



Audio Engineering Society

Convention Paper 10501

Presented at the 150th Convention, Online
2021 May 25–28

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (<http://www.aes.org/e-lib>) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Comparison of Distortion Products in Headphone Equalization Algorithms for Binaural Synthesis

Braxton Boren¹ and Michele Geronazzo^{2,3}

¹Audio Technology Program, American University, Washington, DC

²Dept. of Humanities and Cultural Heritage, University of Udine, Italy

³Dyson School of Design Engineering, Imperial College London, London, UK

Correspondence should be addressed to Braxton Boren (boren@american.edu)

ABSTRACT

Headphone design has traditionally focused on creating a frequency response to make commercial stereo audio sound more natural. However, because of the sensitivity of spatial hearing to frequency-dependent cues, binaural reproduction requires headphones' target spectrum to be as flat as possible. Initial attempts to equalize headphones used a naive inversion of the headphone spectrum, which degraded binaural content because the headphone transfer function (HpTF) changes each time headphones are re-seated. Many different algorithms have been proposed to improve binaural equalization, each of which has been tested over a limited sample of HpTFs. The present study gathered 1550 HpTFs from different institutions into a single dataset for large-scale comparisons of equalization algorithms. A numerical metric was designed to quantify auditory perception of spectral coloration from 'ringing' peaks in the post-equalization HpTF. Using this metric, eight of the most prominent equalization methods have been compared over the aggregate HpTF dataset. High-shelf regularization is shown to outperform all other equalization techniques using either individualized or averaged input spectra. In addition, high-shelf regularization without individual measurements gives less average coloration than direct inversion using individualized equalization.

1 Introduction

For commercial stereo audio, it is not desirable to achieve a completely flat headphone frequency response [1, 2, 3], but for binaural rendering a transparent headphone response is important as a flat response presents a neutral starting point for accurate rendering of the head-related transfer function (HRTF) of a virtual auditory event. Previous work suggests that flatter frequency responses may contribute to improvements in sound externalization [4], auditory distance perception [5, 6], and sound localization [7, 8].

The headphone transfer function (HpTF) consists of

low-frequency resonance effects primarily from the headphone cups and the high-frequency resonances of the listener's pinnae [9, 10, 11]. Early research into HpTFs assumed that the headphone response could be completely removed via a direct inversion filter [12, 13], given by

$$H^{-1}(\omega) = \frac{1}{H(\omega)}. \quad (1)$$

However, in practice the refitting of headphones by the same listener leads to significant geometric changes at

small wavelengths, leading to shifts in the HpTF's high-frequency notches [14, 15, 16].¹ These differences can be immediately audible [18], but directly inverting the old measured response at these frequencies leads to large peaks in the inverse filter which introduce dynamic range loss but no longer equalize the measured notches.² In addition, these peaks are much more noticeable than the notches they are intended to equalize [20, 21]. For these reasons, more recent headphone equalization approaches use some form of averaging, discrimination, or regularization to reduce gain at high frequencies [22, 23, 10, 24].

Pralong and Carlile reported large HpTF differences between subjects above 4 kHz, leading them to recommend individualized headphone measurements [25]. However, 4 kHz also marks the beginning of high-frequency HpTF variations introduced by re-seating the same headphones for the same subject [22, 10]. Thus inter-subject and intra-subject variations are highly intertwined, making it difficult to account for one and not the other. Since many of the headphone equalization algorithms mentioned above reduce their gain at high frequencies, they may also reduce some of the benefits from individualized equalization. Since an averaged filter only accounts for the resonance of the headphones themselves and not individual anthropometry [23, 24], it is an open question whether individualized measurements are necessary if the primary goal is to reduce the presence of 'ringing' distortion peaks at the listener's eardrums [26].

2 Background

2.1 Existing Headphone Equalization Algorithms

The headphone equalization algorithms in use today use a variety of frequency discrimination methods to achieve a natural listening condition. It is possible to hand-tune equalizing peaks and notches to a single fitting of headphones to achieve good results [27], but because of the time-consuming nature of this approach, most equalizations try to achieve a good inverse filter which will not require additional tuning by the user.

¹Some variability within headphone transducers is also present, but these are much smaller than the variation within the measured HpTF [17].

²In-ear headphones display smaller intra-subject variability if they achieve a complete seal [19], but they also induce greater listener fatigue, making them less desirable for many binaural applications.

These techniques tend to look for different ways to minimize the most extreme peaks in the frequency response of the inverse filter.

In particular, these include a frequency-domain peak compression algorithm [22], a statistical approach inverting the 95th percentile of each frequency bin's magnitude for a set of HpTF measurements [10], and a variety of frequency regularization methods, with fixed or adaptable parameters [23, 28]. While all of these algorithms have the same goal (i.e. the reduction of large high frequency peaks), their approaches vary, and in general their parameters have been hand-tuned using informal listening tests over small databases of HpTFs measured at different laboratories.

Using the Spatially Oriented Format for Acoustics (SOFA) [29], we have created a new data format for HpTFs and consolidated many of the existing HpTF databases into a single publicly available dataset with 1550 HpTFs [30]. This allows the large-scale comparison of different algorithms and input parameters under a variety of different conditions.

2.2 Peak Error Metric

Since the reduction of 'ringing' distortion peaks is the stated objective of all the different equalization algorithms listed, it makes sense to rank these algorithms based on the amount of such peaks in a large number of HpTFs after equalization using each method [26]. 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 list of peaks, peaks smaller than a threshold are removed to account for signal noise. We use a threshold value of 1 dB to find perceptually significant peaks, similar to [24]. The peaks are sorted in order of decreasing height, and smaller peaks within $\pm 1/6$ 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 1, this increases the numerical contribution of narrow peaks by emphasizing their difference with the spectrum immediately around them.

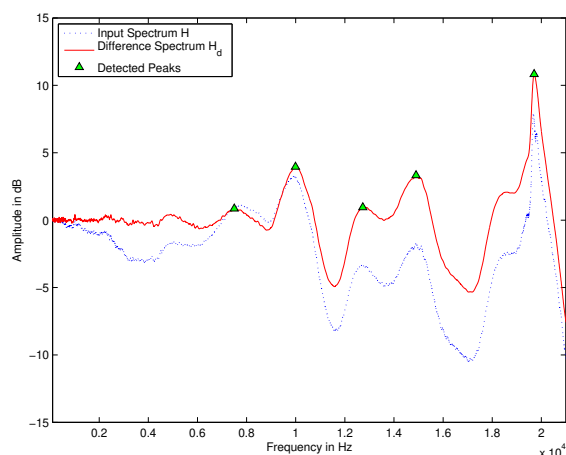


Fig. 1: Example post-equalization difference spectrum and detected peaks.

From these selected peaks, we remove those for which the original input spectrum H is negative (0 dB is defined here as the average level between 100 and 4000 Hz). This leaves a vector of peak locations p , which is used to calculate the peak error metric E_{pk+} by

$$E_{pk+} = \frac{\sum_p H(p)}{3[\log_2(f_{high}/f_{low})]}, \quad (2)$$

where the denominator represents the number of third octave bands within the frequency range of the spectrum.³ Using values of $f_{low} = 50$ and $f_{high} = 21000$, this amounts to a division by about 26. This scales most E_{pk+} values from 0-1, where 0 represents zero peak distortion, and 1 dB per third octave is representative of the most distorted equalization spectra.

3 Cross-Comparison for Single Subjects

Using this metric, we can begin to visualize the effect of naive inversion on a set of HpTF measurements for a single listener. Figure 2 shows the E_{pk+} resulting from equalizing 20 individual measured HpTFs on the

³Calculating the value based on H ensures the contribution of each peak is continuous, whereas because of thresholding the values each of $H_d(p)$ could jump from 0 to 1 dB. However, because of the near-peak rejection within ± 1 dB, this function is not continuous as a whole, which is important to note when optimizing HpTF filters numerically.

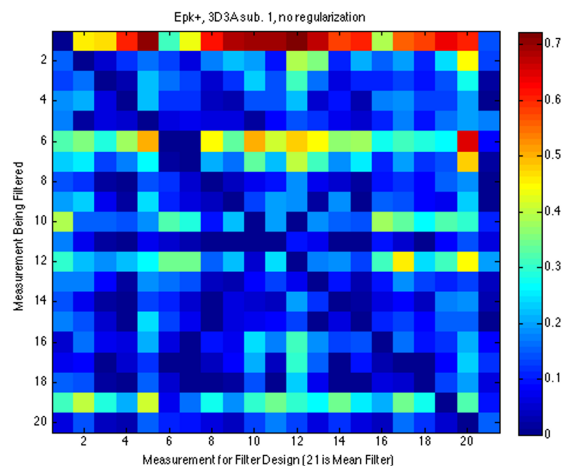


Fig. 2: E_{pk+} for 20 HpTF measurements of the same subject, applied to every other measurement.

same subject by naive inversion of each of the other measurements. Each row of the matrix represents an HpTF (1-20) being filtered, while each column shows the measurement used to produce the inverse filter. Column 21 shows the results of using an averaged filter based on all 20 measurements instead.

From this and similar plots for different subjects in the PHOnA dataset, we can observe several tendencies:

- All E_{pk+} values on the main diagonal are 0, since each measurement's naive inversion gives perfect cancellation with itself.
- Some measurements are easier to filter than others: Rows 1, 6, and 19 for this subject have much higher E_{pk+} than others for most filters applied. Conversely, some rows have fairly low E_{pk+} regardless of the filter applied.
- Some measurements create a better filters than others. For instance, measurement 6, despite its high E_{pk+} value when filtered, creates a lower than usual E_{pk+} when applied to other measurements. This is perhaps due to the inversion process, as certain HpTFs with prominent peaks will tend to create very aggressive filters once their frequency response is inverted.
- The mean filter (Column 21), based on an averaged magnitude spectrum across all measurements, outperforms any individual inverse filter.

Because of the superiority of mean filters, many early efforts emphasized taking many measurements of subject's HpTFs to get the best generic spectrum of the headphones independent of intra-subject variation. However, when a regularized filter was applied to this same subject, all E_{pk+} values in this matrix became 0 (i.e., no peaks above 1 dB were detected in the difference spectrum), though this was not the case for every subject in the database. Thus individual measurements with good regularization may sometimes reduce distortion peaks as well as methods based on an averaged filter.

4 Methods

After having considered cross-comparisons of this sort for different subjects, we measured the E_{pk+} resulting from different proposed equalization techniques across the PHOnA dataset. In addition, for each we compared the result of an averaged filter across all subjects for the same set of headphones, as well as an individualized filter which was based on the averaged spectrum for a single individual with a single set of headphones. This mean filter was then passed as the input to each of 8 different equalization algorithms:

1. LMS High-Shelf Regularized Inversion [24]
2. Frequency-Domain Peak Compression [22]
3. LMS Regularization Based on Input Spectrum [24]
4. "Compare and Squeeze" [31], similar to #3
5. Sigma Inversion [28]
6. LMS Inversion Without Regularization [23]
7. Direct Inversion Only [32]
8. 95th Percentile Inversion⁴ [10]

Here, algorithms #1 and 2 both seek to reduce equalization peaks above a set value (4 kHz in both cases). Algorithms #3-5 apply some form of regularization, but using a smoothed version of the measured HpTF

⁴The last algorithm actually required a set of multiple measurements, out of which it inverted the 95th percentile magnitude at each frequency bin. Thus this algorithm was passed multiple measurements, either all the measurements for a set of headphones, or all the measurements for a single individual with a single set of headphones.

as a target function instead. Algorithms #6-7 apply no regularization, and algorithm #8 finds a target spectrum based on the variance across many measurements.

The direct inversion was intended as a benchmark only, since direct inversion even with a mean filter can still produce large distortion peaks. All algorithms were implemented based either on the original authors' description or using average parameters built in the Matlab functions sent by the authors of the equalization methods. Thus no calibrations nor adjustments of the algorithms were performed since the goal was a comparison of a single technique over a large number of HpTFs measured on different subjects, headphones, and microphones.

Each measurement in the PHOnA dataset was then equalized using both an averaged filter, taking the mean across all measurements of the same headphones, and an individualized filter, taking the mean across all measurements of each subject for a single set of headphones. The E_{pk+} values for all filtered measurements were then averaged into a final E_{pk+} value for each case (individualized/averaged) and each algorithm (1-8). Statistical evaluation was performed on E_{pk+} values by a two-way ANOVA on input filter and headphone equalization method as factors. Post-hoc analysis with Bonferroni correction on p-values were provided.

5 Results

The two-way ANOVA revealed a statistically significant interaction between input filter and equalization algorithm for both channels: left [$F(7,26096) = 59.3$, $p \ll .001$] and right [$F(7,26096) = 67.6$, $p \ll .001$]. The resulting averaged E_{pk+} values for each algorithm (#1-8, ordered left to right) and input type are displayed in Figures 3 and 4 for the left and right channels, respectively. It can be seen that every algorithm performs significantly better with input individualized to the listener than with only the averaged data for the headphones being used (all comparisons exhibited $p \ll .001$).

Looking closer at the results, algorithm #1 (LMS High-Shelf Regularization) clearly outperformed all the others, both in the individualized or averaged case with statistical significance. When individualized input was available, this algorithm left less than 0.1 dB/third octave in the equalized spectrum, by far the least distortion of any of the techniques investigated. Moreover,

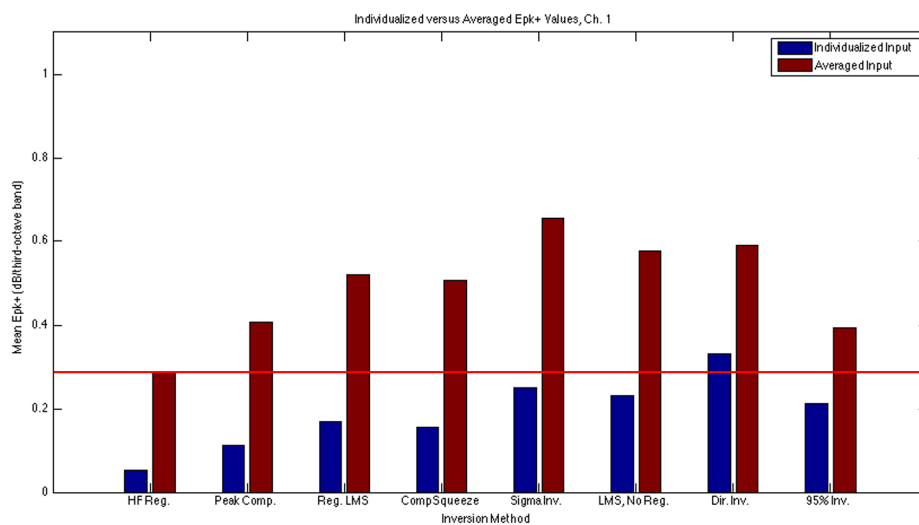


Fig. 3: Averaged E_{pk+} values for all algorithms, using both individualized and averaged input, left channel.

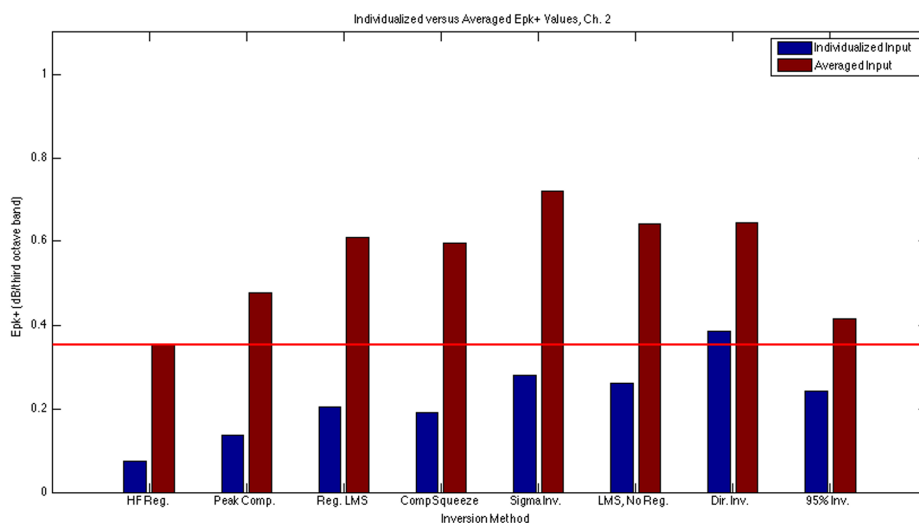


Fig. 4: Averaged E_{pk+} values for all algorithms, using both individualized and averaged input, right channel.

the two-way ANOVA reported no statistically significant differences between algorithms #3 and #4 both in the individualized or averaged case, $p > 0.99$.

However, the LMS High-Shelf Regularization also reduced E_{pk+} to between 0.3-0.4 dB/third octave when using an averaged input across headphones only. The red lines on Figures 3 and 4 show the mean E_{pk+} for

algorithm #1 when using a non-individualized input, and it can be seen that the averaged input for algorithm #1 outperforms the individualized results of direct inversion for both the left and right channels with a statistically significant difference, $p \ll .001$.

6 Discussion

In evaluating the final results of the comparison, it is useful to consider the goals of each algorithm being investigated: algorithms #3-5 all consider the shape of the input spectrum as a basis for their own regularization parameters, resulting in a less drastic equalization, while algorithms #1-2 apply a more uniform criteria at all frequencies above 4 kHz. Algorithms #6-7 apply no regularization (algorithm #8 is in a class by itself as it is based only on the variance across measurements, which will tend to have a regularizing effect at high frequencies since they experience the most difference between measurements). In this sense algorithms #3-5 attempt to be more ‘organic’ in their approach while algorithms #1-2 are a bit more ‘ruthless,’ and the ruthless approach clearly results in lower distortion. These differences were also detected from the ANOVA statistical interaction between input filter and equalization algorithm.

However, distortion is not the only goal of a good HpTF equalization: it is a well-known problem that such regularized approaches, which reduce their gain drastically at high frequencies, end up having a net lowpass effect, as the natural notches of the physical HpTF combine with the notches (but no peaks) of the equalization filter (Fig. 5). It is possible that if we understood the variance in high-frequency peaks better, it might be possible to design new approaches to binaural headphone equalization which combine the peak-avoidance of the ‘ruthless’ algorithms with the better high-frequency content of the ‘organic’ algorithms.

7 Conclusion and Future Work

This study merely compared distortion products in existing algorithms. However, there are many possibilities for improving the state of the art based on large-scale data analysis. First, we would like to apply machine learning techniques to individual algorithms and their output spectra, using hyperparameter optimization techniques for adjusting many input parameters on individual algorithms to reduce distortion for a single headphone. Since the E_{pk+} metric is not continuous, grid search techniques will be more useful than gradient descent, which assumes a continuous multidimensional function.

In addition, since even the ‘ruthless’ algorithms still performed significantly better with individualized data,

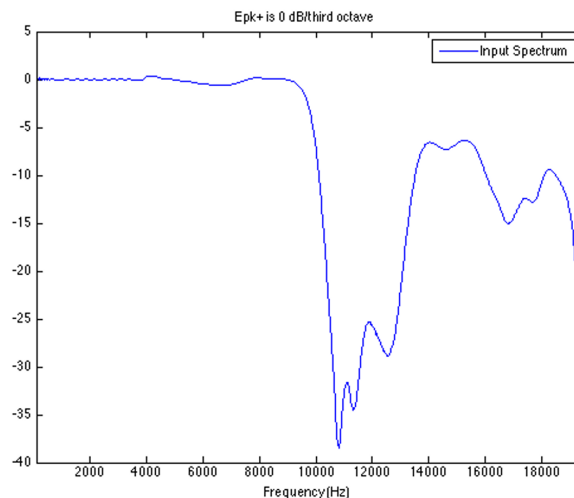


Fig. 5: HpTF filtered using algorithm #1 with $E_{pk+} = 0$, but with a net lowpass effect

it is worth further investigating the previously stated model of “headphone cups affect low frequencies, while individual differences affect high frequencies.” While that statement is obviously true, it may be incomplete: re-seating of the same headphones results in minute changes to the HpTF spectra at very high frequencies, while the changes resulting from individuals’ pinnae differences involve slightly larger wavelengths and may occur at slightly lower frequencies on average, which would explain why the high-shelf regularized approaches still performed significantly better with individualized input data. To investigate this, we plan to perform clustering analysis over peak locations in the final spectra using both individualized and non-individualized filters, to provide better data regarding this point.

Finally, the application of large amounts of HpTF data might be able to be combined with the HRTFs of listeners in situations where the HpTF and HRTF contain similar information. Building off the work of Kelly and Boland [33], if a better database connecting users’ HRTFs and HpTFs could be constructed, it would be possible to create binaural renderings which build off the naturally-occurring notches in the HpTF rather than trying (unsuccessfully) to fill in existing notches and then superimpose new ones in the binaural rendering.

References

- [1] Moller, H., Jensen, C., Hammershoi, D., and Sorensen, M., "Design Criteria for Headphones," *Journal of the Audio Engineering Society*, 43(4), pp. 218–232, 1995.
- [2] Olive, S. E., Welti, T., and McMullin, E., "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] Olive, S. E., Welti, T., and Khonsaripour, O., "A Statistical Model That Predicts Listeners' Preference Ratings of In-Ear Headphones: Part 2 – Development and Validation of the Model," in *Proceedings of the 143rd Audio Engineering Society Convention*, New York, NY, 2017.
- [4] Boren, B. B. and Roginska, A., "The Effects of Headphones on Listener HRTF Preference," in *Proceedings of the 131st Audio Engineering Society Convention*, New York, NY, 2011.
- [5] Sunder, K., Tan, E.-L., and Gan, W.-S., "Effect of Headphone Equalization on Auditory Distance Perception," in *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, 2014.
- [6] Engel, I., Alon, D. L., Robinson, P. W., and Mehra, R., "The effect of generic headphone compensation on binaural renderings," in *AES International Conference on Immersive and Interactive Audio*, York, UK, 2019.
- [7] Schonstein, D., Ferr, L., and Katz, B. F. G., "Comparison of headphones and equalization for virtual auditory source localization," in *Acoustics '08*, pp. 4617–4622, Paris, 2008.
- [8] Satongar, D., Member, A. E. S. S., and Pike, C., "The Influence of Headphones on the Localization," *Journal of the Audio Engineering Society*, 63(10), pp. 799–810, 2015.
- [9] Moller, H., Hammershoi, D., Jensen, C., and Sorensen, M., "Transfer Characteristics of Headphones Measured on Human Ears," *Journal of the Audio Engineering Society*, 43(4), pp. 203–217, 1995.
- [10] Masiero, B. and Fels, J., "Perceptually Robust Headphone Equalization for Binaural Reproduction," in *Proceedings of the 130th Audio Engineering Convention*, London, UK, 2011.
- [11] Takemoto, H., Mokhtari, P., Kato, H., Nishimura, R., and Iida, K., "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), pp. 3832–41, 2012, ISSN 1520-8524, doi:10.1121/1.4765083.
- [12] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. I: Stimulus synthesis," *The Journal of the Acoustical Society of America*, 85(2), pp. 858–67, 1989, ISSN 0001-4966.
- [13] Moller, H., "Fundamentals of binaural technology," *Applied Acoustics*, 36, pp. 171–218, 1992.
- [14] Toole, F. E., "The Acoustics and Psychoacoustics of Headphones," in *Proceedings of the 2nd Audio Engineering Society International Conference*, Anaheim, CA, 1984.
- [15] Kulkarni, A. and Colburn, H. S., "Variability in the characterization of the headphone transfer-function," *Journal of the Acoustical Society of America*, 107(2), pp. 1071–1074, 2000.
- [16] Volk, F., "Inter- and Intra-Individual Variability in the Blocked Auditory Canal Transfer Functions of Three Circum-Aural Headphones," *Journal of the Audio Engineering Society*, 62(5), pp. 315–323, 2014.
- [17] Volk, F., Encke, J., Kreh, J., and Hemmert, W., "Relevance of Headphone Characteristics in Binaural Listening Experiments: A Case Study," in *Proceedings of the 143rd Audio Engineering Society Convention*, New York, NY, 2017.
- [18] Paquier, M. and Koehl, V., "Discriminability of the placement of supra-aural and circumaural headphones," *Applied Acoustics*, 93, pp. 130–139, 2015.
- [19] Olive, S. E., Welti, T., and Khonsaripour, O., "A Statistical Model That Predicts Listeners' Preference Ratings of In-Ear Headphones: Part 1 – Listening Test Results and Acoustic Measurements," in *Proceedings of the 143rd Audio Engineering Society Convention*, New York, NY, 2017.

- [20] Bucklein, R., "The Audibility of Frequency Response Irregularities," *Journal of the Audio Engineering Society*, 29(3), pp. 126–131, 1981.
- [21] Moore, B. C., Oldfield, S. R., and Dooley, G. J., "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz." *The Journal of the Acoustical Society of America*, 85(2), pp. 820–36, 1989, ISSN 0001-4966.
- [22] Hiekkänen, T., Makivirta, A., and Karjalainen, M., "Virtualized Listening Tests for Loudspeakers," *Journal of the Audio Engineering Society*, 57(4), pp. 237–251, 2009.
- [23] Lindau, A. and Brinkmann, F., "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings," *Journal of the Audio Engineering Society*, 60(1), pp. 54–62, 2012.
- [24] Scharer, Z. and Lindau, A., "Evaluation of Equalization Methods for Binaural Signals," in *Proceedings of the 126th Audio Engineering Society Convention*, Munich, Germany, 2009.
- [25] Pralong, D. and Carlile, S., "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), pp. 3785–93, 1996, ISSN 0001-4966.
- [26] Boren, B. B., Geronazzo, M., Brinkmann, F., and Choueiri, E., "Coloration Metrics for Headphone Equalization," in *The 21st International Conference on Auditory Display (ICAD-2015)*, Graz, Austria, 2015.
- [27] Griesinger, D., "Accurate reproduction of binaural recordings through individual headphone equalization and time domain crosstalk cancellation," in *Proceedings of the 147th Audio Engineering Society Convention*, New York, NY, 2019.
- [28] Gomez-Bolaños, J., Mäkiavirta, A., and Pulkki, V., "Automatic regularization parameter for headphone transfer function inversion," *Journal of the Audio Engineering Society*, 64(10), pp. 752–761, 2016, ISSN 15494950, doi:10.17743/jaes.2016.0030.
- [29] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M., "Spatially Oriented Format for Acoustics," in *Proceedings of the 134th Audio Engineering Society Convention*, Rome, Italy, 2013.
- [30] Boren, B. B., Geronazzo, M., Majdak, P., and Choueiri, E., "PHOnA : A Public Dataset of Measured Headphone Transfer Functions," in *Proceedings of the 137th Audio Engineering Society Convention*, Los Angeles, CA, 2014.
- [31] Muller, S., *Digitale Signalverarbeitung für Lautsprecher*, Ph.D. thesis, RTWH Aachen, 1999.
- [32] Wightman, F. L. and Kistler, D. J., "Headphone simulation of free-field listening. II: Psychophysical validation." *The Journal of the Acoustical Society of America*, 85(2), pp. 868–78, 1989, ISSN 0001-4966.
- [33] Kelly, I. J. and Boland, F. M., "Exploiting randomness in acoustic impulse responses to achieve headphone compensation through deconvolution." *The Journal of the Acoustical Society of America*, 133(5), pp. 2778–87, 2013, ISSN 1520-8524, doi:10.1121/1.4798799.