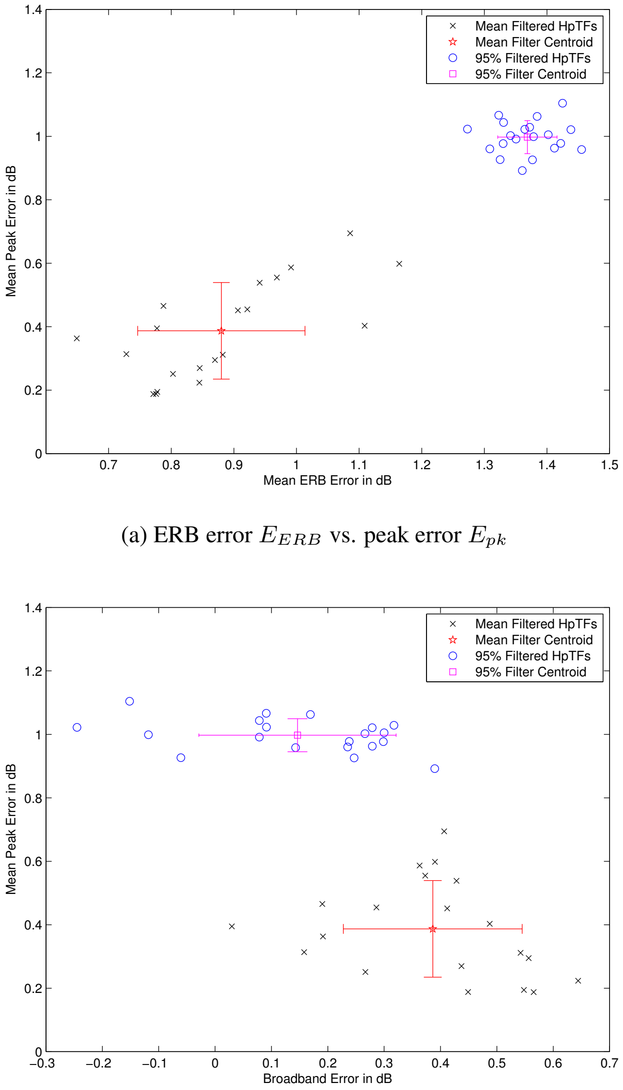
$$E_{pk} = \frac{\sum_p H_d(p)}{3[\log_2(f_{high}/f_{low})]}, \tag{1}$$

where the denominator represents the number of third octave bands within the frequency range of interest. Using values of $f_{low} = 50$ and $f_{high} = 21000$, this amounts to a division by about 26. This operation scales common post-EQ values of E_{pk} into a range of 0-2 dB, similar to E_{ERB}.

4. ANALYSIS

To examine the usefulness of the proposed metric, both peak error E_{pk} and mean ERB error E_{ERB} were calculated for sets of 20 repeated HpTF measurements from the PHOnA HpTF dataset [21]. HpTFs were equalized with two different statistical algorithms derived from the entire set of measurements: a mean inverse filter and an inversion of the 95th percentile of all measurements [16].

Using these data, each HpTF can be plotted on a 2 dimensional space showing its mean ERB error and peak error. For many subjects the additional metric of E_{pk} is sufficient to differentiate narrowband ringing peaks from broadband coloration effects. However, because ERB error may be affected differently by different arrangements of peaks and notches, in some cases ERB error may be highly correlated with peak error if the broadband error is low compared to the error introduced by narrow peaks and notches. Figure 3(a) shows the right ear plots for a single subject, along with the centroids and standard deviations for both algorithms in terms of E_{ERB} and E_{pk}. Viewed in this way, the 95% inversion algorithm seems objectively inferior to a mean filter in terms of either error metric.



(a) ERB error E_{ERB} vs. peak error E_{pk}

(b) Broadband error E_{br} vs. peak error E_{pk}

Figure 3: 2D plots of E_{ERB}(a) and E_{br}(b) on the x-axis against E_{pk} on the y-axis for the same set of equalized HpTFs for a single subject. Centroids and standard deviations for mean and 95% inversion algorithms are displayed in red and magenta, respectively.

Because of this observation, we decided to modify the ERB error metric by removing the contributions of narrowband error. This required us to develop a similar metric for the notch error E_n. It will be seen that it is not necessary to create a separate notch-detection algorithm. Instead, the equivalent peak detection approach is applied to the negative difference spectrum $-H_d = H_c - H_f$. While narrow notch error is not usually perceptible, such notches may contribute to the overall ERB error metric. The notch error value E_n is also strictly non-negative, corresponding to the absolute value of the mean notch depth per third octave band.

Having calculated E_{pk} and E_n, we then define a new broadband error measure, E_{br}, which removes the average narrowband error, defined as the mean of the peak and notch error, from the mean ERB error value:

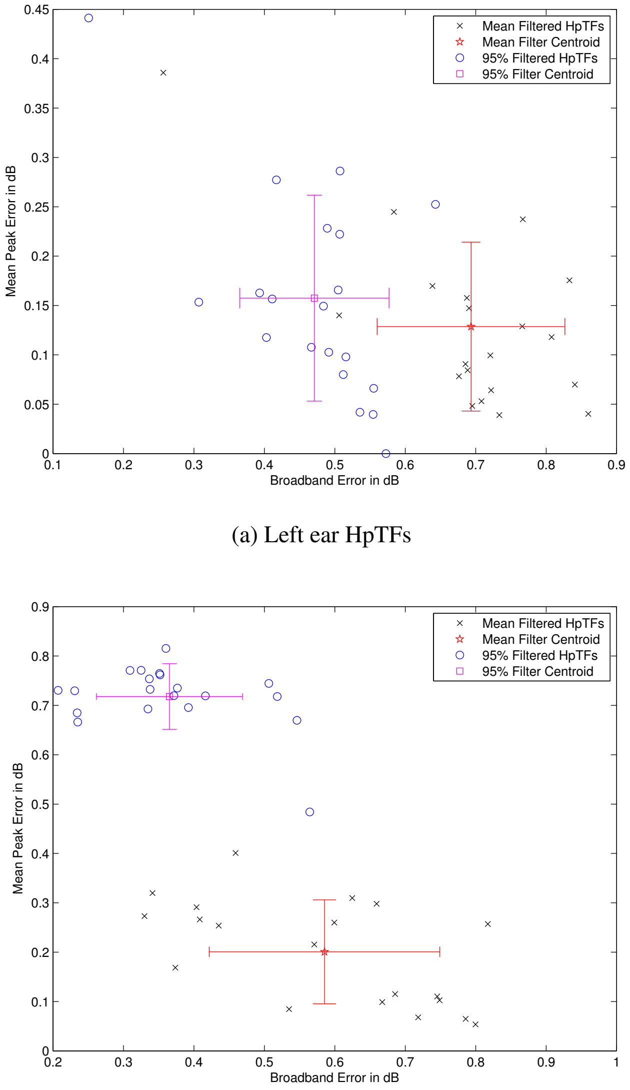
$$E_{br} = E_{ERB} - \frac{E_{pk} + E_n}{2}. \qquad (2)$$

This definition allows the possibility of negative values of E_{br}, especially in cases where a signal has large narrow peaks and notches that cancel each other out within the scope of a given ERB filter. But this is still useful information, as it shows that in these cases the post-EQ spectrum's coloration comes almost entirely from narrowband artifacts rather than broadband filtering effects.

We can see in figure 3(b) that plotting E_{pk} against E_{br}, for the same data as figure 3(a), distinguishes the effects of different forms of coloration in both equalization algorithms. Rather than forming a neat line, we now observe a diagonal split between the two algorithms, with the 95% inversion HpTFs grouping together in the upper-left corner and the mean-filtered HpTFs in the bottom-right. The 95% inverse filter generates uniformly higher values of E_{pk} but slightly lower values of E_{br} than does the mean filter. Because the 95% inversion creates more notches than peaks, the average narrowband error is greater than E_{pk} for most of the spectra.

Figure 4 shows measurements for both ears of another subject from the dataset. In 4(a), there is a left-right split, yielding similar E_{pk} values for both algorithms but slightly greater E_{br} values for the mean-filtered HpTFs. In 4(b), however, there is another pronounced diagonal split, similar to figure 3(b). In the left-right split case, the 95% inversion appears to be a better choice of equalization. However, more perceptual testing and perhaps task-specific information would be needed to say definitively which of the algorithms is performing better in diagonally-split cases.

We were initially surprised that the peak error was higher in most cases for the 95% inverse filter than for a mean filter given that Masiero proposed a 95% inversion to preferentially choose large notches over peaks. The key distinction in this case is that the 95% inversion tries to avoid *large* peaks in the post-EQ spectrum, while the metric proposed here selects *narrow* peaks, whatever their size. Upon qualitative inspection of the spectra in question, we noticed that because of the algorithm's preference for large notches this created a chisel-like effect that also created many small narrow peaks in the spectrum. Increasing the value of the peak threshold would reduce the contribution from these peaks, but it is our opinion that 1 dB is an appropriate threshold for finding perceptually salient peaks. Still, the exact threshold and bandwidth limits of the peak and notch detection function could be subject to further parameterization for other HpTF equalization algorithms.



(a) Left ear HpTFs

(b) Right ear HpTFs

Figure 4: Plots of broadband error E_{br} (x-axis) and peak error E_{pk} (y-axis) for another subject from the PHOnA dataset. Centroids and standard deviations for mean and 95% inversion algorithms are displayed in red and magenta, respectively.

5. CONCLUSION

A rigorous comparison of different equalization schemes must consider not only the separate dimensions of inter-headphone, inter-subject, and intra-subject spectral variations, but also the additional dimensions introduced by each equalization algorithm and its input parameters [14, 15, 16, 17] that have thus far been fine-tuned by hand. Any attempt to compare HpTFs along all these dimensions will inevitably require more data than can be reliably analyzed by visual observation. Computational analysis over a large database may allow us to find subtle trends in large datasets or to investigate the many algorithmic parameters that remain unexplored. In addition, machine learning techniques may be employed to design new equalization techniques based on inputs based on the user's headphones or anthropometry.

However, computational optimization requires a consensus coloration metric to be minimized. We have argued here that the existing mean ERB error metric is not sufficient to fully represent the dual problems of narrow peaks and broadband coloration in HpTF equalization. While it would be possible to minimize only E_{br} or E_{pk}, probably a better solution would be to minimize a linear combination of these two error metrics or perhaps just the Euclidean distance from the origin in the 2-space shown here (adding a zero-floor for negative values of E_{br}). Subjective perceptual testing and evaluation will be needed to determine the correct relationship between these two metrics so that computational processes can come as close as possible to the experience of a human listening over headphones.

Finally, this research direction leads the way for the development of an auditory model for coloration and spectral profiling induced by headphones and for the use of generic HRTFs that will be used in an increasing number of applications, such as spatial audio in web browsers [26].

6. ACKNOWLEDGMENTS

This work was supported by a grant from the Sony Corporation and by the research project "PADVA - Personal Auditory Displays for Virtual Acoustics" (no.CPDA135702) of the University of Padova.

7. REFERENCES

[1] Henrik Moller, Clemen Jensen, Dorte Hammershoi, and Michael Sorensen. Design Criteria for Headphones. *Journal of the Audio Engineering Society*, 43(4):218–232, 1995.

[2] Sean E Olive, Todd Welti, Elisabeth Mcmullin, and Harman International. The Influence of Listeners Experience, Age, and Culture on Headphone Sound Quality Preferences. In *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, 2014.

[3] Braxton B Boren and Agnieszka Roginska. The Effects of Headphones on Listener HRTF Preference. In *Proceedings of the 131st Audio Engineering Society Convention*, New York, NY, 2011.

[4] Kaushik Sunder, Ee-Leng Tan, and Woon-Seng Gan. Effect of Headphone Equalization on Auditory Distance Perception. In *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, 2014.

[5] D Schonstein, L Ferr, and B F G Katz. Comparison of headphones and equalization for virtual auditory source localization. In *Acoustics '08*, pages 4617–4622, Paris, 2008.

[6] EAG Shaw and R Teranishi. Sound Pressure Generated in an External-Ear Replica and Real Human Ears by a Nearby Point Source. *The Journal of the Acoustical Society of America*, 44(1), 1968.

[7] Hironori Takemoto, Parham Mokhtari, Hiroaki Kato, Ryouichi Nishimura, and Kazuhiro Iida. Mechanism for generating peaks and notches of head-related transfer functions in the median plane. *The Journal of the Acoustical Society of America*, 132(6):3832–41, December 2012.

[8] F L Wightman and D J Kistler. Headphone simulation of free-field listening. I: Stimulus synthesis. *The Journal of the Acoustical Society of America*, 85(2):858–67, February 1989.

[9] Henrik Moller. Fundamentals of binaural technology. *Applied Acoustics*, 36:171–218, 1992.

[10] Henrik Moller, Dorte Hammershoi, Clemen Jensen, and Michael Sorensen. Transfer Characteristics of Headphones Measured on Human Ears. *Journal of the Audio Engineering Society*, 43(4):203–217, 1995.

[11] Abhijit Kulkarni and H Steven Colburn. Variability in the characterization of the headphone transfer-function. *Journal of the Acoustical Society of America*, 107(2):1071–1074, 2000.

[12] R. Bucklein. The Audibility of Frequency Response Irregularities. *Journal of the Audio Engineering Society*, 29(3):126–131, 1981.

[13] B C Moore, S R Oldfield, and G J Dooley. Detection and discrimination of spectral peaks and notches at 1 and 8 kHz. *The Journal of the Acoustical Society of America*, 85(2):820–36, February 1989.

[14] Timo Hiekkanen, Aki Makivirta, and Matti Karjalainen. Virtualized Listening Tests for Loudspeakers. *Journal of the Audio Engineering Society*, 57(4):237–251, 2009.

[15] Alexander Lindau and Fabian Brinkmann. Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings. *Journal of the Audio Engineering Society*, 60(1):54–62, 2012.

[16] Bruno Masiero and Janina Fels. Perceptually Robust Headphone Equalization for Binaural Reproduction. In *Proceedings of the 130th Audio Engineering Convention*, London, 2011.

[17] Zora Scharer and Alexander Lindau. Evaluation of Equalization Methods for Binaural Signals. In *Proceedings of the 126th Audio Engineering Society Convention*, Munich, Germany, 2009.

[18] G Wersényi. Evaluation of a Matlab-based Virtual Audio Simulator with HRTF-Synthesis and Headphone Equalization. In *Proceedings of ICAD 12-Eighteenth Meeting of the International Conference on Auditory Display*, pages 221–224, Atlanta, GA, 2012.

[19] D Pralong and Simon Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *The Journal of the Acoustical Society of America*, 100(6):3785–93, December 1996.

[20] Piotr Majdak, Yukio Iwaya, Thibaut Carpentier, Rozenn Nicol, Matthieu Parmentier, Agnieszka Roginska, Yoiti Suzuki, Kanji Watanabe, Hagen Wierstorf, Harald Ziegelwanger, and Markus Noisternig. Spatially Oriented Format for Acoustics. In *Proceedings of the 134th Audio Engineering Society Convention*, Rome, Italy, 2013.

[21] Braxton B Boren, Michele Geronazzo, Piotr Majdak, and Edgar Choueiri. PHOnA : A Public Dataset of Measured Headphone Transfer Functions. In *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, 2014.

[22] Floyd E Toole and Sean E Olive. The Modification of Timbre by Resonances: Perception and Measurement. *Journal of the Audio Engineering Society*, 36(3):122–142, 1988.

[23] David M Green and Christine R Mason. Auditory profile analysis : Frequency, phase, and Weber's Law. *Journal of the Acoustical Society of America*, 77(3):1155–1161, 1984.

[24] Sean E Olive, Peter L Schuck, James G Ryan, Sharon L Sally, and Marc E Bonneville. The Detection Thresholds of Resonances at Low Frequencies. *Journal of the Audio Engineering Society*, 45(3):116–128, 1997.

[25] Fabian Brinkmann, Alexander Lindau, and Stefan Weinzierl. ASSESSING THE AUTHENTICITY OF INDIVIDUAL DYNAMIC BINAURAL SYNTHESIS. In *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, volume 71, pages 62–68, Berlin, Germany, 2014.

[26] Michele Geronazzo, Jari Kleimola, and Piotr Majdak. Personalization support for binaural headphone reproduction in web browsers. In *Proceedings of the 1st Web Audio Conference*, Paris, 2015.