



Audio Engineering Society

Convention Paper 8571

Presented at the 131st Convention
2011 October 20–23 New York, NY, USA

This Convention paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This convention paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. 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.

Towards the Creation of a Standardized HRTF Repository

Areti Andreopoulou, and Agnieszka Roginska

Music and Audio Research Laboratory (MARL), New York University, New York, NY, 10012, USA
aa1510@nyu.edu, roginska@nyu.edu

ABSTRACT

One of the main spatial audio topics, nowadays, involves working towards an efficient individualization method of Head Related Transfer Functions. A major limitation in this area of research is the lack of a large and uniform database that will incorporate as many individualized properties as possible. This paper presents the MARL-NYU file format for storing HRTF datasets, and investigates the necessary normalization steps that assure a uniform and standardized HRTF repository, by compiling selected datasets from four HRTF databases.

1. INTRODUCTION

When a sound wave reaches a listener, it interacts with their head, body and pinnae in complex, constructive and destructive ways. The combination of these filtering effects is captured in the Head Related Impulse Responses (HRIRs), or their frequency domain equivalent Head Related Transfer Functions (HRTFs). Variations in the physiology of the head, body, and especially pinnae of different individuals require the acquisition of personalized HRTFs through numerous measurements for the creation of a realistic virtual auditory scene.

The publicly available HRTF databases are captured with different standards - varying azimuth and elevation increments, covering different ranges, at several filter lengths and sample rates. This makes the compilation of a repository a challenge. Inevitably, the need for definition of a standard in the HRTF acquisition process has a lot of premise. Such an action will increase compatibility between different 3D audio systems, will eliminate data redundancy in HRTF datasets, and will simply file exchanges among researchers in general.

In the absence of a well-defined and widely acceptable procedure of capturing HRTF datasets, this paper proposes a new MATLAB file format for storing HRTF sets, and describes the process of combining different databases into a large, uniform and standardized repository.

2. HRTF DATABASES

The specifications of four of the most common, publicly available HRTF databases are described in this section. These collections were used as the main source for the creation of the compiled repository. Each database was captured in different conditions, covering different ranges, and at varying azimuth and elevation increments.

2.1. The LISTEN database

The Institute for Research and Coordination Acoustic/Music (IRCAM) in collaboration with AKG has released an HRIR measurement database, as part of the Listen research project [15]. The set that consists of 51 subjects was captured using logarithmic sine-sweep signals at 44100 Hz. 10 different elevations were measured starting at -45° and ending at 90° in 15° vertical increments. The number of azimuth locations varies from 24 (15° azimuth increments in 0° elevation) to just 1 (in 90° elevation). The Impulse Responses are publicly available as 512-point minimum-phase filters with the corresponding Interaural Time Differences (ITDs).

2.2. The CIPIC database

The CIPIC database was captured at the Center for Image Processing and Integrated Computing, University of California Davis [1]. The set consists of 43 human subjects plus 2 KEMAR mannequins (with large and small pinnae), measured at 50 different elevations from -45° to 230.625° in 5.625° increments, and at 25 azimuth locations ($\pm 80^\circ$, $\pm 65^\circ$, $\pm 55^\circ$, and from -45° to $+45^\circ$ in 5° increments). Each Impulse Response is 200 samples long, corresponding to a duration of 4.5 ms, and was captured using Golay-Code signals at 44100 Hz. The distance from the speakers to the subject was adjusted to 1 m. The database also includes a set of 27 anthropometric measurements of each subject's head, torso and pinnae.

2.3. The FIU database

The Florida International University DSP Lab has released its own HRTF database in 2010 [7]. The Impulse Responses were captured with the HeadZap system from AuSIM 3D using Golay-Code signals at 96000 Hz. The set includes measurements of 15 human subjects in 6 different elevations (-36° , -18° , 0° , 18° , 36° , and 54°) and at 12 azimuth locations (every 30°)

creating a grid of 72 locations around each subject. The deliverable responses are 256-point minimum phase filters with their corresponding ITD values. The database also includes 3D images and anthropometric measurements of each subject's pinnae.

2.4. The MIT / KEMAR database

The KEMAR database was recorded at the MediaLab Institute of Technology of Massachusetts (MIT) [6]. It was captured using Maximum Length pseudo-random binary Sequences (MLS), with the speaker placed 1.4 m. away from the KEMAR mannequin. A total of 710 locations were recorded; 14 different elevations from -40° to 90° in 10° vertical increments, with the number of corresponding azimuth positions varying from 72 (5° horizontal increments), to just 1 (in 90° elevation). The resulting Impulse Responses are 512-points long at a sampling frequency of 44100 Hz.

3. THE MARL-NYU HRTF FILE FORMAT

As it has been seen so far, the existing HRTF databases were captured in slightly different conditions (at various distances and with different excitation signals), at varying azimuth and elevation increments, covering different ranges, and have several filter lengths and sample rates. In this paper we propose a new file format for storing HRTF datasets that accounts for this variance, allowing for filters to be stored in their original form, without the need of any transformations, yet assuring a smooth navigation between different sets in the case of a composite HRTF repository.

The MARL-NYU file format, created in MATLAB, avoids any data redundancy by storing the filters only in their original, recorded form, reducing therefore significantly the size of the data to be stored and handled. It can hold important information regarding the recording conditions of the Impulse Responses, such as the distance of the subject from the sound source and the type of excitation signal used.

Every location in a subject's set of measurements is stored in an array called *data*. Each position in the *data* array is a structure that encapsulates all the location-specific filter information, namely: the azimuth and elevation positions, the left and right ear HRIRs or HRTFs, the corresponding ITD values, and the distance of the sound source from the subject. Every consecutive measurement is stored in a separate structure and is appended to the original array.

All the information that is specific to a whole set of measurements (dataset), such as the sample rate, the type of filter (e.g. minimum-phase), and the type of excitation signal used (sine sweep, MLS, Golay codes, or other) are stored in a separate structure called *specs*. The *specs* structure, also, holds the name of the subject, and that of the database where this dataset belongs. Additional information, either location or dataset specific, can be stored in the *comments* fields that exist both in the *data* and in the *specs* structures.

The MARL-NYU file format

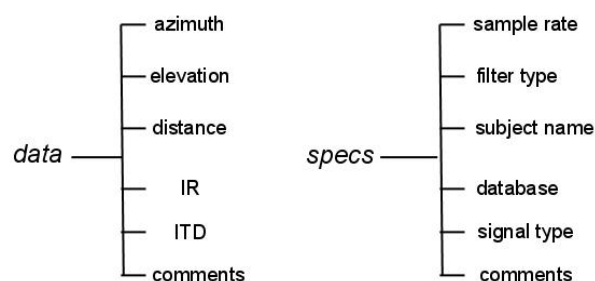


Figure 1. Outline of the MARL-NYU file format for storing HRTF datasets.

The MARL-NYU format is currently fully supported by *ScanIR*, which is an application for multi-channel impulse response measurements in MATLAB, by Boren & Roginska [3].

The flexibility of our suggested file format in accommodating inter-database variations makes the compilation of an HRTF repository less challenging. The resulting HRTF collection consists of 110 datasets selected from the LISTEN, CIPIC, FIU, and MIT databases, and is available for download at <http://marl.smusic.nyu.edu/>.

In the following section we describe in detail the necessary standardization steps that would assure uniformity in such compiled HRTF collections. A discussion of the potential benefits from a large HRTF repository follows.

4. STANDARDIZATION PROCESS

The first step towards the acquisition of a uniform database was to assure that all HRTF filters had the same sampling rate. All databases have a sampling rate of 44100 Hz, except for the FIU, which was captured at 96000 Hz. A down-sampling algorithm implemented in MATLAB was used to fix this discrepancy.

The next step was to address the inconsistency in the azimuth and elevation increments in each database. Most of the available databases advance at varying, location-dependent azimuth and elevation intervals, and cover different ranges. In our compiled repository each HRTF dataset is forced to increase in increments of 15° in both azimuth and elevation.

In order to achieve this, all non-minimum-phase databases had to be converted into minimum-phase ones. We can say that a minimum-phase filter is a variation of the original filter with the initial time-delay removed. Such filters have all their poles and zeros inside the unit circle, and can be obtained through the calculation of the real cepstrum of the original HRIRs, which have arbitrary phase. In the frequency domain, both the original and the resulting filters share the same magnitude spectrum [8]. The corresponding ITD values were also stored.

In the following stage we linearly interpolated between adjacent HRIR filters to create new ones that would match the locations of the desired 15-degrees resolution grid, for each dataset. The ITD values were also adjusted to match the new coordinates of the filters.

The final step concerned the inconsistency in the lengths of the HRIRs in different databases, which varied between 512 and 128 samples. This issue was overcome by zero padding and transforming all minimum-phase filters into the frequency domain. This operation resulted in 256 samples long HRTFs for all datasets in our repository.

The standardization process resulted in a repository where each set consists of 256 samples long, minimum-phase HRTFs, selected at 15° azimuth and elevation (from -45° to 90°) increments, with extracted and interpolated ITD values, and at a sampling rate of 44100 Hz.

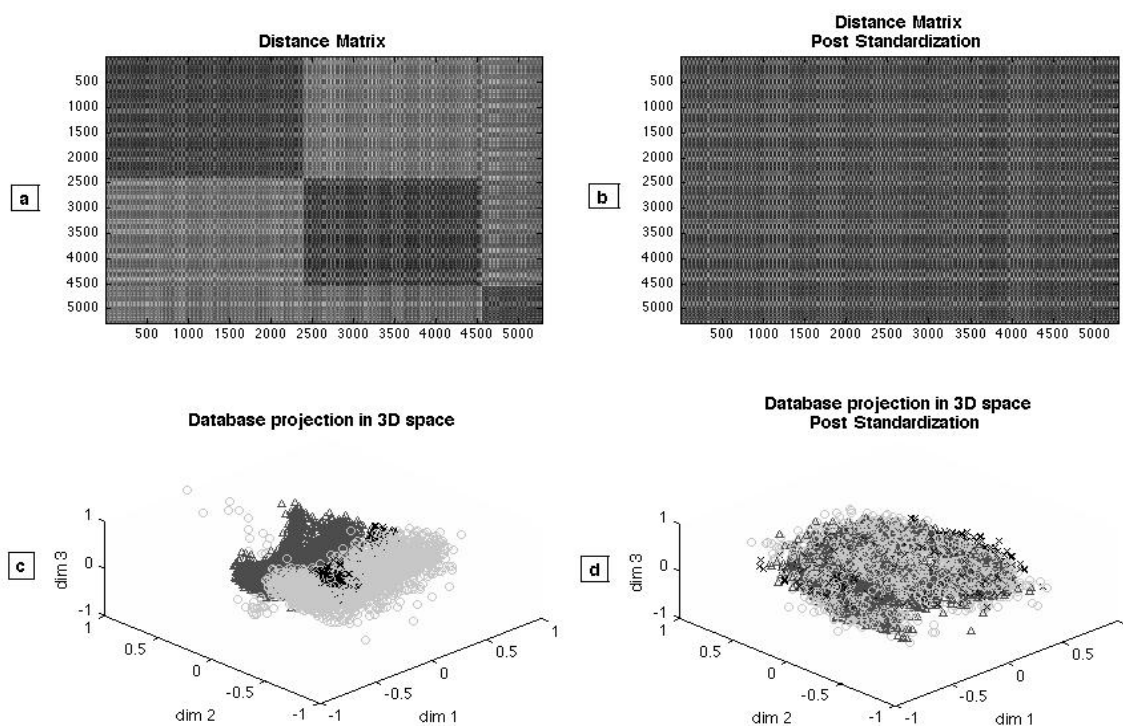


Figure 2. Distance matrices and Multidimensional Scaling projections on a 3D space of the log-magnitude HRTF repository subset. 2a: Distance matrix on the pre-Standardization repository; the dark rectangle patterns clearly mark the three larger HRTF subsets. 2b: Distance matrix on the post-Standardization repository; it does not reveal any database-related patterns, only inter-HRTF-filter relationships. 2c: MDS projection of the repository subset on a 3D space; it is apparent that each database occupies its own distinct sub-area in 3D feature space. 2d: MDS projection of the Standardized repository subset on a 3D space; upon effective standardization, all HRTF filters uniformly projected on the 3D feature space.

5. UNIFORMITY EVALUATION OF THE COMPILED DATABASE

5.1. Uniformity evaluation

In order to evaluate the uniformity of our newly compiled database, the distance matrix was computed on the log magnitude of a subset of the HRTFs. This subset only contained filters on the horizontal plane (0° elevation).

A distance matrix can be realized as a two-dimensional array containing the pairwise distances in a given set of points. For this test, Euclidean distance was used as a metric, resulting in a matrix that was representing the spacing of the HRTF subset in Euclidean space. The distance is zero on the diagonal and symmetric across that.

The plot of the distance matrix revealed patterns in the data collection that were strongly correlated to the characteristics of each initial database. More specifically, in Figure 2a one can observe the three clearly marked, dark gray, square areas lined on the diagonal of the plot, which correspond to the three larger databases on the repository; the LISTEN, CIPIC and FIU respectively.

Multidimensional scaling (MDS) was furthermore applied to the distance matrix, as a way of visualizing the HRTF data. MDS is a mathematical technique used to visualize proximities in data in a lower dimensional space. More specifically, given a distance matrix Δ the goal of the MDS is to find I new vectors $x_1, \dots, x_I \in \mathbb{R}^N$ such that:

$$\|x_i - x_j\| \approx \delta_{i,j} \text{ for all } i, j \in I \quad (1.)$$

where $\delta_{i,j}$ is the distance between the i^{th} and j^{th} points in the distance matrix Δ and $\|x_i - x_j\|$ is a vector norm.

This technique enabled the projection of all HRTF filter points from a 256 to a 3-dimensional space.

As it can be seen in Figure 2c, each dataset occupies its own distinct sub-area in 3D feature space. Such a behavior, although it disallows for a smooth, unified HRTF repository, has been observed in the past, especially in user-selection HRTF studies. In such studies, when more than one HRTF collections were offered as an option to the subjects, the results would reveal that certain subject sub-groups would have a strong preference for one collection over another [12]. Supplementary standardization steps were, therefore, thought to be necessary.

5.2. Further standardization steps and reevaluation

In order to eliminate the remaining normalization issues, further standardization steps were employed. More specifically, subtraction of the mean and division by a standard deviation were applied to each of the four HRTF databases separately.

If $D_L(M \times N)$ is a data matrix that contains all the LISTEN database filters, in which each $d_{Li}(i = 1 : M)$ is a column vector that corresponds to an HRTF filter, the empirical mean vector $\mu(N \times 1)$ can be calculated as:

$$\mu = \frac{1}{M} \sum_{i=1}^M d_{Li} \quad (2.)$$

and the standard deviation one $\sigma(N \times 1)$ as:

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (d_{Li} - \bar{d}_L)^2} \quad (3.)$$

The newly computed distance matrix on the standardized repository did not reveal any strong, database-related patterns, as it can be seen in Figure 2b. Rather it contained information on the similarity patterns of individual HRTF filters. Moreover, when

MDS was re-applied to the collection, all filters were properly centered along the same 3D space in a uniform manner (Figure 2d).

6. DISCUSSION AND CONCLUSIONS

Currently, there are very few HRTF sets available to the spatial audio community, each captured with different standards (at varying azimuth and elevation increments, covering different ranges, at several filter lengths and sample rates), a fact that makes the compilation of a repository a challenge. In this paper a new file format for storing HRTF datasets was described, built to account for all this variance, assuring a smooth navigation between HRIR sets originating from different databases.

The agreement for a universal, well defined HRTF file format is an essential step that needs to be taken by the spatial audio community. Such an action will make the incorporation of databases into a system easier and will simply file exchanges among researchers in general. As it was demonstrated in this paper, simple normalization steps suffice in assuring a standardized HRTF repository for further applications.

In the absence of a well-defined and widely acceptable procedure of capturing HRTF datasets, the MARL-NYU file format, created in MATLAB, was designed to hold filters in their original, recorded form, avoiding the need for any ad hoc transformations, and minimizing data redundancy. By offering the option of holding supplementary, database-specific information such as the sampling rate, filter length, distance from the sound source, type of excitation signal, etc., it allows for uniformity while preserving the original characteristics of each database.

One research field that could potentially benefit from the acquisition of a standardized HRTF repository is that of HRTF individualization, which was developed to improve spatial impression in systems, where the use of subject-specific filters is not an option. Currently, several methods that lead towards this goal are explored, such as HRTF decomposition [5], Principal Component Analysis [9], HRTF clustering [4; 13], the Boundary element method [10], etc. In these methods a large and uniform repository, which would account for as many individualized filtering properties as possible could lead to a robust individualization system that would be less prone to subject-dependent HRTF variations.

Another field that could also benefit from an easily expandable, standardized HRTF repository is Database Matching, where the success of the system is directly dependent on the variety in the available filter collection. In this method each user is paired to an existing library of measurements based on common features and characteristics, either through subjective [11] or objective [14] analysis. In cases where not all different user profiles are covered, this method may fail to provide the best match to a given case, degrading, in such way, localization accuracy and spatial experience during binaural reproduction.

Future work will involve an effort to expand to an even larger HRTF repository that would account for the characteristics of more individuals, and an investigation of the relationships among individual HRTF filters with the use of clustering analysis.

7. REFERENCES

- [1] Algazi, V.R, Duda, R.O, Thompson, D.M, and Avendano, C. "The CIPIC HRTF Database". Proc. of the *IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics*, pp. 99-102, Mohonk Mountain House, New Paltz, NY, October 2001.
- [2] Blauert, J. "Spatial Hearing: The Psychophysics of Human Sound Localization". Mass: MIT Press, 1996.
- [3] Boren, B. and Roginska, A. "Multichannel Impulse Response Measurement in Matlab". Proc of the *131st AES Convention*, New York, NY, October 2011.
- [4] Fahn, C., and Lo, Y.C. "On the clustering of Head-Related-Transfer Functions used for 3-D Sound Localization". *Journal of Information Science and Engineering*, Vol. 19, pp. 141 – 157, 2003.
- [5] Faller, K. J., II, Barreto, A. and Adjouadi, M. "Decomposition of Head-Related Transfer Functions Based on the Hankel Total Least Squares Method". Proc of the *Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pp. 161 - 166, Macro Island, FL, January 2009.
- [6] Gardner B, Martin KD, "HRTF Measurements of a KEMAR", *Journal of the Acoustical Society of America*, vol. 97, issue 6, pp. 3907-3908, June 1995.
- [7] Gupta, N., Barreto, A., Joshi, M., & Agudelo, J. C. "HRTF Database at FIU DSP Lab". Proc of the *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 169–172, Dallas, TX, March 2010.
- [8] Hugeng, D. G., & Wahab, W. "Effective Preprocessing in Modeling Head-Related Impulse Responses Based on Principal Components Analysis". *Signal Processing: An International Journal*, 4(4), pp. 201-212, 2010.
- [9] Hwang, S. and Park, Y. "Interpretations on principal components analysis of head-related impulse responses in the median plane". *Journal of the Acoustical Society of America* 123 (4), pp. EL65-EL71, March 2008.
- [10] Katza, B. F. G. "Boundary element method calculation of individual head-related transfer function". *Journal of the Acoustic Society of America*, 110 (5), Pt. 1. pp. 2440 - 2448, 2005.
- [11] Roginska, A., Wakefield, G. H., and Santoro, T. S. "User Selected HRTFs: Reduced Complexity and Improved Perception". *Undersea Human System Integration Symposium*, Providence, RI, July 2010.
- [12] Roginska, A., Santoro, T. S., and Wakefield, G. H. "Stimulus-dependent HRTF preference". Proc of the *129th AES Convention*, San Francisco, CA, USA, November 2010.
- [13] So, R.H.Y., Ngana, B., Hornerb, A., Braasch, J., Blauert, J. and Leung, K. L. "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: cluster analysis and an experimental study". *Ergonomics*, Vol. 53, No. 6, pp. 767 - 781, June 2010.
- [14] Zotkin, D. N., Hwaiig, J., Duraiswami, R., and Davis, L. S. "HRTF Personalization Using Anthropometric Measurements". Workshop on *Applications of Signal Processing to Audio and Acoustics*, pp. 157 - 160, New Paltz, NY, October 2003.
- [15] <http://recherche.ircam.fr/equipes/salles/listen/index.html>