A Customizable Model of Head-Related Transfer Functions Based on Pinna Measurements

Navarun Gupta navarung@bridgeport.edu
Assistant Professor, Department of Electrical Engineering
University of Bridgeport
Bridgeport, CT 06604, USA

Armando Barreto barretoa@fiu.edu
Associate Professor, Department of Electrical Engineering
Florida International University
Miami, FL 33174, USA

Abstract

This paper proposes a method to model Head-Related Transfer Functions (HRTFs) based on the shape and size of the outer ear. Using signal processing tools, such as Prony's signal modeling method, a dynamic model of the pinna has been obtained, that completes the structural model of HRTFs used for digital audio spatialization.

Listening tests conducted on 10 subjects showed that HRTFs created using this pinna model were 5% more effective than generic HRTFs in the frontal plane. This model has been able to reduce the computational and storage demands of audio spatialization, while preserving a sufficient number of perceptually relevant spectral cues.

Keywords: HRTF, Binaural, HRIR, Pinna, Model

1. INTRODUCTION

HRTFs represent the transformation undergone by the sound signals, as they travel from their source to both of the listener's eardrums. This transformation is due to the interaction of sound waves with the torso, shoulder, head and outer ear of a listener [9]. Therefore, the two components of these HRTF pairs (left and right) are typically different from each other, and pairs corresponding to sound sources at different locations around the listener are different. Furthermore, since the physical elements that determine the transformation of the sounds reaching the listener's eardrums (i.e., the listener's head, torso and pinnae), are somewhat different for different listeners, and so should be their HRTF sets [2].

Currently, some spatialization systems make use of HRTFs that are empirically measured for each prospective user. These "custom" HRTFs are anthropometrically correct for each user, but the equipment, facilities and expertise required to obtain these "measured HRTF pairs", constrain their application to high-end, purpose-specific sound spatialization systems only [2]. For most consumer-grade applications, sound spatialization systems resort to the use of "generic" transfer functions, measured from a manikin with "average" physical characteristics [7], which, evidently is a fundamentally imperfect approach.

This paper reports on our work to advance an alternative approach to sound spatialization, based on the postulation of anthropometrically-related "structural models" [6] that will transform a single-channel audio signal into a left/right binaural spatialized pair, according to the sound source simulation. Specifically, the work reported here proposes linkages between the parameters of the HRTF model and key anthropometric features

of the intended listener's pinna, so that the model, and consequently the resulting HRTFs are easily "customizable" according to a small set of anthropometric measurements.

2. MEASUREMENTS AND IMPLEMENTATION

Current sound spatialization systems use HRTFs, represented by their corresponding impulse response sequences, the Head-Related Impulse Responses, (HRIRs) to process, by convolution, a single-channel digital audio signal, resulting in the two components (left and right) of a binaural spatialized sound. When these two channels are delivered to the listener through headphones, the sound will seem to emanate from the source location corresponding to the HRIR pair used for the spatialization process [4].

In our laboratory, we use the Ausim3D's HeadZap HRTF Measurement System [1]. 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. 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), i.e., the additional delay observed before the onset of the response collected from the ear that is farthest from the speaker position. In addition to the longer onset delay of the response from the "far" or "contralateral ear" (with respect to the sound source), this response will typically be smaller in amplitude than the response collected in the "near" or "ipsilateral ear". The difference in amplitude between HRIRs in a pair is referred to as the Interaural Intensity Difference (IID).

Our protocol records HRIR pairs from source locations at the 72 possible combinations of ϕ = {-36°, -18°, 0°, 18°, 36°, 54°} and θ = {0°, 30°, 60°, 90°, 120°, 150°, 180°, -150°, -120°, -90°, -60°, -30°}. The left (L) and right (R) HRIRs collected for a source location at azimuth θ and elevation ϕ are symbolized by $h_{L,\theta,\phi}$ and $h_{R,\theta,\phi}$, respectively. The corresponding HRTFs are $H_{L,\theta,\phi}$ and $H_{R,\theta,\phi}$. The creation of a spatialized binaural sound (left and right channels) involves convolving the single-channel digital sound to be spatialized, s(n), with the HRIR pair corresponding to the azimuth and elevation of the intended virtual source location:

$$y_{L,\theta,\phi}(n) = \sum_{k=-\infty}^{\infty} h_{L,\theta,\phi}(k) \cdot s(n-k) \quad , \text{ and } \quad y_{R,\theta,\phi}(n) = \sum_{k=-\infty}^{\infty} h_{R,\theta,\phi}(k) \cdot s(n-k) \qquad (1)$$

3. STRUCTURAL MODEL

Structural models of HRTF are based on the premise that each anthropometric feature of the listener affects the HRTF in a way that can be described mathematically [11]. Because such a model has its origin in the physical characteristics of the entities involved in the phenomenon, it should be possible to derive the value of its parameters (for a given source location), from the sizes of those entities, i.e., the anthropometric features of the intended listener. Proper identification of such parameters and adequate association of their numerical values with the anthropometric features of the intended listener may provide a mechanism to interactively adjust a generic base model to the specific characteristics of an individual. One of the most practical models has been proposed by Brown and Duda [6]. Their model is illustrated in Figure 1:

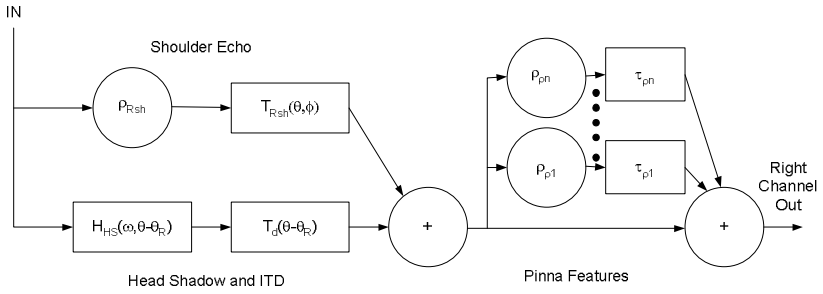

FIGURE1: Right Channel half for Brown & Duda's Structural HRTF Model. The model comprises a symmetric left half (not shown).

4. CUSTOMIZABLE PINNA MODEL

The Head-Shadow sub-model of Duda's structural model was developed by Lord Rayleigh and can be emulated by a one-pole/ one-zero model [5]. This model, incorporating the IID and ITD effects, can account for an approximate localization, particularly in the horizontal plane (elevation, $\phi = 0^o$). However, to produce elevation effects, a good pinna sub-model is required. The definition and anthropometric characterization of the pinna model has remained an open question, so far, and it is the objective of our work. Carlile [7] divides pinna models according to the main phenomenon that they address: Resonating, diffractive and reflective. From these, reflective models have attracted the most attention in the literature.

The intent of our work is to define a functional pinna sub-model that has anthropometric plausibility and then associate its parameters to anthropometric features of the listener's pinna. Taking into account the information available about the existence of a resonant effect implemented by the ear's concha [13] and according to the reflective pinna models discussed previously, we propose that the pinna may, in turn, be modeled as the series connection of an equivalent second-order resonator and a series of characteristic echoes, representing the delayed and attenuated secondary paths taken by the incoming sound, in addition to a "direct path". A block diagram representation of this model is shown in Figure 2.

It should be noted that this pinna model allows for the "direct-path", F0, and each one of the "echoes", F1, F2, and F3, to be affected by a different equivalent resonance.

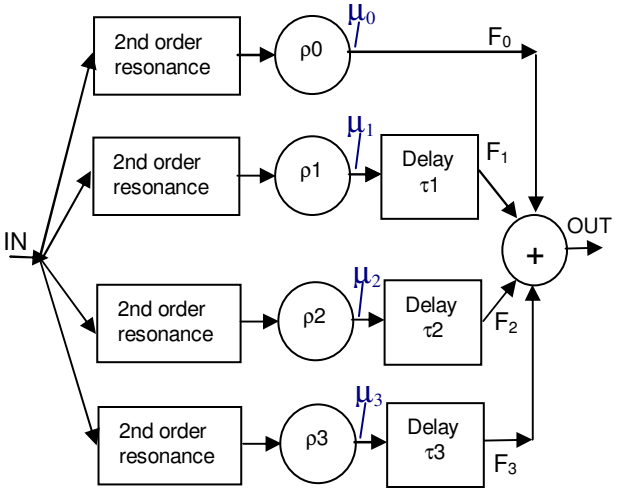

FIGURE 2: Proposed Pinna Model

Accordingly, HRIRs are envisioned as the impulse response of this model, which will be the superposition of four damped sinusoidals (the impulse responses of each of the 2nd order resonators), characterized by their frequency, f, and damping factor, σ. These damped sinusoidals are altered in their amplitude according to the ρ_k parameters, and delayed according to the τk parameters.

Thus, the instantiation of this proposed model will require the identification of the fk, σ_κ, ρ_κ and τ_κ values, to characterize the parameters of the model that successfully approximates an HRIR collected for a given azimuth and elevation, through the output provided by the pinna model. The main challenge in this operation is the fact that the several replicas of the damped oscillation are irreversibly mixed together, partially overlapping in time, in the measured HRIR. This problem was addressed by the sequential application of Prony's modeling algorithm [10, 12] to partial segments of the response. Prony's method approximates a given signal μ(t) as the superposition of p damped sinusoidals:

$$\mu(t) = \sum_{j=1}^{p} \rho_j e^{(\sigma_j t)} \sin(2\pi f_j t + \xi_j) \qquad (2)$$

Figure 3 shows the four F components, obtained using this method, from a measured HRIR

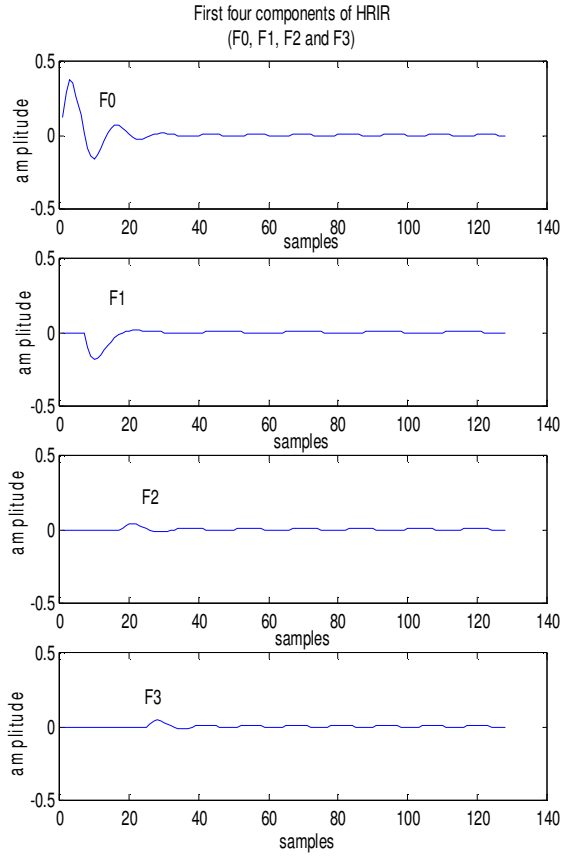

FIGURE 3: Example of HRIR Deconstruction

5. MEASUREMENT OF ANTHROPOMETRIC FEATURES

Key anthropometric features of the ears of the 15 experimental subjects in the study (same 15 subjects for whom the HRIRs were empirically measured with the Ausim 3D system) were captured by means of digital photography (including a distance reference), and laser 3-D scanning, using a Polhemus FastScan handheld scanner.

The features measured are: Ear length (EL), Ear width (EW), Concha width (CW), Concha height (CH), Helix length (HL), Concha area (CA), Concha volume (CV) and Concha depth (CD).

6. ASSOCIATION BETWEEN MODEL PARAMETERS AND ANTHROPOMETRIC MEASUREMENTS

Following the procedure described in the two preceding sections two independent sets of data were available for each pinna of each on of the 15 subjects in the study:

Estimated Model Parameters (for frontal plane sites):

$r_{0\phi}$, $\alpha_{0\,\phi}$, $\rho_{1\phi}$, $\rho_{2\,\phi}$, $\tau_{2\,\phi}$, $\rho_{3\,\phi}$, $\tau_{3\,\phi}$, $\rho_{4\,\phi}$ and $\tau_{4\,\phi}$, for $\phi = -36^o, -18^o, 0^o, 18^o, 36^o$, and 54^o

(Note, here r_0 and α_0 are the magnitude and angle of the poles of the resonator, which define the resonator response F0(n), in terms of its frequency f_0 and its damping factor σ_0)

Measured Anthropometric Features:
EL, EW, CH, CW, CA, CV, CD and HL

Under the assumption that the model parameters depend of the anthropometric features, a general dependency equation may be set, for each model parameter. For example, for the amplitude of the first replica in the pinna model, ρ_1, at $\phi = 54^o$, the following equation may be set up:

$$\rho_{0\phi=54} = KEL(EL) + KEW(EW) + KCH(CH) + KCW(CW) + KCA(CA) + KCV(CV) + KCD(CD) + KHL(HL) + B \quad (3)$$

Coalescing the data from both ears, at the same elevation, (under the assumption of symmetry), 30 equations like the one above can be set up, for each model parameter, at each elevation. Each group of 30 equations can then be analyzed through multiple regression to estimate the values of the constants (KEL, KEW,…KHL, B). The multiple regression analysis was carried out using the Statistical Package for the Social Sciences (SPSS).

7. MODEL EVALUATION AND RESULTS

Using the predictive equations found above, for each subject tested, a Model HRTF was created. Ultimately, the efficiency of the modeled sequences obtained by predicting the model parameters from the anthropometric measurements of the subjects was gauged in listening tests. In these tests, white noise bursts were spatialized using the modeled HRIR sequences, that had been obtained based on the ear measurements of the subject under test, for the six elevations under study. The order in which these elevations where used for the spatialization was randomized. Each elevation was simulated four times (i.e., there were 24 trials for each side of the head.) In each trial the subject would listen to each spatialized sound and then use a graphic user interface to indicate the perceived elevation. Since the spatialization was performed to emulate six specific locations, the absolute value of the angular difference between the perceived elevation and the emulated one would be considered as the elevation error for the trial. The subjects listened to the original, modeled and generic HRTFs. Figure 4 illustrates the average angular error (across all 10 subjects) experienced in the perception of the different emulated elevations for Original, Generic (B&K) and Model HRIRs. The global average error (across all subjects and all elevations) with the original HRTFs, was 23.7^o. The corresponding global average error with modeled HRIRs was 29.9^o. Finally, the global average error when the subjects used the generic HRIRs, collected from the B&K manikin, was 31.4^o.

It should be noted, however, that near the horizontal plane (e.g., between ϕ = -18^o and ϕ = 36^o), the performance of the modeled HRIRs was close to or better than, that of the individually measured HRIRs.

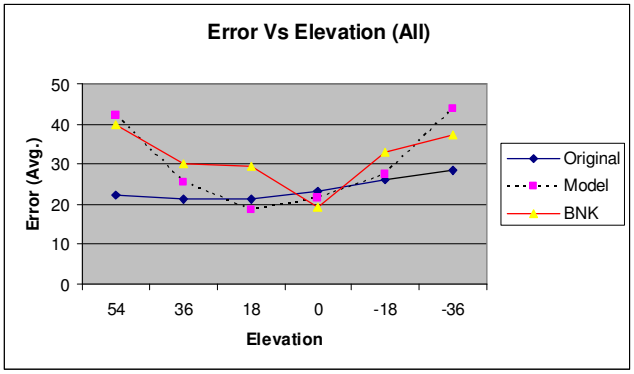

FIGURE 4: Localization performance using 3 different types of HRIRs.

8. CONCLUSIONS

This paper has presented a proposed functional model of the pinna, to be used as the output block in a structural HRTF model. Although this study resorted to the use of a relatively expensive 3-D laser scanner and specialized software to determine some of the anthropometric features of our subjects, which is a prerequisite to the use of the predictive equations developed in this research, it is likely that empirical relationships can be found to obtain these feature values from two-dimensional high-resolution photographs (commonly available) and a few direct physical measurements in the subject.

9. ACKNOWLEDGEMENT

This research was supported by NSF grants, HRD-0317692, IIS-0308155, and CNS-0426125.

10. REFERENCES

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

[2] Begault, D. R., "*A head-up auditory display for TCAS advisories.*" Human Factors, 35, 707-717, 1993.

[3] Begault, D. R., Wenzel, E. M. and Anderson, M. R., "*Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source*", J. Audio Eng. Soc., Vol. 49, No. 10, pp. 904-916, 2001.

[4] Begault, D., "3-D Sound for Virtual Reality and Multimedia", Academic Press, 1994.

[5] Brown, C. P. and Duda, R. O., "*An Efficient HRTF Model for 3-D Sound,*" Proc. 1997 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk, NY, Oct. 1997.

[6] Brown, C. P. and Duda, R. O., "*A Structural Model for Binaural Sound Synthesis*", IEEE Trans. Speech and Audio Processing, Vol. 6, No. 5, pp.476-488, Sep. 1998.

[7] Carlile, S., "*The Physical Basis and Psychophysical Basis of Sound Localization*", in Virtual Auditory Space: Generation and Applications, S. Carlile, Ed., pp. 27-28, R. G. Landes, Austin TX, 1996.

[8] Gardner, W. G. and Martin, K. D., "*HRTF measurements of a KEMAR*". J. Acoust. Soc. Am., 97(6), 1995.

[9] Mills, A. W., "Auditory Localization," in Foundations of Modern Auditory Theory, Vol. II (J, V. Tobias, Ed.), pp. 303-348, Academic Press, New York, 1972.

[10] Osborne, M. R., and Smyth, G. K., "*A modified Prony algorithm for fitting sums of exponential functions*", SIAM J. Sci. Statist. Comput., vol. 16, pp. 119-138, 1995.

[11]Han, H.L., "Measuring a Dummy Head in Search of Pinna Cues", J. Audio Eng., Soc., Vol. 42, No. 1 / 2 , pp. 15 – 37, 1994.

[12] Parks and C.S. Burrus, "*Digital Filter Design*", John Wiley and Sons, p226, 1987.

[13] Shaw, E. A. G., Teranishi, R., "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, 1967.