Audio Engineering Society

# Convention Paper 8902

Presented at the 134th Convention
2013 May 4–7 Rome, Italy



# A standardized repository of Head-Related and Headphone Impulse Response data

Michele Geronazzo[1], Fabrizio Granza[1], Simone Spagnol[1], and Federico Avanzini[1]

[1] *Dept. of Information Engineering, University of Padova, Via Gradenigo 6/B, 35131 Padova, Italy*

Correspondence should be addressed to Michele Geronazzo (`geronazzo@dei.unipd.it`)

**ABSTRACT**

This paper proposes a repository for the organization of full- and partial-body Head-Related Impulse Responses (HRIRs/pHRIRs) and Headphone Impulse Responses (HpIRs) from several databases in a standardized environment. The main differences among the available databases concern coordinate systems, sound source stimuli, sampling frequencies and other important specifications. The repository is organized so as to consider all these differences. The structure of our repository is an improvement with respect to the MARL-NYU data format, born as an attempt to unify HRIR databases. The introduced information supports flexible analysis and synthesis processes and robust headphone equalization.

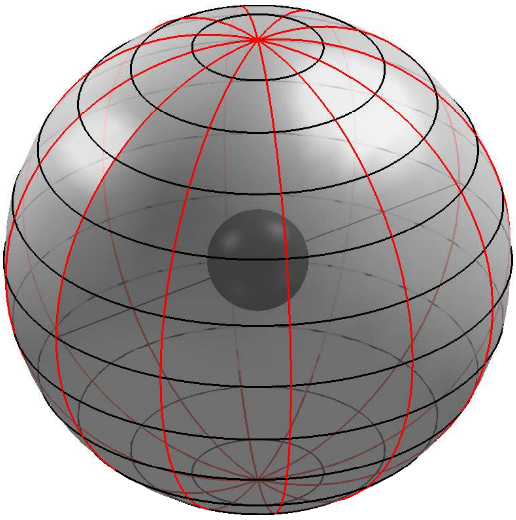
## 1. INTRODUCTION

The analysis and synthesis of 3D audio scenes requires the collection of data such as HRIRs and HpIRs. The former are usually recorded onto a significative number of human subjects and/or dummy heads by varying the position of the sound source with respect to the head, while the latter lead the equalization process of several types and models of headphones. The proposed repository, which represents one fundamental component of our own framework for the analysis and synthesis of head-related transfer function (HRTF) data (see [1] for more details), acts as an organized container for both HRIR and HpIR databases. In addition to full-body HRIRs, recordings of isolated body parts are also stored as partial Head-Related Impulse Responses (pHRIRs), e.g. Pinna-Related Impulse Responses (PRIRs) and HRIRs measured on pinnaless mannequins.

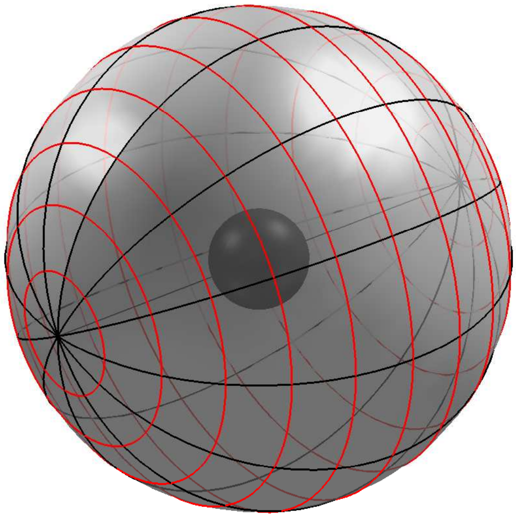
Several research groups provided public-domain HRIR databases. Among these, the CIPIC HRTF database [2] [1] and the LISTEN HRIR database [2]

[1] http://interface.cipic.ucdavis.edu/

[2] http://recherche.ircam.fr/equipes/salles/listen/



(a)



(b)

**Fig. 1:** Coordinate systems: vertical polar (a) and interaural polar (b). The inner grey sphere represents the human head, with the interaural axis evidenced by a thin grey line. Points along a red curve on the outer sphere have constant azimuth angle, points along a black curve have constant elevation angle.

are today the most known and exploited databases. The intrinsic differences among HRIR databases can be mainly summarized referring to the process behind the recordings and their storage, some aspects of which are now discussed.

First of all, the discrete spatial grid where the responses are taken has to be interpreted relatively to the assumed coordinate system: interaural polar or vertical polar. In the former case (e.g. CIPIC database), typically a loudspeaker (or multiple loudspeakers) sequentially plays the sound stimulus moving (or switching to the next one when multiple loudspeakers are used) along a semi-circle with the subject's head in its center until all the sampled positions are stored; then the stimuli are again played after having changed the elevation angle of the semicircle and thus moved in a sheaf of planes with the interaural line as rotation axis. In the latter case (e.g. LISTEN database), the rotation axis of the semi-circle spanned by the loudspeaker(s) is the middorsal line passing through the center of the head, and the azimuth angle is varied first. As a result, as Fig. 1 highlights, the spatial grids spanned by the two systems are different.

At the transmitter side (the loudspeaker), the audio chain has to be calibrated with respect to the stimulus that maximizes the signal-to-noise ratio of the recordings. Some examples of stimuli used in the aforementioned databases are Golay codes and logarithmic sine sweeps. At the receiver side, the position of the microphone plays a predominant role in signal acquisition: the qualitative distinction between blocked-ear-canal, open-ear-canal and eardrum recordings highlights the different acoustical information captured by the measurement setup. Furthermore, after recording all the collected raw data are processed and compensated in different possible ways (e.g. through inverse filtering of the stimulus plus free-field compensation) in order to extrapolate the impulse response. These and other differences among a number of public HRIR databases can be appreciated in Table 1.

On the other hand, no public HpIR databases nor standard HpIR repositories have been proposed to date. Typically, small sets of headphones are tested limited to the purpose of reproducing a *virtual auditory scene* (VAS) for a specific experiment and sub-

| Database | CIPIC [2] | LISTEN | FIU [3] | MIT [4] | AALTO [5] | ARI | PKU&IOA [6] |
|---|---|---|---|---|---|---|---|
| Sampling frequency | 44100 Hz | 44100 Hz | 96000 Hz | 44100 Hz | 48000 Hz | 48000 Hz | 65536 Hz |
| HRIR length | 200pts | 512pts | 256pts | 512pts | 8192pts | 256pts | 1024pts |
| Coordinate system | IP | VP | VP | VP | VP | VP | VP |
| Spatial grid | 25A,50E | 24A,10E | 12A,6E | 72A,14E | 72A,19E | 90A,22E | 72A,14E |
| No. of directions | 1250 | 187 | 72 | 710 | 240 | 1550 | 793 |
| No. of distances | 1 | 1 | 1 | 1 | 2 | 1 | 8 |
| Distance | 1 m | 1.95 m | – | 1.4 m | 0.68,1.35 m | 1 m | 0.2–1.6 m |
| Stimulus type | GC | LS | GC | ML | LS | LS | SG |
| Mic position | BEC | BEC | BEC | BEC | BEC | BEC,BTE | BEC |
| Data format | .mat | .wav,.mat | .txt | .wav | .mat | .mat | .dat |
| No. of subjects | 45 | 51 | 15 | 2 | 21 | 92 | 1 |
| Raw data | no | yes | no | yes | no | yes | no |
| Onset/ITD data | yes | no | yes | no | no | no | no |
| Anthropometric data | yes | yes | yes | no | no | yes | no |

**Table 1:** Specifications of seven public HRIR databases. Legend: IP = interaural polar; VP = vertical polar, GC = Golay codes, LS = logarithmic sweep, ML = maximum length sequence, SG = spark gap impulse, BEC = blocked ear canal, BTE = behind the ear. The *HRIR length* specification refers to the final, post-processed HRIRs. The two values in the *spatial grid* specification refer to the number of azimuth angles in the horizontal plane and the number of elevations in the median plane respectively; the uneven angular spacing is omitted for brevity.

ject [7]. Recent auralization softwares, e.g. AM3D [3], cluster headphones by type ( i.e. earphones, in-ear headphones, closed-back or open-back circumaural), each capturing common behaviours for a generic listener. However, given $H$ headphone models and $L$ listeners it is possible to measure $H \times L$ individual HpIRs. These responses need to be methodically stored in specific data structures sharing common features with individual HRIR databases.

In the following section, we will explain how the just discussed requirements are managed in our HRIR repository and how the same principles lead the formulation of our HpIR repository.

## 2. REPOSITORY STRUCTURE

### 2.1. HRIR and pHRIR databases

An attempt to unify the aforementioned variability in HRIR databases gave birth to the MARL-NYU data format [8]; CIPIC, LISTEN, FIU [3] [4] and KEMAR-MIT [4] [5] databases were stored in this format, which organizes the information into data and specification sections. Our repository takes the MARL-NYU format as a starting point towards the introduction of some missing relevant information:

- the raw HRIR data in addition to the already available compensated version;
- the HRIR's onset sample;
- the coordinate system adopted for the measurements and the relative management of the measurement space.

Furthermore, four more databases were stored and fitted to our repository:

1. the Aalto HRIR database [5] [6], which includes actual source direction data;
2. the ARI HRTF database [7], collecting both in-ear and behind-the-ear measurements;
3. the PKU&IOA HRTF database [6] [8], containing near-field recordings from a KEMAR dummy head;
4. the Aalto PRTF database [9] [9], which collects pinna-related impulse responses in the mid-sagittal plane as pHRIRs.

[3] http://www.am3d.com/
[4] http://dsp.eng.fiu.edu/HRTFDB/main.htm
[5] http://sound.media.mit.edu/resources/KEMAR.html
[6] http://www.acoustics.hut.fi/go/aes133-hrtf/
[7] http://www.kfs.oeaw.ac.at
[8] http://www.cis.pku.edu.cn/auditory/Staff/Dr.Qu.files/Qu-HRTF-Database.html
[9] http://www.dei.unipd.it/∼spagnols/PRTF_db.zip

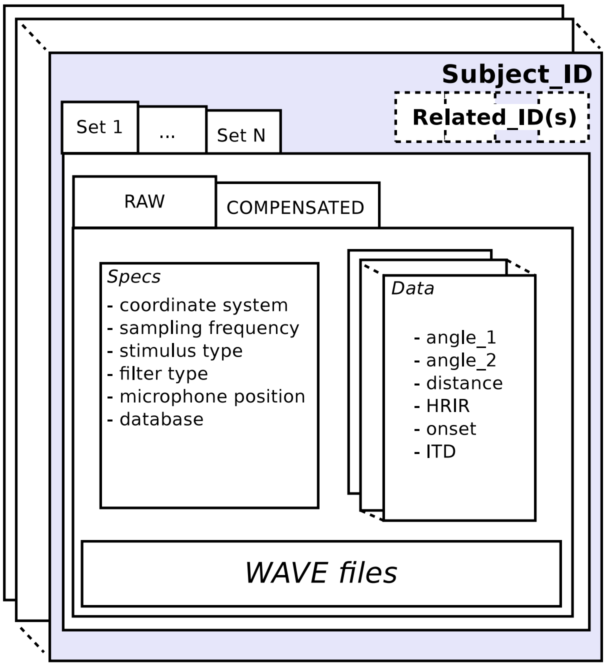



**Fig. 2:** HRIR/pHRIR repository structure.

As Fig. 2 sketches, each single measured subject can be associated to $N$ different HRIR sets. Each set corresponds to a different measurement session: for instance, open-ear canal and closed-ear canal recordings of a same subject make up two different sets. Each HRIR set is in turn associated to a data structure composed of two parts storing raw and compensated information. Both raw and compensated datasets are stored as .mat and .wav files. The .mat structure is divided into *specs* and *data* sections.

The former section reports all the required information about the subject, to whom an unique ID number is assigned, and the adopted measurement technique. The following details are included:

- *coordinate system*: interaural polar or vertical polar;
- *sampling frequency* (in Hz) of the HRIRs;
- *stimulus type*: excitation signal used for measurements;
- *filter type* (in compensated version only): filter used in the compensation of the HRIR (e.g. minimum-phase or fixed filter);
- *microphone position*: e.g. at the entrance of the blocked/open ear canal, at the eardrum, etc.;
- *database* (optional): name of the database to which the responses belong.

A crucial issue with respect to the first point is how to interpret the polar ranges for azimuth $\theta = [-180°, 180°)$ and elevation $\phi = [-90°, 90°]$ of the MARL-NYU database container, because the inclusion of a database likely implies a conversion of the angular ranges. This is the main reason why the information on which coordinate system was assumed is included.

The data section defines the details of each different HRIR measurement. Three fields univocally determine the spatial position of the sound source: azimuth and elevation angles (in degrees), and *distance* (in meters) from the center of the subject's head. According to the coordinate system, azimuth and elevation have different definition and range:

- in the interaural polar system, elevation $\varphi$ is the angle between the horizontal plane and the plane containing both the interaural axis and the sound source ($\varphi \in [-180, +180)$); on this plane, azimuth $\theta$ is the angle between the line of intersection with the median plane and the line passing from the origin of the axes and the sound source ($\theta \in [-90, +90]$);
- in the vertical polar system, azimuth $\theta$ is the angle between the median plane and the plane containing both the vertical axis and the sound source ($\theta \in [-180, +180)$); on this plane, elevation $\varphi$ is the angle between the line of intersection with the horizontal plane and the line passing from the origin of the axes and the sound source ($\varphi \in [-90, +90]$).

Such a differentiation could cause confusion and ambiguity (as it happens in the MARL data format); this is managed in our repository by calling the two angles $angle_1$ and $angle_2$: $angle_1$ is the one with range $[-90, +90]$, while $angle_2$ has $[-180, +180)$ range. Points with equal $angle_1$ describe a circle which is parallel to the median plane (interaural polar case) or parallel to the horizontal plane (vertical polar case).

The remaining fields of the data section are dedicated to the result of the measurement:

- *HRIR*: pair of vectors relative to the left and right ears containing the HRIR samples;
- *onset*: pair of indices indicating the onset sample of each HRIR vector, calculated as the sample which precedes the last zero-crossing before the main impulse of the HRIR;
- *ITD*: difference between the left- and right-HRIR onsets; if the sound source is on the left (right) the ITD is negative (positive).

Finally, the subject's anthropometric data is saved in the *anthropometry* folder and associated to a single set of HRIRs. As a matter of fact, each HRIR set has to be seen as a still frame of the subject's anthropometry during those measurements, which is not guaranteed to remain unchanged in a future measurement session. The data format is not defined at the moment, yet it is desirable to uniform heterogeneous anthropometric information in a coherent data format in line with biometrical standards [10].

The pHRIRs are organized in the same manner as the HRIRs; it is care of who includes the partial responses to keep track in a *comment* field of which structural component is associated to those signals. The HRIR and pHRIR repositories use different subject ID enumerations with the constraint that a subject whose partial responses are included in the pHRIR repository is linked to a corresponding subject in the HRIR repository through the *related_ID* field. Such a subject always exists, possibly with a corresponding HRIR data structure containing his/her anthropometric data only.

### 2.2. HpIR databases

The HpIR describes both the headphone's transducing properties and the headphone-to-eardrum transfer function [7]. In order to reach high localization accuracy in VASs and thus to provide proper binaural signals, headphones have to be carefully equalized.

In the typical sound transmission model, fundamentals of binaural technologies [11] assume as prerequisite for 3D audio rendering

$$Z_{headphones} \approx Z_{radiation} \tag{1}$$

where $Z_{radiation}$ denotes the equivalent impedance outside the ear canal in free-field listening conditions and $Z_{headphones}$ the equivalent impedance outside the ear canal with headphones. This equality holds for wavelengths greater than the ear canal's width, thus approximately under 10 kHz, and gives rise to the so-called *Pressure Division Ratio* (PDR):

$$P_{open}/P_{blocked} = P^{Hp}_{open}/P^{Hp}_{blocked} \tag{2}$$

where $P_{open}$ and $P_{blocked}$ denote the sound pressure at the entrance of the open- and blocked-ear canal respectively, while $P^{Hp}_{open}$ and $P^{Hp}_{blocked}$ denote the same sound pressure observation points when the sound source is a headphone. Headphones with $PDR \approx 1$ satisfy the *free-air equivalent coupling* (FEC) characteristic [11]. In order to verify this basic behaviour, several measurements with different recording conditions should be conducted and stored in our repository.[10]. Furthermore, analyses of headphone-to-external ear transfer functions, ear canal entrance-to-eardrum transfer functions and impedance mismatch need to be carried on in order to understand and control sound artifacts introduced by headphones.

The proposed standardization of HpIR databases follows an organization similar to the one introduced in the previous section for HRIR/pHRIR databases. The key observations that guide the design of such structure (see Fig. 3) are:

- HpIRs are highly sensitive to the positioning of headphones;
- both closed- and open-ear canal measurements are required for the evalution of binaural reproduction;
- how the headphones interact with the external ear is strictly subject-dependent.

The intersubjective variation is particularly marked in the high-frequency range, where important elevation cues generally lie. Thus, an inaccurate compensation likely leads to spectral colorations that affect

[10]Headphones with no FEC, e.g. insert headphones or in-ear monitors, could be employed in spatial audio reproduction as well, provided that the HpIRs are measured with an unblocked ear canal [12].

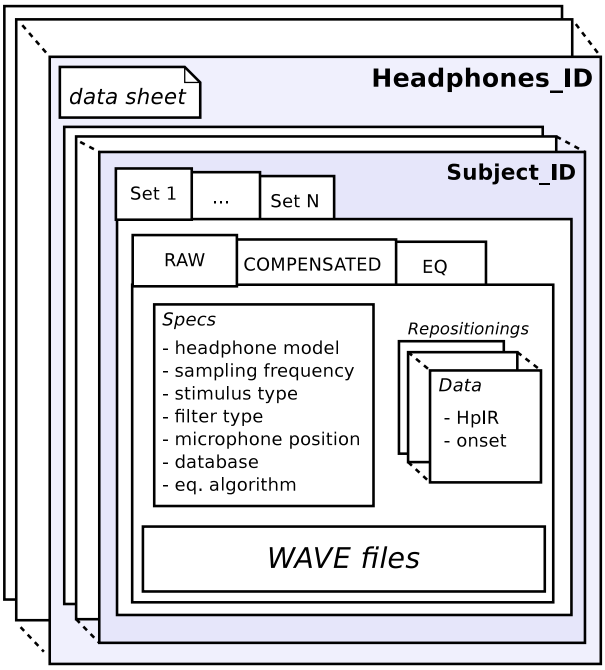



**Fig. 3:** HpIR repository structure.

both source elevation perception and sound externalization [13].

A new file-system level is inserted on top of the *subject* folder: for each pair of headphones, a collection of individual HpIRs and equalization filters are stored. Indeed, one of the purposes of this archive is to compute the equalization filter that compensates the headphone starting from the raw HpIRs.[11] Every *subject* folder contains three subfolders - *raw*, *compensated* and *eq*. The *raw* subfolder contain raw data from the recordings in both .mat and .wav formats. The second subfolder contains the compensated impulse responses, while the latter stores the equalization filter (under the form of an impulse response) obtained from the inverse HpIR through one or more techniques, e.g. considering the mean HpIR measurement with respect to all the repositionments of that device [15].

The .mat structure is divided into *specs* and *data* sections. The former section includes, additionally to the *sampling frequency*, *stimulus type*, *filter type*, and *microphone position* fields defined as for the HRIR repository, the following information:

- *headphone model*, producer included;
- *eq algorithm* (in *eq* version only): equalization algorithm used (e.g. preprocessed inversion transfer function, least-squares minimization of inversion error [14]).

Furthermore, similarly to the HRIR / pHRIR repository, it is possible to include generic HpIRs measured in other laboratories and keep track of this information in the *database* field.

The *data* section is made of an array of structures of length $R$ (= number of repositionings for that pair of headphones) with each cell containing the data from a single recording:

- *HpIR*: pair of vectors containing the HpIR samples, relative to the left and right ear/headphone respectively;
- *onset*: pair of indices indicating the left and right HpIR's onset sample.

Finally, specific characteristics of the headphones such as transducer type, acoustic coupling, and design are stored in their data sheet. This information resides in the root directory of that device and no re-organization has been made until now.

## 3. CONCLUSIONS AND PERSPECTIVES

The HRIR and HpIR repository described in this paper represents a further step towards both a binaural impulse response database standardization and a basic tool for the analysis, synthesis and rendering of 3D audio. The mixed structural modeling formulation introduced in [1] requires a well-defined repository in order to support the analysis and design of novel synthetic filter models and HRIR/pHRIR selection processes. Thanks to the proposed solution, a flexible mixture of acoustic responses and synthetic models potentially increases the possible pHRIR combinations together with their psychoacoustic accuracy.

The inclusion of computer-simulated HRIRs / pHRIRs calculated from mesh models of human

[11] Although various techniques have been proposed in order to face the equalization issue (see [14] for a review), modeling the correct equalization filter is still a hot open research theme.

heads [16] and spatially discretized so as to be added to our repository would raise the number of available subjects in a limited amount of time. The unification of heterogeneous datasets also facilitates normalization and cluster analysis of impulse response data (e.g. through principal component analysis).

Finally, the storage of individual headphone compensation filters promotes the personal use of headphones. This is especially true in experimental environments, yet HpIR modeling and cluster analysis would allow the extension of this approach to a commercial domain.

## 4. REFERENCES

[1] M. Geronazzo, S. Spagnol, and F. Avanzini. A modular framework for the analysis and synthesis of head-related transfer functions. In *Proc. 134th Conv. Audio Eng. Soc.*, Rome, Italy, May 2013.

[2] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF database. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, pages 1–4, New Paltz, New York, USA, October 2001.

[3] N. Gupta, A. Barreto, M. Joshi, and J. C. Agudelo. HRTF database at FIU DSP lab. In *Proc. 35th IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP 2010)*, pages 169–172, Dallas, TX, USA, March 2010.

[4] W. G. Gardner and K. D. Martin. HRTF measurements of a KEMAR. *J. Acoust. Soc. Am.*, 97(6):3907–3908, June 1995.

[5] J. Gómez Bolaños and V. Pulkki. HRIR database with measured actual source direction data. In *Proc. 133rd Conv. Audio Eng. Soc.*, San Francisco, CA, USA, October 2012.

[6] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *IEEE Trans. Audio, Speech, Lang. Process.*, 17(6):1124–1132, August 2009.

[7] D. Pralong and S. Carlile. The role of individualized headphone calibration for the generation of high fidelity virtual auditory space. *J. Acoust. Soc. Am.*, 100(6):3785–3793, December 1996.

[8] A. Andreopoulou and A. Roginska. Towards the creation of a standardized HRTF repository. In *Proc. 131st Conv. Audio Eng. Soc.*, New York, NY, USA, October 2011.

[9] S. Spagnol, M. Hiipakka, and V. Pulkki. A single-azimuth pinna-related transfer function database. In *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, pages 209–212, Paris, France, September 2011.

[10] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon. A survey on ear biometrics. *ACM Trans. Embedded Computing Systems*, 9(4):39:1–39:33, March 2010.

[11] H. Møller. Fundamentals of binaural technology. *Appl. Acoust.*, 36:171–218, 1992.

[12] M. Hiipakka, T. Kinnari, and V. Pulkki. Estimating head-related transfer functions of human subjects from pressure-velocity measurements. *J. Acoust. Soc. Am.*, 131(5):4051–4061, May 2012.

[13] B. Masiero and J. Fels. Perceptually robust headphone equalization for binaural reproduction. In *Proc. 130th Conv. Audio Eng. Soc.*, London, UK, May 2011.

[14] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *Proc. 126th Conv. Audio Eng. Soc.*, Munich, Germany, May 2009.

[15] R. Nishimura, H. Kato, P. Mokhtari, and H. Takemoto. Effects of headphone calibration functions on subjective impressions of spatial sound. In *Proc. 20th Int. Congr. Acoust. (ICA 2010)*, Sydney, Australia, August 2010.

[16] B. F. G. Katz. Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. *J. Acoust. Soc. Am.*, 110(5):2440–2448, November 2001.