7 - 12 SEPTEMBER 2014, KRAKÓW
FORUM ACUSTICUM

Measurement of a head-related transfer function database with high spatial resolution

Thibaut Carpentier, Hélène Bahu, Markus Noisternig, Olivier Warusfel
UMR STMS IRCAM-CNRS-UPMC, 1 place Igor Stravinsky, 75004 Paris, France.

Summary
This paper describes a database of high spatial resolution head-related transfer functions (HRTF) measurements for 54 subjects (42 males, 12 females) and 3 dummy heads. The head-related impulse responses (HRIR) have been measured in IRCAM's anechoic chamber using the exponentional sweep sine technique and a sampling rate of 96 kHz. Microphones were positionned at the entrance of the blocked ear canal. The spatial sampling scheme is based on a Gaussian grid and includes 1680 directions with full azimuth range (0° to 360°), and elevation ranging from -51° to +86°. The angular step size is approximately 6 degrees in both dimensions. The subjects head position and orientation are tracked with an infrared optical motion capture system. The HRIR are publicly available in the standardized SOFA file format.

PACS no. 43.66.Pn, 43.60.Ek

1. Introduction

Head Related Transfer Functions (HRTFs) describe the linear filtering of a free-field sound from a given direction caused by the physical propagation and diffraction around the head, body and ears of a listener. When represented in time domain these functions are typically referred to as Head Related Impulse Responses (HRIR). They comprise the different sound localization cues and are thus essential for the design and the evaluation of spatial audio systems (e.g. binaural synthesis, room auralization, virtual reality applications, etc.).

Several research groups measured HRTF data and made them available for other researchers (see e.g. [1–8]). This work presents a new HRIR database measured on human subjects and three different dummy heads that provides:

- a large number of human subjects;
- a high resolution sampling grid;
- a high sampling rate;
- low harmonic distortions and signal-to-noise ratio well suited for binaural sound reproduction.

The article is organized as follows: Section 2 presents the measurement apparatus and the data acquisition procedure. Section 3 discusses the obtained HRIR data and the corresponding database.

2. Measurement setup

2.1. Infrastructure

All measurements were performed in IRCAM's full anechoic chamber. This sound isolated cuboid-shaped room (room-within-a-room structure supported on neoprene mounts; all interior surfaces and mechanical structures are covered with anechoic wedges) provides a useable volume of 103 m^3 ($5.7\times4.3\times4.2$ m^3), a lower limiting frequency of 75 Hz, and approximately 18.5 dB SPL (A Weighting) background noise. The anechoic chamber is equipped with a pivoting arc positioning system and a turntable that allow for positioning a sound source at any arbitrary position on a sphere around the head (see Fig. 1). The head is centered and aligned by means of three coincident laser beams pointing at the entrance of the ear canals and the rotation axis of the turntable, respectively. Head movements are tracked during measurements using an infrared camera motion capture system.

2.1.1. Turntable

Grating floor sections were installed in the center of the anechoic chamber to support a Brüel & Kjaer 9640 turntable system (see Fig. 1). This stepper motor turntable provides an angular resolution of 1° and is connected to a ball-bearing mounted rotary plate that supports the relatively high axial load. An adjustable-height stool is mounted on the plate, which is equipped with a slim back and neck rest. The rotation axis of the turntable is aligned by means of a ceiling-mounted vertical pointing laser beam.

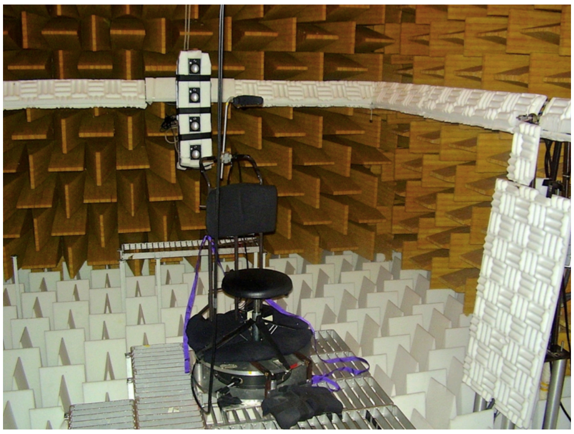

Figure 1: View of the anechoic chamber. Turntable with stool in the foreground; 4 loudspeakers mounted on the pivoting arc in the background.

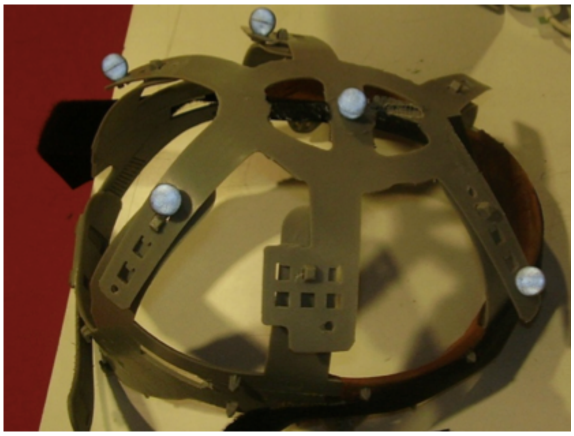

Figure 2: Safety helmet frame with reflective markers for head motion capture.

2.1.2. Pivoting arc

Measurement loudspeakers are mounted on a stepper-motor driven pivoting arc that is controlled by a 10 bit PIC microcontroller. It provides an angular resolution of $360/2^{10} \approx 0.35°$. During measurements elevation angles are monitored with two high-precision electronic angular position sensors. The axis of rotation is indicated by two laser beams, which are aligned with the left and right stepper motor axes, respectively. The position of the subject – sitting on the stool – is then adjusted such that his/her interaural axis is aligned with the arc's rotation axis. To damp the vibrations of the arc a tension belt is stretched along the mechanical structure. The structure itself is further covered with absorbent material to minimize acoustic reflections. The arc supports up to four loudspeakers at a distance of 2.06 m from the center of the measurement sphere (see Fig. 1).

2.1.3. Motion tracking system

Head movements were monitored during measurements using six OptiTrack V100:R2 infrared motion capture cameras. The motion tracking software can identify rigid bodies defined by a set of reflective markers. We mounted five reflective markers onto the inner frame of a safety helmet (see Fig. 2) that can be easily adjusted to a subject's head. The center marker was used as a reference point and was aligned with both the system's vertical rotation axis and the center of the subject's head. Prior to the actual measurement session the tracking system was re-calibrated each time when a rigid body was adjusted to a subject's head. An additional webcam was installed so that the experimenter could monitor the subject during measurements from outside the anechoic chamber.

2.2. Audio equipment

2.2.1. Loudspeakers

Non-coaxial two-way ELAC 301 loudspeakers were used for the HRIR measurements. The distance between tweeter and woofer is approximately 6 cm; the cross-over frequency is 3.2 kHz. The tweeter-woofer distance corresponds to an elevation-angle bias of less than 2° that will be neglected for the remainder of this paper. The loudspeakers are driven by a 4-channel Yamaha P2040 amplifier that delivers up to 20 watts into 8 Ω loads.

2.2.2. Ear molds and microphones

For each subject individualized silicon ear molds were fabricated by a hearing aid specialist. They were cut to fit to the external auditory meatus [3], varnished for rigidification, and drilled to hold the Knowles FG26107 C34 miniature microphones and cable connectors (see Fig. 3). The microphones are connected to a custom made low-noise preamplifier and signal conditioner with a stabilized voltage regulated power supply to reduce total harmonic distortions. Audio signals are then transmitted to a RME Fireface 800 digital audio interface, which is also used for measurement signal playback.

2.3. Software components

2.3.1. Architecture overview

The measurement environment is divided into three main functional units: (i) automatic sequence control and monitoring of the measurement process with Matlab, (ii) real-time audio processing and communication with peripheral audio devices with Max/MSP, and (iii) real-time head motion capture using Natural Point's TrackingTools software.

The software units communicate trough the following internet protocols: Matlab sends instructions and

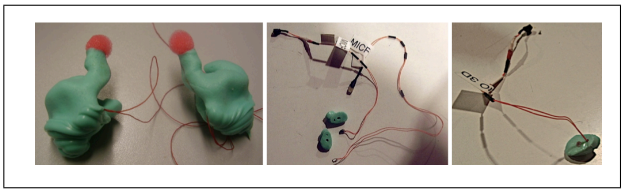

Figure 3: Different stages of the crafting of the ear molds: (left) silicon ear mold; (middle) drilled holes for microphone and cable connector; (right) ear mold with mounted microphone.

receives notifications ("error" or "success") to/from Max/MSP applying the Open Sound Control (OSC) protocol; the Virtual Reality Peripheral Network (VRPN) protocol is used for streaming the tracking data from Natural Point's interface to Max/MSP. The interaction in between the functional units is completely defined by the data exchange protocols and allows for the substitution of any element by a suitable replacement unit, which adheres to the defined protocol. The hardware and software architecture is illustrated in Fig. 4.

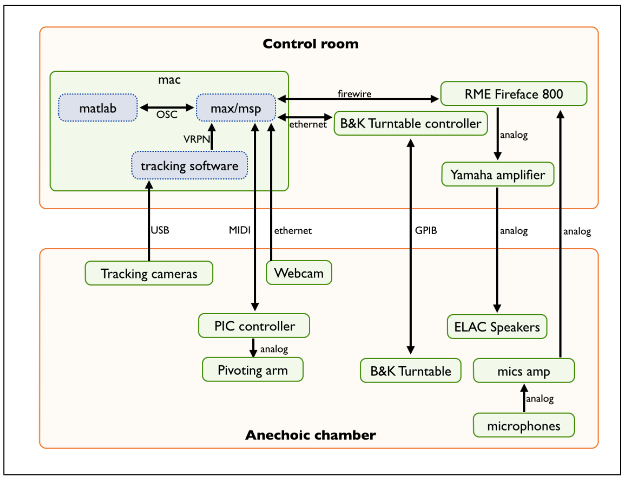

Figure 4: Overview of the measurement setup: hardware devices (green solid line); software components (blue dashed line).

2.3.2. Audio acquisition

HRIRs are measured with the exponential sweep sine method [9] at 96 kHz sampling rate. The benefits of sweep-based techniques over other acoustical measurements have been extensively discussed in literature (see e.g. [9, 10]). Basically, they allow for high signal-to-noise ratios (SNR), suppression of the harmonic distortion artifacts, and they are less vulnerable to the effects of time variance.

The sweep length is chosen to be 2^{16} samples (i.e. 683 ms) for a frequency range from 0 to 48 kHz. A 50 samples fade-out is applied to guarantee for a zero-crossing at the end of the signal. The length of the sweep is chosen as a trade-off between a good signal-to-noise ratio and the shortest possible duration of the overall measurement procedure. Short sweeps further reduce the risk of head movements during measurements so that the time-invariance hypothesis holds. A pause of 150 msec is set after the emission of each sweep; this pause is sufficiently long given (i) the typical length of HRIRs (about a dozen of milliseconds), and (ii) the low latency of the system.

To obtain the HRIR from sweep measurements the recorded signal has to be deconvolved with the excitation signal. The measurement system performs the deconvolution on the fly and allows for in-situ monitoring of the results. All computations are performed with double-precision floating point arithmetic; the numerical noise resulting from spectral inversion of the sweep is less than -110 dB. After deconvolution, the SNR is estimated through backward integration of the energy. Measurements with an estimated SNR below a threshold of 35 dB SNR are repeated. On average, SNRs are approximately 75 and 55 dB for ipsilateral and contralateral sides respectively (for 0° elevation). All data (raw recordings, sweep signal, HRIR data, etc.) are saved as 64-bit Matlab files along with miscellaneous textual metadata.

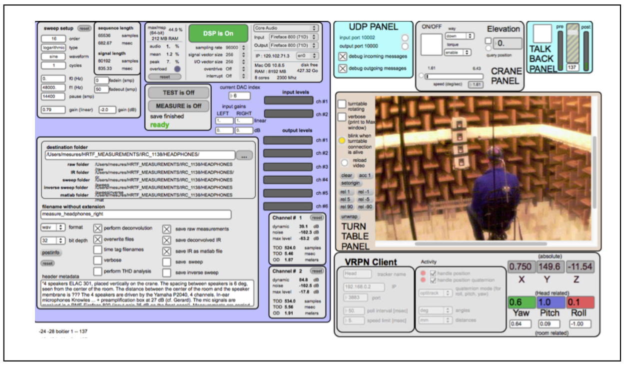

Figure 5: Experimenter interface in Max/MSP: (left) sweep settings, input/output signal vu-meters, real-time SNR estimations, etc.; (right) robotic arc elevation control, webcam monitoring, head-tracking data.

2.3.3. Control of mechanical devices

The B&K turntable system (type 9640) consists of a controllable turntable (type 5960) and a turntable controller (type 5997). A software plugin was developed to control the turntable from Max/MSP applying the GPIB (IEEE-488) protocol. The plugin receives acknowledgments from the B&K controller and unblocks HRIR measurements only when the turntable has reached the indicated position.

Max/MSP further controls the pivoting arc by sending MIDI commands to the PIC microcontroller. The actual angular position is monitored with the high-precision electronic angular position sensors that are connected to a mass balance. A control loop ensures that the arc reaches the target elevation.

3. Description of the database

3.1. Measurement grid

The Fourier-Bessel Expansion (FBE), or also Spherical Harmonics Expansion (SHE), of the HRTF data has been proven useful for many applications, such as spatial interpolation and range extrapolation (see e.g. [11] for further details). To accurately evaluate the discrete spherical wave spectrum of the measured HRTF data proper sampling of the sphere is indispensable. The number and distribution of sampling points limits the useable angular bandwidth, i.e. the maximum harmonic order N of the spherical data. Various spherical sampling schemes have been compared in [12]. Rafaely [13] shows that a Gaussian sampling grid with $2\,(N+1)^2$ sampling points allows for the exact computation of the spherical wave spectral coefficients. Although other sampling schemes offer the same property with less sampling points, they are often difficult to implement and/or it takes too much time to use them with a scanning array due to the non-regular angular distribution of points on (or even inside) the sphere.

When decomposing HRTFs onto a basis of discrete spherical harmonics, spatial aliasing limits the upper frequency for analysis to $kr < N$ (see e.g. [13]), where k is the wavenumber and r the radius of the listener's head. With a typical head radius of $r \simeq 10$ cm and an upper frequency limit of $f \simeq 16$ kHz the SHE order should be chosen $N \geq 29$. For the measured database, a nearby Gaussian grid with $N = 29$ was used. A Gaussian grid is uniformly distributed in azimuthal direction and close to a uniform angular distribution in elevation. The constant azimuthal step for the chosen order N is exactly 6°, which is compatible with the angular resolution of the B&K turntable. The elevation step is almost constant to approximately 5.9°. Therefore, the four available loudspeakers have been mounted with a vertical offset of 6°. This configuration allows for a good approximation of the theoretical positions (see Tab. I) while minimizing the number of displacements of the pivoting arc (which is rather slow). Due to practical constraints, such as the presence of a grating floor and the anechoic wedges on the ground floor, the lowest measured elevation is limited to −50.5°. This results in a polar gap of 360 non-measured directions (20%) with respect to the theoretical 29^{th} order Gaussian grid.

It is important to note that a 29^{th}-order Gaussian grid does not include data on the equator of the sphere. However, it is essential for many applications to provide HRTF data in the horizontal plane. Therefore, we performed an additional measurement with the pivoting arc at 0° elevation (measuring the four loudspeakers at -12, -6, 0 and 6° elevation). In summary the used measurement grid provides 1680 directions (see Fig. 6). With the described measurement setup, the data acquisition takes about 90 minutes per subject.

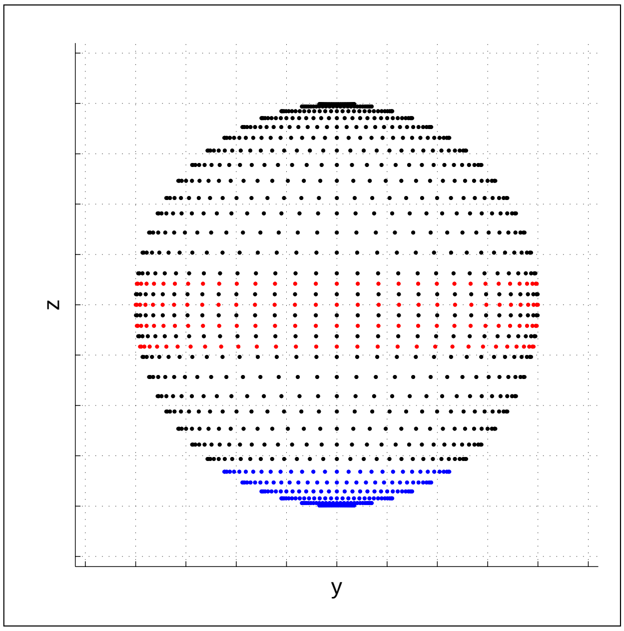

Figure 6: Measurement grid: (black dots) upper part of the Gaussian grid; (blue dots) lower part of the Gaussian grid that was not measured; (red dots) additional measurement points (not Gaussian).

3.2. Content of the database

This first release of the database includes head-related impulse responses for 54 subjects (42 men and 12 women), as well as 3 dummy heads (a Neumann KU 100, and a Brüel & Kjaer type 4100D with and without pinna). Free-field measurements of the loudspeakers and microphones responses are also available. Individual loudspeaker impulse responses were measured with a high-precision pre-polarized Brüel & Kjaer 1/2" free-field microphone (type 4189-L-001) and can be used for both free-field and diffuse-field equalization of HRTFs.

The downloadable package includes additional technical documentation and Matlab programs for screening the data. The database will be upgraded in the near future adding (i) HRIR data for additional subjects, (ii) anthropometric measurements, and (iii) high-precision 3D scans of the head and pinna for some of the subjects.

3.3. Data format

HRIR data are stored in the standardized AES-X212/SOFA "Spatial Acoustic Data File Format" (see e.g. [14] and http://www.sofaconventions.org for details). SOFA builds on netCDF (Network Common Data Form, http://www.unidata.ucar.edu/) and provides data compression, network transfer, file hierarchy, and partial data access over networks via OPeNDAP (Open-Source Project for a Network Data Access Protocol, http://www.opendap.org).

Data are made public-domain (for research and educational purposes) and are available at http://www.hrtf.ircam.fr.

4. Conclusions

We presented a new HRIR database measured for 54 human subjects that further provides tracking data for the actual head position during the measurements. All HRTF measurements were performed at the entrance of the blocked ear canals. This uniform database aims at enabling studies on interpersonal differences and facilitating research on individualized HRTFs.

Future work will focus on correcting the data according to the actual sound source directions, and completing the database with 3D meshes and anthropometric data. It is also planned to further investigate the impact of incomplete data (missing directions in the lowest part of the auditory sphere) on spherical harmonics analysis and modeling.

Acknowledgement

This work was funded by the French FUI project BiLi ("Binaural Listening", www.bili-project.org, FUI- AAP14) with support from "Cap Digital Paris Region".

References

[1] F.L. Wightman, D.J. Kistler: Headphone simulation of free-field listening (I): Stimulus synthesis. J. Acoust. Soc. Am. **85** (1989) 858 – 867.

[2] P. Majdak, M. J. Goupell, B. Laback: 3-D localization of virtual sound sources: effects of visual environment, pointing method, and training. Attention, Perception, and Psychophysics, **72** (2010) 454 – 469.

[3] H. Møller, M.F. Sørensen, D. Hammershøi, C.B. Jensen: Head-Related Transfer Functions of Human Subjects. J. Audio Eng. Soc, **43** (1995) 300 – 321.

[4] J. Blauert, C. Brueggen, A.W. Bronkhorst et al: The AUDIS catalog of human HRTFs. J. Acoust. Soc. Am. **103** (1998) 2901 – 2902.

[5] V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano: The CIPIC HRTF database. Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 99 – 102.

[6] B. Xie, X. Zhong, D. Rao, Z. Liang: Head-related transfer function database and its analyses. Science in China Series G: Physics, Mechanics and Astronomy **50** (2007) 267 – 280.

[7] O. Warusfel: Listen HRTF Database, 2003. recherche.ircam.fr/equipes/salles/listen

[8] J.G. Bolaños, V. Pulkki: HRIR database with measured actual source direction data. Proc. 133rd Convention of the Audio Engineering Society, 2012.

[9] A. Farina: Simultaneous measurement of impulse response and distorsion with a swept-sine technique. Proc. 108th Convention of the Audio Engineering Society, 2013.

[10] S. Müller, P. Massarani: Transfer-function measurement with sweeps. J. Audio Eng. Soc. **49** (2001) 443 – 471.

[11] M. Pollow, K.-V. Nguyen, O. Warusfel, T. Carpentier, M. Müller-Trapet, M. Vorländer, M. Noisternig: Calculation of head-related transfer functions for arbitrary field points using spherical harmonics decomposition. Acta Acustica united with Acustica, **98** (2012) 72 – 82.

[12] M. Noisternig, F. Zotter, B. F. G. Katz: Reconstructing sound source directivity in virtual acoustic environments. In : Principles and Applications of Spatial Hearing. Y. Suzuki, D. S. Brungart, H. Kato (eds.). World Scientific Publishing, 2011.

[13] B. Rafaely: Analysis and design of spherical microphone arrays. IEEE Transactions on Speech and Audio Processing, **13** (2005) 135 – 143.

[14] P. Majdak, R. Nicol, T. Carpentier, Y. Suzuki, H. Wierstorf, H. Ziegelwanger, M. Noisternig: Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions. Proc. 134th Convention of the Audio Engineering Society, 2013.

Table I: Theoretical elevations (for 29th order Gaussian grid) versus actually measured elevations. Black font: upper part of the Gaussian grid; blue font: lower part of the Gaussian grid that was not measured; red font: additional measurement points (not Gaussian). All values are expressed in degrees.

theoretical	-85.4826	-79.6307	-73.7443	-67.8500	-61.9528	-56.0542	-50.1547	-44.2548	-38.3546
actual	-	-	-	-	-	-	-50.5	-44.5	-38.5
deviation	-	-	-	-	-	-	0.3453	0.2452	0.1454
theoretical	-32.4541	-26.5535	-20.6528	-14.7521	-8.8513	-2.9504	2.9504	8.8513	14.7521
actual	-32.5	-26.5	-20.5	-14.5	-8.5	-3.0	3.0	9.0	15.0
deviation	0.0459	0.0535	0.1528	0.2521	0.3513	0.0496	0.0496	0.1487	0.2479
theoretical	20.6528	26.5535	32.4541	38.3546	44.2548	50.1547	56.0542	61.9528	67.8500
actual	20.5	26.5	32.5	38.5	44.0	50.0	56.0	62.0	67.5
deviation	0.1528	0.0535	0.0459	0.1454	0.2548	0.1547	0.0542	0.0472	0.3500
theoretical	73.7443	79.6307	85.4826	-	-	-	-		
actual	73.5	79.5	85.5	-12.0	-6.0	0.0	6.0		
deviation	0.2443	0.1307	0.0174	-	-	-	-		