Audio Engineering Society

Convention Paper 9126

Presented at the 137th Convention
2014 October 9–12 Los Angeles, USA

PHOnA: A Public Dataset of Measured Headphone Transfer Functions

Braxton Boren[1], Michele Geronazzo[2], Piotr Majdak[3], and Edgar Choueiri[1]

[1] *3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ, 08544, USA*

[2] *Dept. of Information Engineering, University of Padova, Padova, 35131, Italy*

[3] *Acoustics Research Institute, Austrian Academy of Sciences, Vienna, 1040, Austria*

Correspondence should be addressed to Braxton Boren (bboren@princeton.edu)

ABSTRACT

A dataset of measured headphone transfer functions (HpTFs), the Princeton Headphone Open Archive (PHOnA), is presented. Extensive studies of HpTFs have been conducted for the past twenty years, each requiring a separate set of measurements, but this data has not yet been publicly shared. PHOnA aggregates HpTFs from different laboratories, including measurements for multiple different headphones, subjects, and repositionings of headphones for each subject. The dataset uses the spatially oriented format for acoustics (SOFA), and SOFA conventions are proposed for efficiently storing HpTFs. PHOnA is intended to provide a foundation for machine learning techniques applied to HpTF equalization. This shared data will allow optimization of equalization algorithms to provide more universal solutions to perceptually transparent headphone reproduction.

1. INTRODUCTION

While in general it is not optimal for headphones to have a completely flat frequency response, because of the spectral sensitivity of auditory localization a flat response is important for 3D audio applications [1, 2]. With a static fit against the listener's ears, it is theoretically possible to completely remove the spectrum of the headphones via inverse filtering [3, 4].

Practically speaking, however, tiny geometrical changes between headphones and listener result in resonance differences that in turn lead to significant changes in the high frequencies of the headphone transfer function (HpTF) [6]. For low fre-

Table 1: HpTF Databases contained in PHOnA

Institution	Subjects	Repositionings	Headphones
ARI Vienna	120	5	1
DSTO Australia [5]	3	20	1
ITA Aachen [6]	15	8	2
Princeton University 3D3A Lab	4	20	1
TU Berlin [7]	95	10	1
University of Padova [8]	18	10	3

quencies, inter-subject variability is limited up to $\approx$ 4 kHz because at these frequencies headphones act as an acoustic cavity only introducing a constant level variation [6]. On the contrary, in the higher spectrum, headphone position and the listener's anthropometry may give rise to significant spectral distortions due to two reasons:

- standing wave buildup inside the headphone cups
- the outer ear's resonances that yield an individual spectrum for each HpIR

Because of the large inter-subject variations in HpTFs, individual headphone equalization is recommended [9], but given the magnitude of intra-subject variations, individual equalization can still result in significant spectral distortions [10]. In some cases it has been shown that a naive HpTF inversion may actually make it more difficult for subjects to localize sounds in 3D space [11].

Several researchers over the last 20 years have examined the variability in HpTF measurement by conducting extensive studies of inter- and intra-subject variation [12, 9, 5, 13, 14, 6, 7], but so far each study has required the time-consuming process of measuring multiple HpTFs. The availability of a large amount of individual measurements can accelerate the development of headphone correction procedures that are listener-specific [12, 9] and robust to headphone placements [10]. All these elements are important for the development of authentic high-fidelity 3D audio systems.

2. DATA AGGREGATION

2.1. Gathering Existing Data

Since HpTFs exhibit an overall variation level on a similar scale as HRTFs, it is instructive to observe the trend of publicly available HRTF datasets released over the past 15 years.

Early measured HRTF datasets, which require specialized equipment in an anechoic environment, were recorded and released by individual labs [15, 16]. This criterion is less applicable to HpTF measurements, whose fixed costs are much lower. However, in recent years there has been a larger push for aggregating many existing HRTF databases into a single publicly available repository, allowing machine learning techniques to be applied to a larger dataset [17]. It has been proposed by Geronazzo *et al.* [8] that this same aggregation should also be applied to existing HpTF databases.

With this motivation in mind, we have created the PHOnA dataset (Princeton Headphone Open Archive) to make the large amounts of existing HpTF data publicly available to researchers worldwide. So far we have compiled six individual databases into a single dataset, as shown in table 1. The dataset contains HpTFs from ARI Vienna, DSTO Australia, ITA Aachen, TU Berlin, University of Padova, and Princeton University. Each lab's data are organized based on the number of subjects measured, number of intra-subject measurements, number of headphones measured, and whether microphone responses are available to equalize the responses. The full dataset is available from the Princeton 3D and Applied Acoustics Lab's website.*

*http://www.princeton.edu/3D3A/Phona.html

2.2. Heterogeneity

The most common methods employed in the acquisition of HpIRs involve the use of an artificial ear, such as the B&K 4153; a dummy head, such as a KEMAR mannequin; or human listeners. But using an artificial ear neglects the acoustic effects of individuals' pinnae, which is an important component for perceptually transparent headphone listening. Individual recordings are preferable because they measure intra-subject variations from multiple positionings of headphones for the listener.

The data contained in the PHOnA dataset are quite heterogeneous, with different labs using different headphones, audio I/O equipment, and settings. As the algorithms proposed for HpTF equalization have so far been based on the data available to a single institution, it is possible for a proposed technique to be overfit to the data available. The release of a larger dataset makes it possible to compare equalization methods from one lab as applied to all the data available, which should allow for more universal techniques.

For instance, a perceptually relevant equalization algorithm was developed [6] to preferentially produce notches rather than peaks in the final spectrum produced by headphones after equalization, since peaks are more noticeable. This can be seen by looking at measurements of multiple fittings of a pair of headphones on a single listener. The aggregate spectra can be used to design a filter for maximum transparency on any given measurement from the set.

In figure 1(a) we show a left ear HpTF from the dataset, obtained by three methods: mean equalization from 20 measurements, perceptual equalization, and no equalization. It is clear that both equalization techniques drastically improved the final spectrum with respect to the unequalized version, especially at frequencies below 4 kHz. However, the mean-equalized spectrum contains a 10 dB peak that will most likely be heard as a high-frequency ringing. The perceptual equalization removed that peak at the expense of adding notches, which, on the other hand will probably be less noticeable [18].

In figure 1(b) we have another measurement for the same listener, but with a different fitting of the headphones. Again, both equalization algorithms give a much flatter spectrum than the unequalized response. As compared to the first measurement, the peaks are less pronounced in the mean-equalized spectrum. However, the notches are about the same in the perceptually equalized spectrum, and are probably more noticeable than the mean equalized response in this case [18].

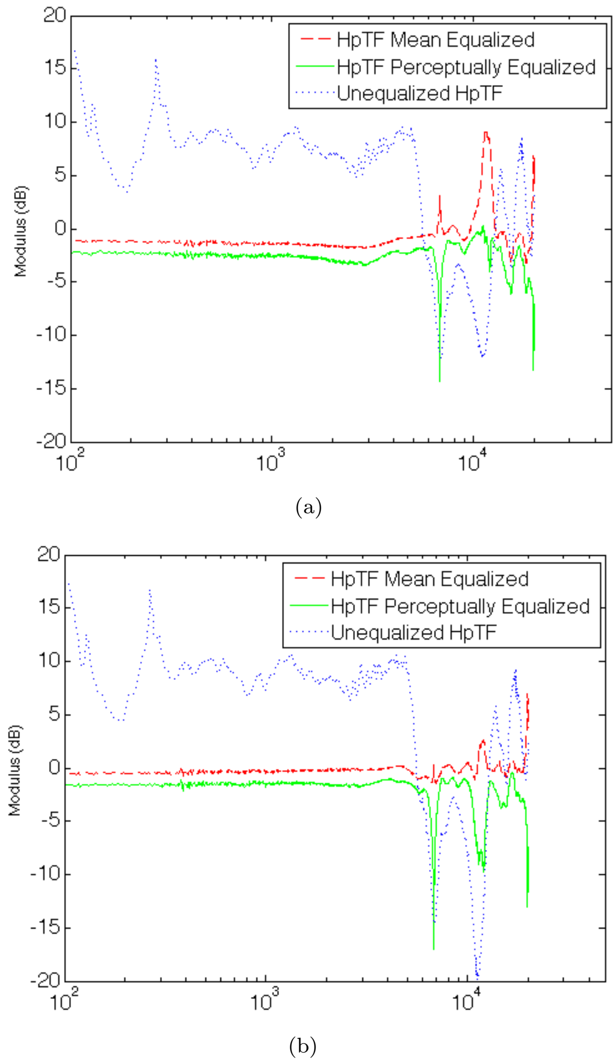

Fig. 1: Two HpTFs for the same listener obtained using mean equalization, perceptual equalization, and no equalization

There are significant variations across many different dimensions related to setups and subjects. Thus a uniform container is required for the systematic application of quantitative measures of equalization performance for different algorithms.

3. STANDARDIZATION PROCESS

We chose to organize the files in PHOnA in the spatially oriented format for acoustics (SOFA) [19]. One of the main motivations behind the SOFA project is to facilitate the personalization of binaural rendering, which is clearly in line with the PHOnA project.

We have taken into account various aspects of each HpTF database in order to propose headphone conventions for SOFA, i.e. definitions of data and metadata consistently describing particular HpTF measurement setups.

3.1. The SOFA

The aim of SOFA is the representation of spatially oriented data in a general way in order to promote interchangeability and extendability. The AES-X212 HRTF file format standardization project* is based on the SOFA format and is currently under review for approval by the AES standardization committee. SOFA files have .sofa extension and data are serialized into a binary stream with netCDF-4 (Unidata)† software libraries which ensure requirements such as open access, efficient data compression, self description, network transparency, and machine independence.

In SOFA, a measurement setup with arbitrary geometry is described by the relation between the objects *listener* and *source*, which are positioned inside the global coordinate system of the object *room*. Acoustic sensors, i.e. microphones, are defined by the object *receiver*, which are related to the receiver. On the other hand, acoustic excitation sources, i.e. loudspeaker drivers, or individual loudspeakers in a speaker array, are defined by the object *emitter*. In SOFA, receivers and emitters are described in a local coordinate system relative to those of the listener and source, respectively.

The most common measurement setups are provided in the form of predefined description conventions. Conventions *SimpleFreeFieldHRIR* defines a set of recommendations for a typical HRTF measurement setup in an assumed free field. It is one of the few well-defined SOFA conventions, together with *GeneralFIR* and *GeneralTF*, two general specifications. Further SOFA conventions have been recently proposed and are under evaluation: *MultiSpeakerBRIR*, which defines binaural room impulse responses (BRIRs) measured with an arbitrary number of receivers and emitters in an echoic single room.

3.2. HpTF measurement of a single listener

In this study, we propose SOFA conventions for storing headphone impulse responses (HpIRs) of a single listener, namely, *SingleHeadphoneIR*.

The HpTF measurement setup is different from the HRTF measurement setup. First, the emitter (i.e. the headphones) inevitably perturbs the acoustic waves before they reach the receivers (i.e. the microphones) that are physically connected to the resonating body of the listener (external ear and ear canal), shaping a new geometry. As a consequence, the emitter encloses the acoustic meatus and becomes itself part of a unique resonating object, which shares some acoustic properties of the listener, based on how the headphones are positioned. Thus, a repeated measurement will probably yield significant differences in the HpIR. Second, there is a one-to-one correspondence between an emitter and a receiver: the excited emitter produces the relevant signal in the corresponding receiver and the crosstalk to all other receivers is negligible. These properties have been taken into account in the SOFA convention *SingleHeadphoneIR*.

Furthermore, we have considered metadata specific for headphone measurements. For example, the emitter's placement and measuring conditions (e.g. open- vs. blocked-entrance ear canal measurements) depend strongly on headphone type (e.g. circumaural or supraaural headphones, earphones, insert earphones, bone conducted headset and assistive hearing devices – see [20] for an exhaustive review on headphone design).

The measured HpIRs are represented as FIRs, with a single HpIR set of a listener per file; thus, the number of receivers (i.e., microphones) defines the

*http://www.aes.org/standards/meetings/init-projects/aes-x212-init.cfm

†http://www.unidata.ucar.edu/software/netcdf/

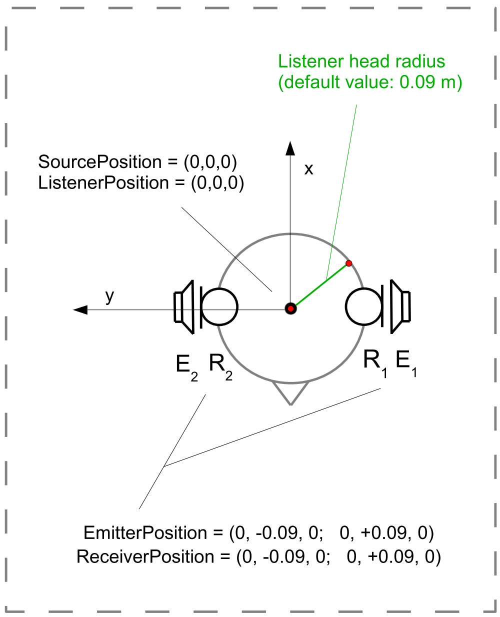

Fig. 2: HpTF measurement setup considered in the SOFA convention *SingleHeadphoneIR*

number of emitters, with a strict one-to-one correspondence between emitters and receivers. Each measurement in the file is considered as a repetition of the first measurement under modified conditions.

In the first proposed version of *SingleHeadphoneIR*, the following data and metadata are proposed:

- *General attributes*:
  - *SOFAConventions*: SingleHeadphoneIR
  - *SOFAConventionsVersion*: 0.1
  - *Datatype*: FIR
  - *Roomtype*: free field
  - *SourceManufacturer*: the name of the manufacturer of the headphones. Optional.
  - *SourceModel*: the model name of the headphones given by the manufacturer. Optional.
  - *EmitterFormFactor*: refers to the headphone form factor (e.g. circumaural or supra-aural). Optional.
  - *EmitterDesign*: refers to the ear-cup design (e.g. open or closed). Optional.
  - *EmitterSensitivity*: refers to the electro-acoustic transducer sensitivity (transfer factor) in mV/Pa. Optional.
  - Other as defined by the general SOFA specifications.
- *R*: Arbitrary number of receivers, with a default value of 2;
- *E*: Defines the number of emitters and must be the same as R;
- *M*: Arbitrary number of measurements available for that listener, with a default value of 1;
- *ListenerPosition*: The position of the listener, with a default of (0 0 0);
- *SourcePosition*: The position of the headphones, which are usually are located at the same position as the listener, thus, the default is (0 0 0);
- *ReceiverPosition*: Defines the position of the microphones (in meters), with a default of (0 -0.09 0; 0 +0.09 0).
- *EmitterPosition*: Defines the position of the individual headphone drivers, which are usually located near to the microphones, i.e., receivers. The default is the same position as that of the receivers: (0 -0.09 0; 0 +0.09 0).
- *MeasurementDate*: vector of size M, defines the date and time of the particular measurement as the number of seconds from 1970-01-01 00:00:00.

4. CONCLUSIONS AND FUTURE WORK

We have presented the PHOnA dataset with the goal of providing high-quality HpTFs to research laboratories working in headphone reproduction and 3D audio. The shared data can be used to evaluate the extent and significance of spectral variations due to individuals' pinnae, headphone selection, and different fittings of headphones. This large dataset, based on measurements from many different institutions, aims at providing maximum generality and reduced likelihood of overfitting an equalization technique to a small sample of HpTFs. The proposed SOFA convention, *SingleHeadphoneIR*, will improve portability and enhance data sharing across different institutions. The existence of an open HpTF dataset will allow better quantitative evaluation of existing equalization methods.

For future work, an objective metric will be developed to quantify the perceptual coloration in equalized spectra for existing equalization methods. Such a coloration index, if validated in perceptual tests, could be used as a cost function to train machine learning algorithms to develop optimized equalization filters for headphone listening. This will allow the development of more robust equalization in cases where individual HpTF measurements are available for designing inverse filters. It will also allow an investigation of optimal methods for non-individualized equalization techniques for a given model of headphones. The PHOnA dataset is intended to encourage the broadest application of 3D sound to the largest part of the population: both high-end consumers who want the best fidelity, and late adopters who have never experienced 3D audio before.

5. 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.

6. REFERENCES

[1] H. Moller, C. Jensen, D. Hammershoi, and M. Sorensen. Design Criteria for Headphones. *Journal of the Audio Engineering Society*, 43(4):218–232, 1995.

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

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

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

[5] K. Mcanally and R. Martin. Variability in the Headphone-to-Ear-Canal Transfer Function. *Journal of the Audio Engineering Society*, 50(4):263–266, 2002.

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

[7] A. Lindau and F. 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.

[8] M. Geronazzo, F. Granza, S. Spagnol, and F. Avanzini. A standardized repository of Head-Related and Headphone Impulse Response data. In *Proceedings of the 134th Audio Engineering Society Convention*, Rome, Italy, 2013.

[9] D. Pralong and S. 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.

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

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

[12] H. Moller, D. Hammershoi, C. Jensen, and M. Sorensen. Transfer Characteristics of Headphones Measured on Human Ears. *Journal of the Audio Engineering Society*, 43(4):203–217, 1995.

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

[14] M. Paquier and V. Koehl. Audibility of headphone positioning variability. In *Proceedings of the 128th Audio Engineering Society Convention*, London, 2010.

[15] V. Algazi, R. Duda, D. Thompson, and C. Avendano. The CIPIC HRTF database. In *2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics*, number October, pages 99–102, New Paltz, NY, 2001.

[16] N. Gupta, A. Barreto, M. Joshi, and J. Agudelo. HRTF Database at FIU DSP Lab. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 169–172, Dallas, TX, 2010.

[17] A. Andreopoulou and A. Roginska. Towards the Creation of a Standardized HRTF Repository. In *Proceedings of the 131st Audio Engineering Society Convention*, New York, NY, 2011.

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

[19] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, and H. Ziegelwanger. Spatially oriented format for acoustics: A data exchange format representing head-related transfer functions. In *Audio Engineering Society Convention 134*. Audio Engineering Society, 2013.

[20] J. Borwick. *Loudspeaker and Headphone Handbook*. Taylor & Francis, 2001.