Title
Structural Composition and Decomposition of HRTF's

Authors
Algazi, Ralph
Duda, Richard O
Morrison, Reed P.
et al.

Publication Date
2001

STRUCTURAL COMPOSITION AND DECOMPOSITION OF HRTFS

V. Ralph Algazi, Richard O. Duda, Reed P. Morrison and Dennis M. Thompson

CIPIC
U.C. Davis
Davis, CA 95616, USA
`vralgazi@ucdavis.edu`

ABSTRACT

The analysis and modeling of the response of parts of the body provides valuable insight into many features of the head-related transfer function (HRTF). In spatial sound simulations, partial models, such as the spherical head model, can also generate simple and effective approximate localization cues. In this paper, we consider the composition of an approximate HRTF from the responses of structural components by making use of detailed measurements of isolated pinnae and of a pinna-less head and torso. We determine that such a composition is sensitive to additional geometric parameters that can be obtained from anthropometry. We show that, with such parameters, simple composition rules can produce a good correspondence between measured and composite HRTFs.

1. INTRODUCTION

Head-related transfer functions (HRTFs) depend explicitly on frequency and the spatial location of the source relative to the listener, and implicitly on the size and shape of the listener's body. Much of the complexity of HRTFs stems from the combination of the separate effects of the torso, head, and pinnae, each of which scatters incident sound waves in different but overlapping frequency bands.

It has recently been shown that the complete HRTF can be computed by solving the wave equation subject to the boundary conditions presented by the surface of the body [1, 2]. Although this approach is very promising and may ultimately replace acoustic techniques for measuring HRTFs, it is still important to understand how the various components of the body influence HRTF behavior. The use of models provides an important way to gain this understanding.

Lord Rayleigh's classical model of the head as a rigid sphere explains much of the behavior of the interaural time difference (ITD), as well as some of the behavior of the interaural level difference (ILD) [3]. The Lopez-Poveda/Meddis model of the concha as a spiral-shaped cylinder explains the behavior of the so-called "pinna notch" [4]. Shaw's more complex model of the pinna provides insight into the various pinna resonances and general high-frequency HRTF behavior [5]. Our own work with ellipsoidal models for the torso explains the effect of torso reflections on low-frequency HRTF behavior [6].

Genuit first proposed the representation of HRTFs as a combination of filter structures, each component accounting for the contributions of a corresponding anatomical structure [7]. We shall refer to such a representation as a structural model. As a simple example, one might hypothesize that a spherical head model can be used to estimate the sound field in the vicinity of the pinna, and that a pinna model can then be used to estimate the sound field that enters the ear canal. Thus, the HRTF might be represented by a cascade of two filters, one for the head and the other for the pinna. Indeed, we take exactly such an approach in this paper, except that the first filter will account for the combined effects of the head and torso.

Because each filter block corresponds to a specific anatomical structure, filter parameters can be related to specific anthropometric measurements, such as the head radius or the concha diameter. This allows the use of anthropometry to adapt a generic HRTF model to a particular listener, or to account for the effects of posture on the HRTF. However, any decomposition of the HRTF into a combination of filters, where each filter represents the effects of one anatomical structure in isolation, is a heuristic approximation. There are interactions between the waves scattered from the different structures, and it is not clear that it is really possible to recover the full HRTF from HRTFs for different parts of the body measured in isolation.

Our purpose in this paper is to illustrate the composition of an HRTF from transfer functions for structural parts, and to show that structural modeling can, in fact, provide a good approximation to real HRTFs. During the past three years we have assembled a substantial database of measured HRTFs with corresponding anthropometry [8]. More recently, we have measured the spatial responses of a number of isolated pinnae. We first report on the measurement of the responses of the isolated pinnae and of the head-torso combination, and identify key features that the pinna, head and torso contribute to the total HRTF. We then determine and discuss the geometry of the relative positions of these components determined by anthropometry. Based on this geometry, we propose an approximate HRTF as a simple cascade model of the structural components. Although psychoacoustic evaluation remains to be done, we show that such a composition rule can produce a good approximation to the measured HRTF.

2. STRUCTURAL DECOMPOSITION

Investigation of the separate effects of the pinna and the head/torso combination is easily done for the KEMAR mannequin, which has removable pinnae. Measurement of the HRTF without pinna identifies the contributions of the head/torso combination. The re-

---

This work was supported by the National Science Foundation under grants IRI-96-19339 and ITR-00-86075. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

sponse of the isolated KEMAR pinna can also be measured by a method that we describe below. For live subjects, the response of isolated pinnae can be measured using plaster castings made from molds of the pinnae. It is also feasible to measure an approximate pinna-less HRTF for live subjects by having them wear a shower cap that flattens the pinnae, with microphones mounted at the position of the ear canals. We now describe the methods and results obtained in such a decomposition.

2.1. Pinna-related transfer function (PRTF)

1. Measurements

   We have developed an apparatus and a method for measuring the response of the isolated pinna on a plane. The apparatus provides a good approximation to the ideal case of a pinna mounted on an infinite plane. The pinna is placed on a circular plate surrounded by a rectangular table (see Fig. 1). The circular plate mounted on the turntable allows for the continuous rotation of the pinna. The ear canal of the pinna, positioned at the center of rotation on the plate, is at the center of a hoop with a one-meter radius. The plane defined by the hoop is perpendicular to the plane of the plate. The impulse response of the pinna at a specific azimuth and elevation is measured by playing bursts of sound (Golay codes) through the speakers while recording with the microphone placed in the ear canal of the pinna [9]. The resulting response is compensated by making similar measurements with the pinna removed from the plane. Thus, the PRTF measures the change in response produced by the scattering introduced by the pinna.

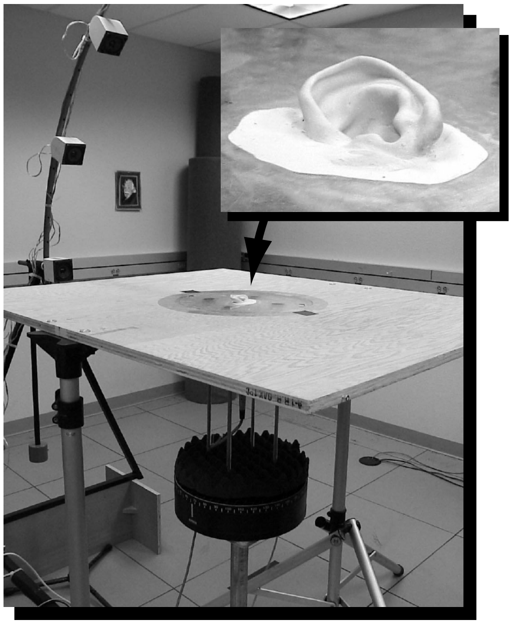


Figure 1: *Measurement of an isolated pinna on a plane*

2. Coordinate Systems

   In the measurement of HRTF data we have adopted a head-centered interaural-polar coordinate system. A sound source location in such a system is characterized by the azimuth angle θ, elevation angle ϕ, and range r. The azimuth angle is the angle between a vector to the sound source and the vertical median plane, where $\theta = +90^\circ$ corresponds to a source pointing directly into the right ear and 0° corresponds to a source pointing directly at the front of the head.[1]

   The coordinate system of the pinna-related transfer function, or PRTF, is ear-canal-centered rather than head-centered. Such a system (see Fig. 2) is exactly the same as the head-centered system except for a small azimuth and range offset that is negligible if the size of the head is small compared to the range to the sound source.

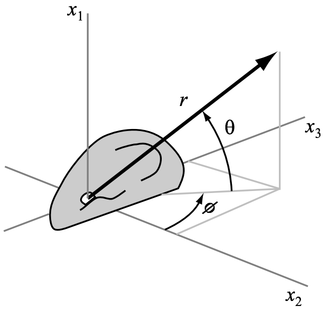


Figure 2: *The interaural-polar coordinate system*

   Because the source is never located below the pinna on the plate, the PRTF data set corresponds to ipsilateral HRTF data. Therefore for a right pinna the azimuth ranges from 0° to $+90^\circ$, and for a left pinna the azimuth ranges from 0° to -90°.

3. Analysis of the pinna response

   The absence of the head and torso contributions enhances and simplifies the appearance of the salient features due to the pinna. Fig. 3 shows the frequency response for the KEMAR large pinna for a 25° cone of constant azimuth (“cone of confusion”). The response was measured every 5°, smoothed with a critical-band filter, and displayed as an image, with dB magnitude mapped to grayscale. The resulting response is substantially simpler than that for the complete HRTF. We can see the pinna notches at low elevations, both front and back. Fig. 3b shows the frequency response for 0° elevation.[2] The PRTF varies slowly as a function of azimuth until -150°, where a pronounced attenuation or shadow due to the pinna becomes apparent.

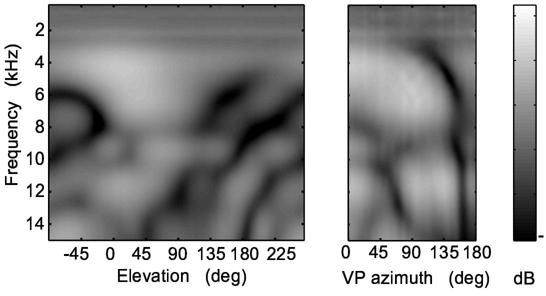
2.2. Response of the head and torso (HAT)

The HAT response H_H of the pinna-less KEMAR mannequin was also measured densely in both azimuth and elevation. Fig. 4 shows the results for a 25° cone of confusion. In the frequency domain, the dominant features are arch-shaped notches (Fig. 4a) that correspond to torso reflections in the time domain (Fig. 4b) [6]. Note that the amplitude of these notches fades out at high frequencies when the sound source is at higher elevations. This corresponds to

---

[1] Note that these angles are different from the angles in a vertical-polar coordinate system, although we use the same terminology for them.

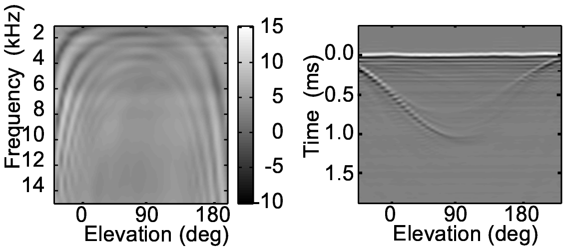
[2] In Figure 3b, we use a vertical-polar azimuth rather that the interaural-polar azimuth to show the response in back as well as in front.



(a) 25° cone of confusion (b) Horizontal plane

Figure 3: *HRTF spectrum for the isolated pinna*

a broadening of the torso reflection that results from dispersion as the incident wave is scattered from upper torso surfaces. Although the exact behavior of these features is subject dependent, qualitatively similar characteristics are exhibited in all of the data that we have examined [6].



(a) HRTF (b) HRIR

Figure 4: *Pinna-less head and torso response*

3. HAT AND PINNA COMPOSITE GEOMETRY

The orientation of the pinna with respect to the head varies significantly from person to person, and has a strong effect on the HRTF[10]. In the HAT coordinate system, the pinna orientation can be specified by three Euler angles: a outward flare θ_f and downward dip θ_d that bring the reference plane for the pinna tangent to the head at the entrance to the ear canal, followed by a counterclockwise rotation θ_r about the normal vector (see Fig. 5). Given these angles, and given the azimuth and elevation for the HAT data, one can determine the corresponding azimuth and elevation for the isolated pinna data.

4. COMPOSITE HRTFS

In the reflection and scattering of sound waves that result in the full HRTF, the interaction of the components is complex, since the torso reflections and scattering will impinge on the pinnae from low-elevation directions and combine or interfere with the direct sound to the pinna. However, examination of the HAT response, of the pinna response and of the full HRTF suggests a simple model and composition rule.

The impulse response HAT data in Fig 4b leads us to hypothesize that sound waves reach the ear by two distinct paths, and suggests the model shown in the upper part of Fig 6. We write the transfer function for the pinna-less head and torso as the sum $H_H = H_H^d + H_H^i$, where H_H^d accounts for the direct path and H_H^i accounts for the indirect or reflected path. Because the waves for these two paths arrive at the ear from different directions, they are filtered differently by the pinna. Thus, we write the complete HRTF as

$$H = H_H^d H_P^d + H_H^i H_P^i = H_H^m H_P^d \qquad (1)$$

where H_P^d is the PRTF for the direct path, H_P^i is the PRTF for the indirect path, and $H_H^m = H_H^d + H_H^i (H_P^i / H_P^d)$. Now, if the PRTF were independent of direction, H_H^m would simplify to H_H and H would simplify to the product $H = H_H H_P^d$. We shall show empirically that this simple composition rule turns out to provide a surprisingly good approximation. A possible explanation is based on a rough case analysis. The greatest differences in the directions of arrival occur when the source is at moderate or high elevations. In this case, the energy from the indirect path is scattered by the irregular shape of the upper torso, and the resulting dispersion of the impulse response reduces its high-frequency content. (Inspection of Fig 4a confirms that the torso notches at high elevations are very weak above 4 kHz, where the pinna begins to be significant.) Thus, at high elevations the magnitude of H_H^i is so small that differences between H_P^i and H_P^d have little effect. At low elevations, the bandwidth of H_H^i is wider, but the directions of arrival are sufficiently close that $H_P^i \approx H_P^d$ and once again $H_H^m \approx H_H$. Thus, in both cases we can approximate H by $H_H H_P^d$.

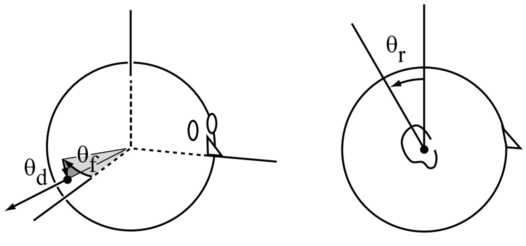


Figure 5: *Pinna location and orientation*

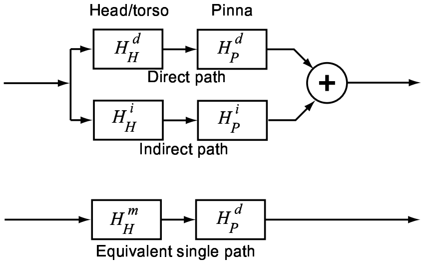


Figure 6: *Models for HRTF composition*

4.1. Composite HRTF for KEMAR

We illustrate the composition process and the resulting composite KEMAR HRTF in Fig. 7. The HAT response for the 25° cone of confusion is shown in Fig. 7a. As we have discussed, the PRTF has to be remapped to align the geometry of the pinna with that of the head and torso. We measured the following values for the large

KEMAR pinna: $\theta_f = 4^\circ$, $\theta_d = 11^\circ$ and $\theta_r = 5^\circ$. Fig. 3a shows the PRTF at an azimuth of 25°. Transformation of that data to the HRTF coordinates result in the pinna response shown in Fig. 7b. In addition to shifting the response, the remapping produces some changes in the depth of the notches for locations in back (elevations above 90°). The composition of pinna and HAT response is illustrated in Fig. 7c and can be compared to the measured HRTF of Fig. 7d. We have assumed that $H_P^i = H_P^d$, so that the composite HRTF is computed simply as the sum in dB of the HAT response and the geometrically corrected PRTF response. Visually, the two HRTFs are in good correspondence. The complex behavior of the composite HRTF can be seen to be the consequence of the considerably simpler behaviors of the components.

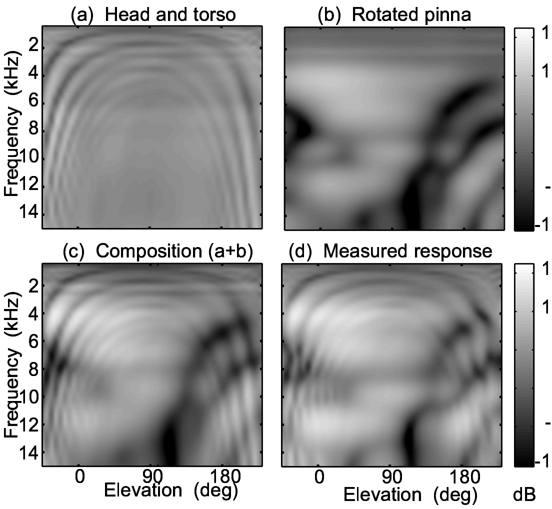


Figure 7: *Composition of HRTFs for a* 25° *cone of confusion*

5. DISCUSSION AND CONCLUSIONS

Structural analysis of the HRTF is a promising approach to understanding and modeling the distinct contributions of the head, torso and pinna. The composition of such partial contributions requires other anthropometric parameters that specify the relative position and orientation of the components. Once these additional parameters have been specified, simple composition rules may provide an effective substitute for a measured HRTF.

Additional work is needed to fully investigate and evaluate this approach. Preliminary investigations at other azimuth angles produced similar results for azimuth angles above 15°. For azimuths between 0° and 15°, the pinna notches are not well reproduced when the source is in back. This is perhaps to be expected, because the difference between mounting the pinna on a plane and on a head becomes significant near the median plane. Obviously, the approach breaks down completely on the contralateral side, where we cannot measure the contribution of the isolated pinna at all. However, it is possible that the contralateral HRTF can be adequately approximated from the ipsilateral HRTF by using a spherical-head model [11].

The ultimate measure of the effectiveness of any approximate HRTF is how faithfully it recreates spatialized sound. A formal psychoacoustic evaluation of such a complete composite HRTF is a future task. However, these empirical results are encouraging, and bring us closer to our goal of substituting simple geometric or mathematical models for complex measurements.

6. REFERENCES

[1] Katz, B. F. G., “Measurement and Calculation of Individual Head-Related Transfer Functions Using a Boundary Element Model Including the Measurement and Effect of Skin and Hair Impedance,” PhD dissertation, Graduate Program in Acoustics, Pennsylvania State University, 1998.

[2] Kahana, Y., Nelson, P. A., Petyt, M. and Choi, S., “Numerical Modeling of the Transfer Functions of a Dummy-head and of the External Ear, Proc. AES 16th Int. Conf. Spatial Sound Reproduction, Rovaniemi, Finland, 1999.

[3] Blauert, J., Spatial Hearing, Revised Edition, MIT Press, Cambridge, MA, 1997.

[4] López-Poveda, E. A., and Meddis, R., “A Physical Model of Sound Diffraction and Reflections in the Human Concha,” J. Acoust. Soc. Amer., Vol. 100, 3248-3259, 1996.

[5] Shaw, E. A. G., “Acoustical Features of the Human Ear” in Binaural and Spatial Hearing, R. H. Gilkey and T. A. Anderson eds. Lawrence Erlbaum Associates, 1997.

[6] Algazi, V. R., Avendano, C., and Duda, R. O., “Elevation Localization and Head-related Transfer Function Analysis at Low Frequencies,” J. Acoust. Soc. Am., Vol. 109, 1110-1122, 2001.

[7] Genuit, K., “Ein Modell zur Beschreibung von Außenohrübertragungseigenschaften”, doctoral dissertation, Rheinisch-Westfälischen Technichen Hoch-schule Aachen, Germany, 1984.

[8] Algazi, V. R., Duda, R. O., and Thompson, D., “The CIPIC HRTF Database”, WASPAA ’01 (2001 IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics), Mohonk Mountain House, New Paltz, NY, 2001.

[9] Algazi, V. R., Avendano, C., and Thompson, D., “Dependence of Subject and Measurement Position in Binaural Signal Acquisition,” J. Audio Eng. Soc., Vol. 47, 937-947, 1999.

[10] Carlile, S., Jin, C. and Harvey, V., “The Generation and Validation of High Fidelity Virtual Auditory Space,” Proc. 20th Ann. Int. Conf. IEEE Engineering in Medicine and Biology, Vol. 20, No. 3, pp. 1090-1095, 1998.

[11] Avendano, C., Duda, R. O., and Algazi, V. R., “Modeling the Contralateral HRTF,” Proc. AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, 313-318, 1999.