Audio Engineering Society

Convention e-Brief 357

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

A database of head-related transfer function and morphological measurements

Rahulram Sridhar, Joseph G. Tylka, and Edgar Y. Choueiri

3D Audio and Applied Acoustics Laboratory, Princeton University, Princeton, NJ, 08544

Correspondence should be addressed to Rahulram Sridhar (`rahulram@princeton.edu`)

ABSTRACT

A database of head-related transfer function (HRTF) and morphological measurements of human subjects and mannequins is presented. Data-driven HRTF estimation techniques require large datasets of measured HRTFs and morphological data, but only a few such databases are freely available. This paper describes an on-going project to measure HRTFs and corresponding 3D morphological scans. For a given subject, 648 HRTFs are measured at a distance of 0.76 m in an anechoic chamber and 3D scans of the subject's head and upper torso are acquired using structured-light scanners. The HRTF data are stored in the standardized "SOFA format" (spatially-oriented format for acoustics) while scans are stored in the Polygon File Format. The database is freely available online.

1 INTRODUCTION

Head-related transfer functions (HRTFs) of an individual describe the idiosyncratic filtering of incident acoustic waves by the individual's morphology and are widely used in synthesizing binaural signals for spatial audio reproduction. The most accurate way to acquire HRTFs is via acoustical measurements in an anechoic chamber [1, Ch. 2]. Since this is commercially infeasible, alternative techniques for estimating HRTFs have been proposed, many of which are summarized by Xie [1]. Most techniques require morphological data that includes either measurements of specific anthropometric features [2] or complete 3D scans of the individual's morphology [3]. Data-driven techniques additionally require corresponding measured HRTFs of a large number of individuals. These HRTFs typically serve as benchmarks for validating different techniques either objectively or via subjective listening tests, and also serve as training data for data-driven techniques. For example, a recent data-driven technique to compute HRTFs directly from head scan point clouds requires measured HRTFs and 3D head scans of many individuals as training data [4].

Many publicly available databases exist that include measured HRTFs for many human subjects and mannequins.[1] However, only a small subset of databases also include corresponding morphological data. In particular, the RIEC database[2] includes 3D head scans [5], the SYMARE database additionally includes torso scans [6], and the more recent database from ITA includes scans of the pinnae only [7]. The CIPIC [8] and ARI[3] databases each include a finite selection of anthropometric measurements, while the LISTEN database[4]

---

[1] See, for example: `https://www.sofaconventions.org/mediawiki/index.php/Files.`

[2] `http://www.riec.tohoku.ac.jp/pub/hrtf`

[3] `http://www.kfs.oeaw.ac.at/hrtf`

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

includes only a sparse set of similar measurements.

Unfortunately, some of the scans from the RIEC database appear to have significant holes, and registration and alignment issues. Furthermore, many of the measured head-related impulse responses (HRIRs) from both the RIEC and CIPIC databases have undesirable pre-responses prior to the main impulses, which may make the data unreliable for use with data-driven techniques without sufficient post-processing. Rugeles Ospina et al. [9] also show that anthropometric features measured directly from 2D photographs, as done for the CIPIC and LISTEN databases, are generally less accurate than corresponding measurements made using 3D scans.

Recognizing the growing need for measured HRTF and 3D morphological data, we have begun an on-going project to measure HRTFs and 3D scans of humans and mannequins, which we compile into a publicly available database. In Sec. 2, we present details of the measurement procedures. In Sec. 3, we describe the signal processing performed on the measured data. We visualize a sample of the data in Sec. 4 and summarize our work in Sec. 5. The database is available online from the 3D Audio and Applied Acoustics Laboratory at Princeton University.[5]

2 MEASUREMENT PROCEDURE

2.1 Head-related transfer functions

We conduct acoustical measurements in a $3.6 \times 2.35 \times 2.55$ m ($l \times w \times h$) anechoic chamber with 8-inch deep (equal to $1/4$ wavelength at ~ 425 Hz) anechoic foam wedges. In the chamber, we place a circular arc which stands vertically and is aligned to be concentric with the "origin" of the chamber (i.e., where the center of the subject's head is ultimately placed). We attach to the arc eight loudspeakers (Genelec 8030A), which are equally-spaced (in 15° increments) between -30° and 75° elevation, and we include a ninth loudspeaker mounted on a separate stand at -57° elevation.[6] Specifically, we align the high-frequency drivers of the loudspeakers with these elevations such that the distance from each high-frequency driver to the origin is approximately 0.76 ± 0.005 m. We also place, directly below the origin, a custom-built chair that is affixed to a computer-controlled turntable (Outline ET250-3D), whose axis of rotation passes through the origin of the chamber. The chair, which is designed to have a minimal effect on incident acoustic waves, consists of a drum-throne seat with backrest and a thin "headrest" structure that provides a reference for positioning the subject's head, in order to minimize head movements during measurements. An image of the setup is shown in Fig. 1.

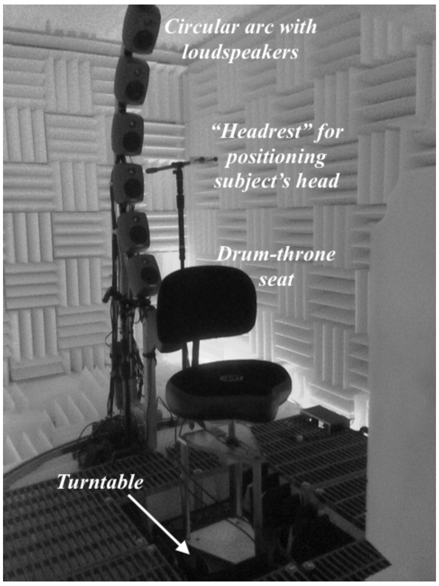

Fig. 1: Setup used to make HRTF measurements.

.

Prior to making measurements, we calibrate and equalize the binaural microphones (Theoretica Applied Physics BACCH-BM Pro). We first adjust, for each channel, the microphone gain (using a B&K Type 4231 calibrator and DP-0978 adapter) such that a 94 dBSPL (rms) sine tone produces a -11 dBFS (peak) signal. We then place the microphones at the origin of the chamber, facing the arc and parallel to the horizontal plane, in order to measure a set of nine "reference" impulse responses (RIRs), one for each elevation. For these measurements, we remove the seat cushion, backrest, and headrest, and then cover the remaining metal structure of the chair with anechoic foam wedges in order to minimize the acoustical influence of the chair-structure on the measurements.

We measure the RIRs by sending to the loudspeak-

---

[5] `https://www.princeton.edu/3D3A/HRTFMeasurements.html`

[6] Here, we use the same spherical coordinate system as that defined in AES69-2015 [10].

ers a series of partially-overlapping exponential sine sweeps [11] (generated in Plogue Bidule) and recording the resulting signals with the microphones. The delay between each successive sweep is 200 ms, yielding distinct impulse responses of up to 200 ms in duration. All measurements are conducted at a sampling rate of 96 kHz and the sweep signals are generated with a nominal frequency range of 20 Hz to 48 kHz, a duration of 500 ms, and an amplitude (at 1 kHz) of 70 dBSPL (rms). The measured signal-to-noise ratio is approximately 38.5 dB for each microphone. The RIRs are used to equalize the transfer functions of each loudspeaker-microphone pair, as described in Sec. 3.1.

For each subject, we measure binaural impulse responses (BIRs) for 648 directions: all nine loudspeaker elevations for each of 72 azimuths (equally spaced between 0° and 355°). We seat the subject on the chair such that the center of the subject's head coincides with the origin. We then place the binaural microphones at the entrances to the subject's blocked ear canals and measure BIRs using the same multiple exponential sine sweeps described above. The subject is rotated in 5° increments and the sweeps are repeated until the measurements are complete. In total, these measurements (including rotation time) takes $\sim$ 11 minutes.

2.2 Morphology

We include in our database corresponding head and torso scans captured using a PrimeSense Carmine 1.08 sensor (3.4 mm spatial, x/y, resolution and 1.2 cm depth resolution at a working distance of 2 m) and pinnae scans captured using an Artec Space Spider structured-light scanner (up to 0.1 mm resolution at a working distance ranging from 0.2 to 0.3 m). To scan the head and torso, each subject is asked to sit on a swivel chair and rotate slowly while a scanning operator adjusts, by hand, the sensor's position and orientation as necessary. The same procedure is repeated with the subject wearing a wig cap, which prevents the subject's hair from obstructing the pinnae. Finally, the scanning operator similarly scans each of the subject's pinnae independently while the subject is stationary.

The head and torso scans are acquired using the Skanect Pro software and the pinnae scans using the Artec Studio 12 Professional software. In all cases, markers of various color, size, and shape are placed on the subject prior to scanning in order to facilitate alignment of scans, as described in Sec. 3.2. Total scanning time (to acquire all scans) for a subject is $\sim$ 10 minutes.

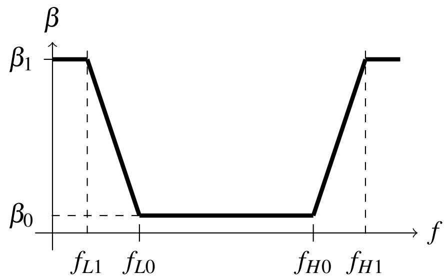

Fig. 2: Frequency-dependent regularization function of the inverse filters for HRTF equalization.

3 DATA PROCESSING

3.1 Head-related transfer functions

The head-related impulse responses (HRIRs) are obtained by equalizing, for each subject and for each loudspeaker-microphone pair, the measured binaural impulse responses (BIRs) by the corresponding reference impulse responses (RIRs). We first apply a 42 ms rectangular window to all of the raw BIRs and RIRs. We then generate inverse filters for the RIRs using frequency-dependent regularization, such that the transfer function of the inverse filter is given by [12]

$$G(f) = \frac{H^*(f)}{H^*(f)H(f) + \beta(f)}, \tag{1}$$

where $(\cdot)^*$ denotes complex conjugation, f is frequency, H is the transfer function of a measured RIR, and β is the regularization function. This function is defined by a set of parameters, which are defined graphically in Fig. 2, and whose default values are given by

$$\begin{aligned} \beta_0 &= 0, & f_{L0} &= 100\text{ Hz}, & f_{H0} &= 30\text{ kHz}, \\ \beta_1 &= 10^{-3}, & f_{L1} &= 50\text{ Hz}, & f_{H1} &= 32\text{ kHz}. \end{aligned}$$

These values were found to sufficiently limit any preresponses in the equalized HRIRs (see Sec. 4), while retaining a wide usable bandwidth. Finally, we convolve the BIRs with these inverse filters and apply a Tukey window to generate HRIRs that have an approximate duration of 10 ms. The BIRs, RIRs, and HRIRs are all included in the database as separate SOFA (spatially-oriented format for acoustics) files [10].

3.2 Morphology

For each subject, we first generate, using Skanect Pro, a watertight mesh of each head and torso scan (with and without the wig cap) and export the mesh, simplified to contain 3×10^5 faces, in the Polygon File Format (PLY). Each scan is then imported into Artec Studio 12, where we translate and rotate it such that the y-axis [10, Fig. 1] is also approximately the interaural axis, with the coordinate system origin located approximately halfway between the entrances to the two ear canals. We also align the x-axis by first determining, via direct observation on the subject, the point on the nose that is at the same height off the ground as the entrance to one of the ear canals, and subsequently ensuring that the x-axis passes through this point. The aligned scans are then re-exported in the PLY format and stored in the database as "consumer-grade" scans.

A copy of the head and torso scan with the subject wearing a wig cap is then modified by replacing the pinnae with those acquired using the Artec scanner. Replacement is achieved by manually erasing the pinnae captured using the PrimeSense sensor and algorithmically aligning those captured using the Artec scanner with the head scan. Alignment is facilitated with the help of markers placed on the subject that are captured in both the head and torso, and pinnae scans. The resulting scans are similarly reoriented, resampled, and exported in the PLY format, all from within Artec Studio 12. These exported scans are stored in the database as "reference" scans.

4 DATA VISUALIZATION

To verify that the our measured HRTFs are free of artifacts (e.g., pre-responses prior to the main impulses), we generate a surface plot of ITD estimated from measured HRTFs using a thresholding approach (see, for example, Katz and Noisternig [13]) with a 20% threshold. Figure 3 shows this surface plot for one of the subjects in our database. This plot shows a plausible ITD surface, as it is generally smooth and free of discontinuities, suggesting that the data are free of significant artifacts, noise, or any other errors.

Figure 4 shows, as an example, a composite of reference and consumer-grade scans of a B&K HATS 4128C mannequin. The two scans may be distinguished by observing the finer details in the pinna of the reference scan. We see that the scanning procedure and processing described in Sec. 2.2 and Sec. 3.2, respectively, yield scans that are generally free of holes and other surface issues.

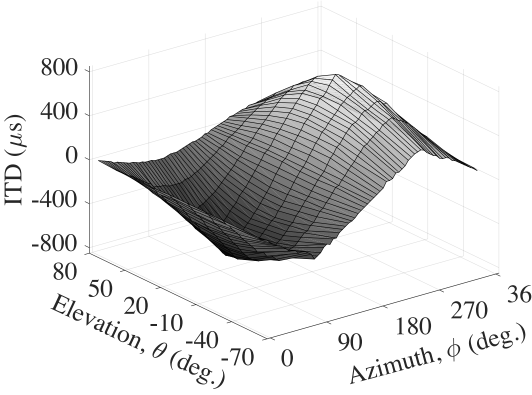

Fig. 3: Surface plot of ITD in μs for one of the subjects in our database.

5 SUMMARY

We have begun an on-going project to measure HRTFs and 3D scans of the head and upper torso of human subjects and mannequins, and compile the data into a freely available database. The project primarily aims to address the lack of such data. We describe details of the measurement procedures used to acquire the data and the subsequent signal processing performed. We also visualize a sample of the data to illustrate that the HRTFs and 3D scans included in our database are free of significant artifacts and noise. The HRTFs are stored in the SOFA format (with separate files for BIRs, RIRs, and HRIRs, as described in Sec. 3.1) and scans are stored in the PLY format.

Acknowledgements

We wish to thank all subjects who provide consent to have their HRTFs and morphological measurements made. We also thank A. Sheron and R. Sorenson for helping with early design ideas for the custom-built chair, and J. Hollingsworth for his help with machining parts of the chair. This work was sponsored by the Sony Corporation of America, and is approved by the Institutional Review Board for human subjects research at Princeton University.

References

[1] Xie, B., *Head-related transfer function and virtual auditory display*, J. Ross Publishing, 2013.

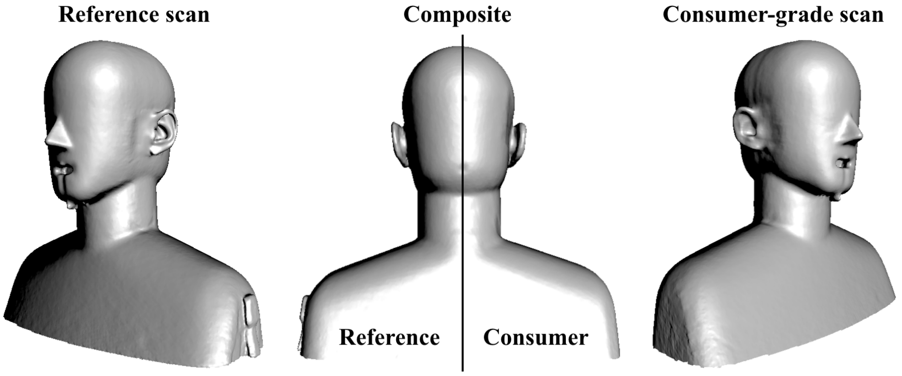

Fig. 4: Composite of reference and consumer-grade scans of the B&K HATS 4128C mannequin.

.

[2] Bilinski, P., Ahrens, J., Thomas, M. R., Tashev, I. J., and Platt, J. C., “HRTF magnitude synthesis via sparse representation of anthropometric features,” in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pp. 4468–4472, IEEE, 2014.

[3] Gumerov, N. A., O’Donovan, A. E., Duraiswami, R., and Zotkin, D. N., “Computation of the head-related transfer function via the fast multipole accelerated boundary element method and its spherical harmonic representation,” *The Journal of the Acoustical Society of America*, 127(1), pp. 370–386, 2010.

[4] Sridhar, R. and Choueiri, E., “A method for efficiently calculating head-related transfer functions directly from head scan point clouds,” in *Audio Engineering Society Convention 143*, Audio Engineering Society, in press.

[5] Watanabe, K., Iwaya, Y., Suzuki, Y., Takane, S., and Sato, S., “Dataset of head-related transfer functions measured with a circular loudspeaker array,” *Acoustical science and technology*, 35(3), pp. 159–165, 2014.

[6] Jin, C. T., Guillon, P., Epain, N., Zolfaghari, R., Van Schaik, A., Tew, A. I., Hetherington, C., and Thorpe, J., “Creating the Sydney York morphological and acoustic recordings of ears database,” *IEEE Transactions on Multimedia*, 16(1), pp. 37–46, 2014.

[7] Bomhardt, R., de la Fuente Klein, M., and Fels, J., “A high-resolution head-related transfer function and three-dimensional ear model database,” in *Proceedings of Meetings on Acoustics 172ASA*, volume 29, p. 050002, ASA, 2016.

[8] Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C., “The CIPIC HRTF database,” in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pp. 99–102, IEEE, 2001.

[9] Rugeles Ospina, F., Emerit, M., and Katz, B. F., “The three-dimensional morphological database for spatial hearing research of the BiLi project,” in *Proceedings of Meetings on Acoustics 169ASA*, volume 23, p. 050001, ASA, 2015.

[10] AES69-2015, “AES69-2015: AES standard for file exchange - Spatial acoustic data file format,” 2015.

[11] Majdak, P., Balazs, P., and Laback, B., “Multiple exponential sweep method for fast measurement of head-related transfer functions,” *Journal of the Audio Engineering Society*, 55(7/8), pp. 623–637, 2007.

[12] Farina, A., “Advancements in Impulse Response Measurements by Sine Sweeps,” in *Audio Engineering Society Convention 122*, 2007.

[13] Katz, B. F. and Noisternig, M., “A comparative study of interaural time delay estimation methods,” *The Journal of the Acoustical Society of America*, 135(6), pp. 3530–3540, 2014.