A SINGLE-AZIMUTH PINNA-RELATED TRANSFER FUNCTION DATABASE

Simone Spagnol,
Department of Information Engineering
Università degli Studi di Padova
Padova, Italy
spagnols@dei.unipd.it

Marko Hiipakka,
Aalto University School of Electronical Engineering
Department of Signal Processing and Acoustics
Espoo, Finland
Marko.Hiipakka@aalto.fi

Ville Pulkki,
Aalto University School of Electronical Engineering
Department of Signal Processing and Acoustics
Espoo, Finland
Ville.Pulkki@aalto.fi

ABSTRACT

Pinna-Related Transfer Functions (PRTFs) reflect the modifications undergone by an acoustic signal as it interacts with the listener's outer ear. These can be seen as the pinna contribution to the Head-Related Transfer Function (HRTF). This paper describes a database of PRTFs collected from measurements performed at the Department of Signal Processing and Acoustics, Aalto University. Median-plane PRTFs at 61 different elevation angles from 25 subjects are included. Such data collection falls into a broader project in which evidence of the correspondence between PRTF features and anthropometry is being investigated.

1. INTRODUCTION

Anthropometric features of the human body have a key role in Head-Related Transfer Function (HRTF) characterization. While non-individualized HRTFs represent a cheap and straightforward mean of providing 3-D perception in headphone reproduction, listening to non-individualized spatialized sounds may likely result in evident sound localization errors such as incorrect perception of elevation, front-back reversals, and lack of externalization [1], especially when head tracking is not utilized in the reproduction [2]. On the other hand, individual HRTF measurements on a significant number of subjects may be both expensive and inconvenient.

Structural modeling of HRTFs ultimately represents an attractive solution to these shortcomings. As a matter of fact, if one isolates the contributions of the listener's head, pinnae, ear canals, shoulders, and torso to the HRTF in different subcomponents - each accounting for some well-defined physical phenomenon - then, thanks to linearity, one can reconstruct the global HRTF from a proper combination of all the considered effects. Relating each subcomponent's temporal and/or spectral features (in the form of digital filter parameters) to the corresponding anthropometric quantities would then yield a HRTF model which is both economical and individualizeable [3].

Following such structural assumption, the present work focuses on the pinna contribution to the HRTF. It is undoubted that the pinna plays a primary part in the perception of source elevation; even so, the relation between acoustic phenomena due to the pinna - mainly resonances and sound reflections [4] - and anthropometry is yet not fully understood. In [5] a promising correspondence between reflection points on pinna surfaces and frequencies of notches occurring in the high-frequency range of the HRTF spectrum was informally found by analyzing median-plane responses from the CIPIC database [6] along with photos of four subjects' pinnae. Still, in order for a more extensive and accurate analysis to be carried out, an alternative database is needed, the most relevant reasons being:

- the presence in the CIPIC HRTFs of the head and shoulders' contributions, which cannot be fully eliminated a posteriori;
- the need of highly detailed photographs of the subjects' pinnae for an effective image processing method that extracts the possible reflection contours to be designed;
- the public unavailability of the remaining subjects' pinnae photographs, necessary to perform such an analysis.

This paper presents such a database, which we refer to as Pinna-Related Transfer Function (PRTF) Database, primarily focusing on the choices and tools through which the final responses were collected. Furthermore, some early results on the so obtained PRTFs' features will be sketched. All of the following work was carried out at the Department of Signal Processing and Acoustics, Aalto University. The database, accompanied by detailed photographs of the subjects' pinnae and the measurement setup, is publicly downloadable from the first author's website at http://www.dei.unipd.it/~spagnols/PRTF_db.zip as a .zip archive.

2. MEASUREMENT PROCEDURE AND APPARATUS

In an ideal situation, the PRTF is the response of the pinna mounted on an infinite plane [7]. In our measurements, an ad hoc pinna isolation device that approximates the ideal case was built and used. The test subjects' torsos and shoulders were isolated by a 1-m $\times$ 1-m, 15-mm thick wooden board having a 24-cm-diameter circular hole in the middle of it that approximately fits the size of the human head. A polycarbonate sheet with grinded edges and a 6-cm-diameter circular hole in the middle was fixed with a dozen

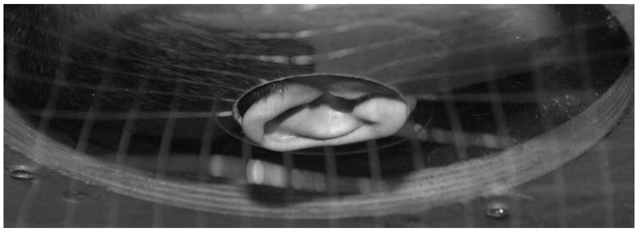

Figure 1: *The pinna isolation device used for PRTF measurements.*

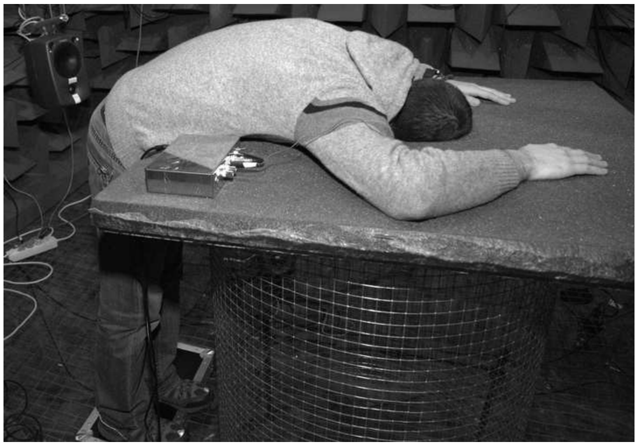

Figure 2: *Subject position during the measurements. On top left, the boom-controlled loudspeaker used for sweep reproduction.*

flat head screws to the board in order to completely cover the hole for the head while letting the subjects' right pinnae come out of the other side of it (see Figure 1). Furthermore, a thick layer of foam with a head-profile-shaped cut in the middle was glued to the upper side of the board with the purpose of adding comfort to the subjects. A piece of such layer could be taken off accordingly with the specific subject's build.

The isolation device was brought right in the middle of an anechoic chamber and placed over an acoustically transparent, one meter high cylindrical metallic fence having 1.75mm thread width in order to avoid reflections from prospective table legs. A controlled boom mounted on the room's ceiling had the purpose of moving the sound source (a Genelec 8030A loudspeaker) along a circumference centered in the pinna hole and laying on the plane parallel to the isolation device. The loudspeaker was positioned upside down, so that the woofer was at the level of the forementioned plane while the tweeter was under it, allowing high-frequency components to directly join the pinna hole without reflecting on the border of the isolation device. Furthermore, the distance between the loudspeaker and the pinna hole was approximately 1.6 m, so we can assume the incident wave to be plane for frequencies above 3 kHz (the loudspeaker's crossover frequency). This assumption may not be guaranteed below 3 kHz, yet the relative little importance of pinna features below this threshold makes this problem negligible. Since the boom was not acoustically transparent and other loudspeakers were fixed to the chamber's walls, let us label the environment as low-echoic rather than anechoic. In spite of this, as Subsection 2.1 will mention, all the data will be adequately windowed so as to discard reflections occurring on the

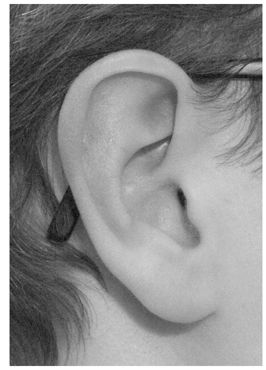

Figure 3: *Subject* 08*'s right pinna.*

room's equipment.

25 subjects (18 men and 7 women), mostly students and staff of Aalto University, participated to the measurements. A Knowles FG-23329 microphone carefully stuffed in the middle of a hollow earplug was placed inside the right ear canal of each subject in turn. Then, the subject was asked to stand in front of one side of the panel (eventually with the help of a pedestal to let his waist reach the level of the isolation device), bend 90 degrees forwards and lay his head on the right side in order to let his pinna pass the hole (see Figure 2). The required 90° head-neck rotation could be reached thanks to the thick layer of foam which allowed the right shoulder to sink at a lower level than the left. This way, the plane spanned by the loudspeaker's rotation approximately corresponded to the subject's median plane. The pinna position was then adjusted both by instructing the subject on how to move his head and by manual intervention through a big hole in the fence. Finally, vertical orientation was adjusted by manually rotating the subject's head to let his ear axis point at a precise mark on one of the chamber's walls. Subjects were told to remain as still as possible, yet their movements were not monitored during the actual measurement session.

The responses were measured via the logarithmic sweep (or logsweep) method [8]. The used sine sweep had 48 kHz sampling frequency, 1 s duration, and spanned the frequency range from 20 Hz to 22 kHz. By controlling the boom rotation and sweep reproduction from a Max/MSP patch running on a workstation just outside the anechoic chamber, sweep responses for 61 different elevation angles were recorded at 48 kHz sampling frequency in approximately six minutes' time. The selected elevation angles, considering the interaural-polar coordinate system (see [6]), spanned the range from -60 to 240 degrees at 5-degree steps. The boom constantly rotated during the measurements, hence high frequencies were measured from a slightly different elevation than low frequencies. However, since the angular speed was almost constantly less that one degree per second, the impact on measurements looks negligible.

In addition, free-field responses were taken by placing the microphone-stuffed earplug inside of a small foam cut, positioning it in the middle of the pinna hole of the isolation device, and repeating the measurement procedure in the same way as for the test subjects.

Pictures of the subjects' right pinnae were also taken before or after the measurements (see e.g. Figure 3). The distance and orientation of the camera with respect to the pinna was kept as

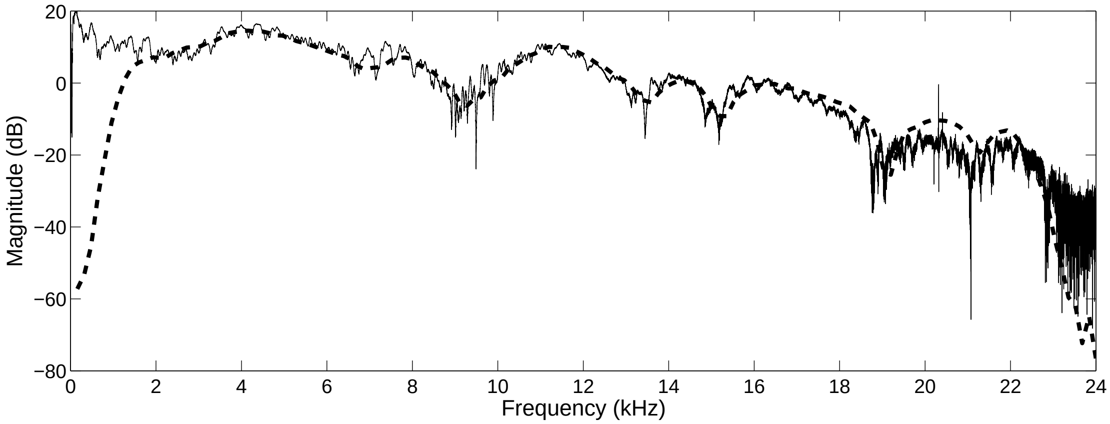

Figure 4: *Original sweep magnitude response (solid line) and post-processed PRTF magnitude (thick dashed line).*

constant among subjects as possible through the help of a tripod. Also, each subject's pinna height (variable d_5 in [6]) was measured and tracked down for resizing purposes. This information, along with each subject's sex and evidenced anomalies in the experiment with respect to the optimal situation, can be found in the online database. As for anomalies, Subject 06's pinna did not completely pass the hole, Subject 13 had a piercing on the helix which could not be taken off, and Subject 18 had the earplug slightly displaced at the end of the measurements.

2.1. Data post-processing

According to the logsweep method, inverse filtering was performed on the measured sweeps (including free-field sweeps) in order to obtain the corresponding impulse responses. Specifically, the inverse response of the excitation signal was first computed and then low-passed and high-passed with fifth-order digital Butterworth filters to compensate for the original zero sound pressure level below 20 Hz and above 22 kHz in the sweep signal. Since the pinna has no effect below 3 kHz and sounds above 15-20 kHz are hardly perceptible by humans we let the high-pass and low-pass Butterworth filters' cutoff frequency be loose, that is 1.2 and 21.6 kHz respectively. Hence, each impulse response was calculated by convolving such band-passed inverse filter with the measured sweep.

Subsequently, a 300-sample Hann window was applied to each impulse response with the aim of cutting off late reflections possibly occurring on the subject's legs, the pedestal, or the room equipment. The window was centered in the first positive peak p exceeding a heuristic amplitude threshold in the impulse response, so that the windowed impulse lasts approximately 3 ms from p.

Finally, free-field compensation of the subjects' impulse responses had to be performed. To this end, for each elevation e, a 10^{th}-order minimum-phase IIR filter which approximates the magnitude of the inverse free-field response at source elevation e was designed through the least-squares fit procedure provided by the Yule-Walker method of ARMA spectral estimation [9]. As we expected, all free-field responses had similar and almost flat - except for a ripple around 2.5 kHz probably due to the loudspeaker's crossover frequency - magnitude plots, with no tangible diffraction occurring on the wooden board. This result certifies the transparency of the measurement setup. Straightforward filtering of the subject's impulse response at elevation e through the so built IIR filter gave the free-field compensated, final pinna-related impulse response (PRIR) that is stored in the database.

Figure 4 shows the magnitude plots of an original recorded sweep and the corresponding post-processed PRTF. It can be clearly seen how the general notch/resonance structure of the typical PRTF is preserved, excluding the very upper and lower frequency ranges which are, however, not of interest to us.

3. EARLY RESULTS AND DISCUSSION

It should be mentioned that, since data was collected for a single azimuth value only, there is no guarantee that integrating a future pinna model based on these responses in a complete structural model would give an appropriate representation of the HRTF. In other words, the PRTF for elevation e and azimuth 0° may have a totally different look than the PRTF for elevation e and e.g. azimuth 60°. However, informal inspection of different HRTF sets revealed how generally there is no pronounced variation in median-plane reflection and resonance patterns across the angular range when the azimuth's absolute value is increased from 0° up to about 30°. Hence we may assert that under the assumption that the source moves in the vicinity of the median plane, pinna effects solely depend on source elevation.

Through direct inspection of the PRTF magnitude plots of all 25 subjects, a couple of observations can be made. First, when the source is ahead of the frontal plane (in our case when $-60^\circ \leq e < 90^\circ$), the PRTF behaviour is quite complex and greatly varies from subject to subject. However, commonly known features evidenced in previous works on PRTFs [10, 11], such as the 4-kHz omnidirectional resonance mode and the notch whose frequency ($6 - 10$ kHz) increases with elevation, appear in the vast majority of subjects (see e.g. Subject 08 in Figure 5). In some cases (e.g. Subject 15), however, the reflection structure is unclear, the magnitude plot presenting valleys which happen to be excessively shallow.

Secondly, while all PRTFs greatly differ among subjects when the source is ahead of the frontal plane, their behaviour is similar for all other elevations. Specifically, allowing some degree of approximation:

- for $90^\circ \leq e \leq 125^\circ$ the majority of PRTFs show a descending magnitude plot with one major resonance around

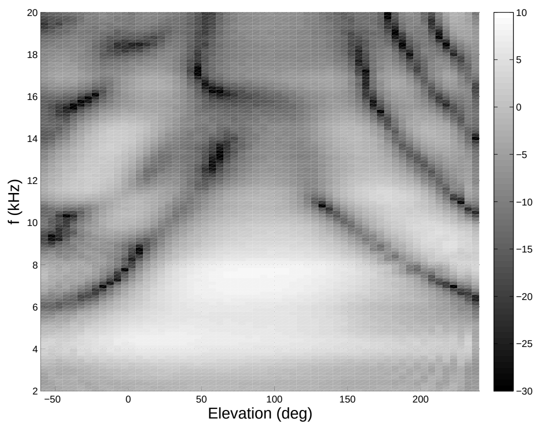

Figure 5: *Magnitude plot (in dB) of Subject* 08*'s PRTF.*

7 kHz and no evident notches;

- from about $e = 130^\circ$ one frequency notch appears at around 10 kHz, eventually followed by others at higher frequencies when the source is about to cross the horizontal plane at $e = 180^\circ$ (this notch was found in [11] too);
- PRTFs for the last elevation angles, especially $e = 240^\circ$, show a more complex magnitude structure with 3 or more notches below 15 kHz (also reported in [11]).

These features can all be detected in Figure 5. The absence of evident notches when the source is above the listener may easily be attributed to the presence of the helix which "masks" the concha, evading direct reflections on it. Conversely, the presence of complicated patterns at $e = 240^\circ$ comparable to those for sources ahead of the frontal plane may be both attributed to reflections on different pinna contours such as the upper part of the helix, the tragus or the crus helias, or to possible unwindowed reflections on the subject's legs.

Finally, even after post-processing some PRTFs still present a "noisy" spectrum. This artifact may likely be associated to subjects' slight movements during the sweep reproduction or to a rattling noise coming from the metallic fence which was reported by a few subjects right after their measurement session. However, besides being isolated cases only, the main features of PRTFs remain preserved.

4. CONCLUSIONS AND FUTURE WORK

A database of Pinna-Related Transfer Functions was presented in this paper. The measurement setup and procedure was described in details, along with the polishing operations applied to obtain the final PRTFs from the measured responses. The early results and assumptions traced in the last section need of course to be further investigated, especially for what concerns PRTF behaviour in the elevation range $-60^\circ \le e \le 90^\circ$ where pinna modifications happen in greater number. Future work includes adaptation of a separation algorithm [12] that extracts the reflective and resonant components from each PRTF to the present database in order to analyze each component separately and study the relation between its features and anthropometry, the final aim of such work being customization of a structural HRTF model.

5. ACKNOWLEDGMENTS

The first author would like to acknowledge the entire spatial sound research group at Aalto University for the constant support throughout setting up and performing this experiment, especially Javier Gomez Bolaños for his precious assistance in photography issues. This work has been supported by The Academy of Finland and by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement No. 240453.

6. REFERENCES

[1] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, 1996.

[2] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *J. Acoust. Soc. Am.*, vol. 42, no. 2, pp. 480–488, August 1967.

[3] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp. 476–488, 1998.

[4] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Series B, Biological Sciences*, vol. 168, no. 1011, pp. 158–180, August 1967.

[5] S. Spagnol, M. Geronazzo, and F. Avanzini, "Fitting pinna-related transfer functions to anthropometry for binaural sound rendering," in *IEEE International Workshop on Multimedia Signal Processing*, Saint-Malo, France, October 2010, pp. 194–199.

[6] R. V. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, 2001, pp. 1–4.

[7] R. V. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTFs," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, 2001, pp. 103–106.

[8] S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 443–471, June 2001.

[9] B. Friedlander and B. Porat, "The modified Yule-Walker method of ARMA spectral estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-20, no. 2, pp. 158–173, March 1984.

[10] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Am.*, vol. 44, no. 1, pp. 240–249, 1968.

[11] Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *Journal of Sound and Vibration*, vol. 300, no. 3-5, pp. 552–579, 2007.

[12] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and modeling of pinna-related transfer functions," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 2010.