HRTF Database at FIU DSP Lab

Navarun Gupta[1], Armando Barreto[2], Manan Joshi[1] and Juan Carlos Agudelo[3]

Department of Electrical and Computer Engineering[1]
University of Bridgeport, Bridgeport, CT 06604

Electrical and Computer Engineering Department[2]
Florida International University, Miami, FL 33174

Wells Fargo & Company[3]
Charlotte, N.C.

ABSTRACT

Head Related Transfer Functions (HRTFs) are signal processing models that represent the modifications undergone by the acoustic signal as it interacts with the listener's body. HRTFs can be used to generate binaural sound as they contain all the information about the sound source's location. This paper describes a database of head-related transfer functions based on ear/head measurements measured at the FIU DSP Lab. HRTF data from 15 subjects at 12 different azimuths and 6 different elevations are included. This database also includes 3-D images of the subject's pinnae (outer ears). Anthropometric measurements of the various parts of the pinna are included.

Index terms – HRTF, binaural, database, pinna, HRIR

1. INTRODUCTION

HRTFs are signal processing models that represent the modifications undergone by the acoustic signal as it travels from a sound source to each of the listener's eardrums. These modifications are due to the interaction of the acoustic waves with the listener's torso, shoulders, head and pinnae, or outer ears. As such, HRTFs are somewhat different for each listener. For a listener to perceive synthesized 3-D sound cues correctly, the synthesized cues must be similar to the listener's own HRTFs.

One can measure individual HRTFs using specialized recording systems; however, these systems are prohibitively expensive and restrict the portability of the 3-D sound system. HRTF-based systems also face several computational challenges. To overcome these problems, sometimes generic HRTFs are used [1]. Due to the loss of individual characteristics, binaural sounds generated using generic HRTFs suffer from higher errors and lower accuracy in localization. Another approach is to customize HRTFs based on anthropomorphic measurements. The sound entering the pinna undergoes several reflective, diffractive and resonant phenomena, which determine the HRTF. Using signal processing tools and statistical analysis, empirical equations can be derived that describe how the HRTFs can be determined by the shape and size of the pinna, head and torso. Pinna plays an important role in localization of sound in the frontal and median plane. Elevation effects can be created by modeling HRTFs based on pinna measurements [2]. The database at FIU contains 3-D images of the pinnae of 15 subjects and their corresponding individual HRTFs. This database can be used to model HRTFs based on the various measurements of the pinna such as the concha volume, concha area and helix length (see fig. 4).

2. MEASUREMENTS

2.1. HRTF Measurement

The effective empirical measurement of HRTF pairs is carried out in the following sequence: A speaker is placed at known relative positions with respect to the subject for whom the HRTFs are being determined, and a known, broad-band audio signal is used as excitation. At FIU's DSP laboratory, the Ausim3D's HeadZap HRTF Measurement System [3] is used. This system measures a 256-point impulse response for both the left and the right ear using a sampling frequency of 96 KHz. Golay codes are used to generate a broad-spectrum stimulus signal delivered through a Bose Acoustimass speaker. The response is measured using miniature blocked meatus microphones placed at the entrance to the ear canal on each side of the head. Under control of the system, the excitation sound is issued and both responses (left and right ear) are captured. Since the Golay code sequences played are meant to represent a broad-band excitation equivalent to an impulse, the sequences captured in each ear are the impulse responses corresponding to the HRTFs. Therefore these responses are called Head-Related Impulse Responses (HRIRs). The system provides these measured HRIRs as a pair of 256-point minimum-phase vectors, and an additional delay value that represents the Interaural Time Difference (ITD).

AUSIM3D HeadZap system is designed for use in reflective, noisy settings typical of offices and laboratories. Instead of using an array of speakers, HeadZap uses a single speaker mounted on a vertically sliding arm. The loudspeaker's position is fixed for every ring of measurements around the subject, changing only every time measurements for a new elevation are needed. The subject is seated on a rotating stool and adjusts his/her position according to the desired HRTF location.

To increase the signal-to-noise ratio (SNR), the subject is outfitted with blocked meatus microphones which gather considerably more signal than do probe microphones, and permit louder test sequences to be used. To overcome the problem of reflection from the walls and other objects in the measurement room, HeadZap only extracts the direct path contribution from the measurements. The measurements are carried out in a room padded with foam to minimize reflections. Extraction of raw HRTF data is performed by windowing the measurements such that no room reflections are incorporated in the compilation of the final data.

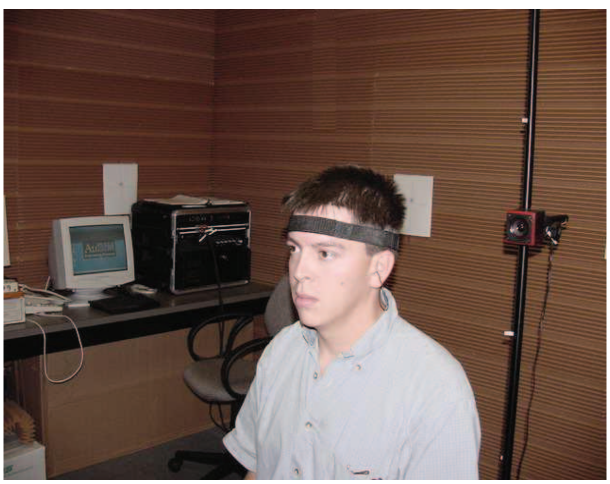


Figure 1. HRTF measurements using the HeadZap System at FIU's DSP Lab

The measurement sequence must be completed for each of the source locations for which an HRIR pair is desired. The different sound source positions are characterized by two descriptive angles of the spherical coordinate system: azimuth and elevation. The horizontal plane is divided into twelve sections, resulting in HRTFs separated by 30° in azimuth. The vertical plane is divided into six intervals at elevations of: +54°, +38°, +18°, 0°, -18° and –36°. Thus, for each subject, 72 impulse responses are measured using HeadZap.

2.2. Measurements of the Pinna

The outer ear (pinna) is a complex shaped organ responsible for shaping much of the HRTF in the higher frequencies. The folds of the pinnae cause minute time delays within the range of 0-300 μs [2] that cause the spectral content at the eardrum to differ significantly from that of the sound source. The complex shape of the pinna causes reflections, diffractions and resonances that give rise to a unique HRTF for each individual. The concha is the most important part of the outer ear when it comes to reflections and resonances. To measure the concha's depth and volume, a 3D scanner was used.

Figure 2. Handheld Scanner being used for making a 3D image of a subject's ear.

The FastScan handheld laser scanner from Polhemus was ideal for scanning the pinna. The scanning involved smoothly sweeping the "Wand" over the subject's ears, in a manner similar to spray painting. Because the subject's head might move during the scanning, the receiver of this scanner is fixed to the subject's head, and it becomes the reference point while scanning the ear. The scanner digitizes the three-dimensional surface of the ear, and after processing the data, the resulting 3D scan of the ear is exported in Autocad (.dxf) and 3D Studio max (.3ds) format.

The resolution of the scanner is 1.0 mm under ideal conditions [4]. However, depending on the scanning environment, and other factors, this accuracy is seldom achieved. Smoothing the data before exporting it also reduces the resolution of the scan. On average, the scans of the ears achieved an accuracy of about 2.5 mm.

To measure the various parts of the ear, AutoCad and 3D-Studio max were used. Autocad was useful in measuring the area of the concha and the 2-D measurements (such as lengths and widths). 3-D Studio max was used to measure the volume of the concha. One useful feature of the 3D Studio max was the “Cap Holes”. This was used frequently to patch up any “holes” in the scan before measuring the volume.

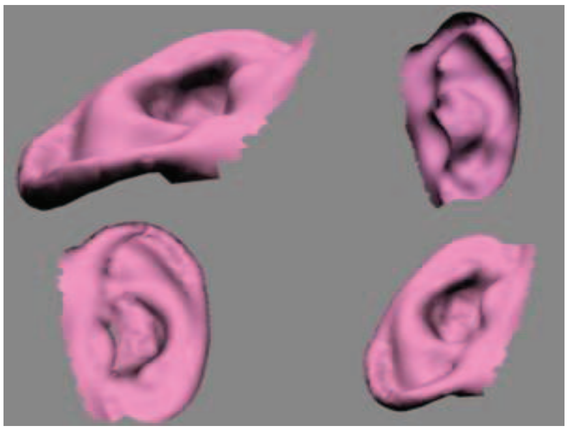
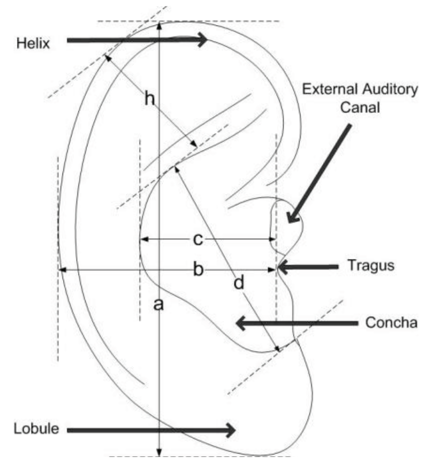


Figure3. 3D images of the ear from various perspectives

3. TEST SUBJECTS

Fifteen students from FIU were used as test subjects. The HRIRs of these subjects were measured using FIU’s Headzap system. Key anthropometric features of the ears of the 15 experimental subjects in the study were captured by means of digital photography (including a distance reference), and laser 3-D scanning, using a Polhemus FastScan handheld scanner. Figure 3 shows a sample 3-D reconstruction from a 3-D scanned file, and a schematic drawing identifying some of the key anthropometric features estimated for each of the 30 ears involved in this study.

(a)



(b)

Figure 4. Anthropometric Characterization of subject ears (a) Sample rendering of a 3-D laser scan; (b) Definition of key anthropometric features

Table 1 provides the names of the anthropometric features indicated in Figure 4, indicating the average values calculated for the 30 ears measured in this study and also the average values for those same parameters from the ears studied in the development of the CIPIC HRTF Database [5]. Measurement of parameters such as Concha Area, Concha Depth and Concha Volume required the use of 3-D modeling programs, such as 3D Studio Max and Autocad.

Identifier in Schematic	Abbreviation Equations	Feature Name	Average This Study	CIPIC average
a	E_L	Ear Length	6.90 cm	6.41 cm
b	E_W	Ear Width	3.25 cm	2.92 cm
d	C_H	Concha Height	2.75 cm	2.59 cm
c	C_W	Concha Width	1.89 cm	1.58 cm
No ID	C_A	Concha Area	3.41 cm^2	Not available
No ID	C_V	Concha Volume	2.72 cm^3	Not available
No ID	C_D	Concha Depth	0.82 cm	1.02 cm
h	H_L	Helix Length	2.40 cm	1.51 cm

Table 1. Average and Identifiers for the key anthropometric features of the ear used

4. HRTF DATA

The FIU DSP Lab HRTF Database is located at: http://dsp.eng.fiu.edu/HRTFDB/

Users can download the HRTF files in zip format from the database. The zip files contain HRTF data from 15 subjects. Once unzipped, the following files can be retrieved: Az180, Az150, Az120, Az90, Az60, Az30, Az00, Az_30, Az_60, Az_90, Az_120 and Az_150. Each file contains HRIRs (Head-Related Impulse Responses) from an azimuth that is reflected in the number next to Az (for example; file Az120 has HRIRs from Azimuth 120). These files can be opened using Microsoft Word or Matlab.

For each Azimuth, there are six elevations (54°, 36°, 18°, 0°, -18° and -36°). At each elevation there are two HRIRs (left and right ear). Each HRIR contains 256 samples (sampled at 96 KHz). Each file also contains the ITD. ITD stands for interaural time difference. This is the time delay (in samples) between the sound's arrival between the left and the right ear. Thus, each Azimuth file has 6 (elevations) * 2 (Left and Right) * 256 (points in HRIR) + 6 (ITDs) = 3078 points.

5. CONCLUSION

FIU DSP Lab HRTF database is unique in the sense that it contains detailed anthropometric data about the pinnae. This data was also sampled at 96 KHz – a significantly higher rate than the other databases [5]. The study of HRTFs is a rapidly growing area with potential uses in virtual environments, auditory displays, entertainment industry, human-computer interface for visually impaired, aircraft warning systems and many others. This database can be used to improve the quality and accessibility of HRTFs by creating customizable HRTFs based on pinna measurements.

6. REFERENCES

[1] Gardner, Bill and Martin, Keith. "HRTF measurements of a KEMAR dummy-head microphone", MIT Media Lab, Perceptual Computing, 1994.

[2] Batteu, D. W., "The Role of the Pinna in Human Localization", Proc. R. Soc. Lon. B., vol. 168, pp. 158-180, 1967.

[3] AuSIM, Inc., "HeadZap: AuSIM3D HRTF Measurement System Manual", AuSIM, Inc., 4962 El Camino Real, Suite 101, Los Altos, CA 94022, 2000.

[4] Polhemus, Handheld Laser Scanner, http://www.polhemus.com/fastscan.htm

[5] Algazi, V. R., Duda, R. O., Avendano, C., "The Cipic HRTF database", IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2001.