EXTRACTION OF PINNA SPECTRAL NOTCHES IN THE MEDIAN PLANE OF A VIRTUAL SPHERICAL MICROPHONE ARRAY

Ankit Sohni, Chaitanya Ahuja, and Rajesh M. Hegde

Indian Institute of Technology, Kanpur
Department of Electrical Engineering
Email: {sohani, chahuja, rhegde}@iitk.ac.in

ABSTRACT

In this paper, a fast method for the extraction of pinna spectral notches (PSN) in the median plane of a virtual spherical microphone array is discussed. In general, PSN can be extracted from the Head Related Impulse Response (HRIR) measured by a spherical array of microphones. However, the PSN extracted herein are computationally complex and also not accurate at lower elevation angles. This work proposes a novel approach to reconstruct the HRIR using microphones over the median plane of a virtual spherical array. The virtual spherical array itself is simulated using the Fourier Bessel series (FBS). Subsequently, these HRIRs are used to extract the PSN. This method is computationally efficient since it is done over the median plane rather than over the complete sphere. On the other hand, it is also accurate due to the utilization of the Fourier Bessel series in the extraction of the PSN. Experimental results obtained on the CIPIC database indicate a high degree of resemblance to the actual pinna walls, even at the lower elevation angles. The results are motivating enough for the method to be considered for resolving elevation ambiguity in 3D audio.

Index Terms— Head Related Transfer Function, Head Related Impulse Response, Fourier Bessel Series, Pinna Spectral notches

1. INTRODUCTION

Individualized Head Related Transfer Function (HRTF) is required to accurately render spatial audio for each person in virtual auditory space (VAS). Rendering sound using non-individualized HRTF results in localization errors such as Front-Back Ambiguity (FBA), Inside-the-Head localization (IHL), and elevation angle misperception [1],[2]. HRIR captures the effects of interaction of sound with the anatomy of the human body which includes head diffraction, torso reflections, and pinna reflections [3]. Due to differences in the human anatomy from person to person, HRIR is different for each individual. But measuring HRIR is a tedious and time consuming process. Hence it is not practical to create a data set of HRIR through standard method described in [4]. The human anatomy is responsible for sound localization. An example is the case of the head being responsible for the lateralization of the sound source in the horizontal plane [5]. Head diffraction causes Interaural Time Difference (ITD) and Interaural Level Difference (ILD) between the sound waves arriving at both ears. It has been shown that ITD and ILD are primary binaural cues for the horizontal plane localization [5]. In the median plane (the plane from where the entrance of ear canal of both the ears are co-located), both the cues are nearly equal to zero. Hence localization in the median plane relies on monaural spectral cues [6],[7]. According to Batteau [8], pinna shape causes multiple reflections of sound waves. The delay between direct wave and reflected wave results in spectral notches. These spectral notches smoothly vary with the elevation angle and are highly dependent on pinna dimensions. Hence localization in the median plane is strongly dependent on the pinna geometry [9]. For different subjects, the location of the spectral notches is different for the same elevation angle, which corresponds to the inter-personal variations in the spectral notches.

The organization of the paper is as follows. Section 2 describes the median plane HRTF representation using FBS. It also describes the reconstruction of continuous median plane HRTF from modal parameters. Section 3 describes the spectral notch extraction algorithm from HRTF reconstructed using FBS. Section 4 illustrates the extracted pinna spectral notches plotted on the pinna image. Smooth pinna spectral notch variations with elevation angles can also be observed herein. This is followed by conclusion in Section 5.

2. REPRESENTATION OF MEDIAN PLANE HRTF THROUGH FOURIER BESSEL SERIES

The mathematical representation of median plane HRTF denoted by $H(f, \phi)$ which is function of frequency f and elevation angle ϕ can be modeled by FBS. In this section, we develop the FBS representation of HRTF in the spherical harmonics domain.

2.1. Spatial Modeling of HRTF

The 3D sound field recorded by spherical array of microphones due to the source located at the entrance of the subject's ear canal is not only the function of frequency f but also that of the spatial coordinates (r, θ, ϕ). Therefore, the recorded HRTF captured by spherical microphone array can be decomposed into its corresponding spherical harmonics as

$$H(k; r, \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} H_n^m(k; r) Y_n^m(\theta, \phi) \tag{1}$$

where $H_n^m(k; r)$ is the Spherical Fourier Transform (SFT) which depends on wave number $k = \frac{\omega}{c}$ for frequency $\omega = 2\pi f$ and speed of sound c, and range r. $Y_n^m(\theta, \phi)$ is called the spherical harmonics of order n and degree m. $Y_n^m(\theta, \phi)$ is defined as

$$Y_n^m(\theta, \phi) = \alpha_n^m P_n^{|m|}(\cos\theta) e^{jm\phi}; 0 \leq \theta \leq \pi, 0 \leq \phi < 2\pi$$
$$\text{with } \alpha_n^m = \sqrt{\frac{2n+1}{4\pi} \frac{(n-|m|)!}{(n+|m|)!}} \tag{2}$$

This work was funded by SERB, Department of Science and Technology (DST), GoI.

where $P_n^{|m|}(\cdot)$ is the associated Legendre polynomial. Since HRTF is generally recorded for fixed range r. Under the far field assumption ($r \geq 1m$), HRTF will be independent of range r and only remain the function of spatial angle (θ, ϕ) and frequency f.

$$H(f;\theta,\phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} H_n^m(f) Y_n^m(\theta,\phi) \tag{3}$$

2.1.1. *HRTF modeling over the median plane*

The decision to choose elevation angle between θ and ϕ is crucial in terms of representation of median plane HRTF. The complex exponents preserve much of the variations as compared to associated Legendre polynomials under finite truncation and convergence is rapidly achieved [10]. This is also important in terms of computational complexity. Therefore, ϕ is chosen as the elevation angle. In the median plane ($\theta = \frac{\pi}{2}$), HRTF can be represented from Equations 2 and 3 as

$$H(f, \theta = \frac{\pi}{2}, \phi) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} \alpha_n^m H_n^m(f) P_n^{|m|}(\cos\frac{\pi}{2}) e^{jm\phi} \tag{4}$$

$P_n^{|m|}(\cos\frac{\pi}{2})$ is non-zero for all odd values of m when n is odd. It is also non-zero for all even values of m when n is even. Therefore, Equation 4 can be re-arranged as

$$H(f,\phi) = \sum_{m=-\infty}^{\infty} [\sum_{n=0}^{\infty} \alpha_{2n+|m|}^m H_{2n+|m|}^m(f) P_{2n+|m|}^m(\cos\frac{\pi}{2})] e^{jm\phi}$$

$$H(f,\phi) = \sum_{m=-\infty}^{\infty} C_m(f) e^{jm\phi} \tag{5}$$

It can be inferred from Equation 5 that spatial component of the median plane HRTF can be represented by Fourier series. Therefore, $C_m(f)$ can also be called the Fourier coefficient.

2.2. Spectral Modeling of Fourier Coefficient

The orthogonal sets of Bessel family of functions are used to model the Fourier coefficients due to strong correlation between spectral component of HRTF ($C_m(f)$) and the family of Bessel functions of first kind as proposed in [11].

$$C_m(f) = \sum_{k=1}^{\infty} C_{mk} J_{|m|}(\beta_k^{|m|} \frac{f}{f_{max}}) \tag{6}$$

where $J_n(\cdot)$ is the Bessel function (first kind) of order n. $\beta_1^n, \beta_2^n \cdots$, β_k^n are the positive roots of $J_n(x) = 0$.

2.3. Decomposition of median plane HRTF into Fourier Bessel Series

By combining Equations 5 and 6, the median plane HRTF can be decomposed into Fourier Bessel series as

$$H(f,\phi) = \sum_{m=-\infty}^{\infty}\sum_{k=1}^{\infty} C_{mk} J_{|m|}(\beta_k^{|m|} \frac{f}{f_{max}}) e^{jm\phi} \tag{7}$$

where C_{mk} is the complex modal parameter, and represents the Fourier Bessel coefficient. By using the following orthogonality condition of the Bessel family

$$\int_{x=0}^{1} J_{|l|}(\beta_k^{|l|} x) J_{|l|}(\beta_{k'}^{|l|} x) dx = \frac{[J_{|l+1|}(\beta_k^{|l|})]^2}{2} \delta_{kk'} \tag{8}$$

The modal parameter C_{mk} can be calculated as

$$C_{mk} = \frac{1}{\pi [J_{|m+1|}(\beta_k^{|m|})]^2} \int_{0}^{f_{max}} \int_{-\pi}^{\pi} f H(f,\phi) J_{|m|}(\beta_k^{|m|} \frac{f}{f_{max}}) e^{-jm\phi} df d\phi \tag{9}$$

It is important to note that the HRTF is generally measured over discretized spherical grid. In particular, median plane HRTF is uniformly sampled on hemisphere from $-45°$ to $225°$ in the CIPIC database. Therefore, modal parameter C_{mk} must be calculated from discrete spatial and spectral HRTF. Additionally, the intregations in Equation 9 take the form of summations in Equation 10 as

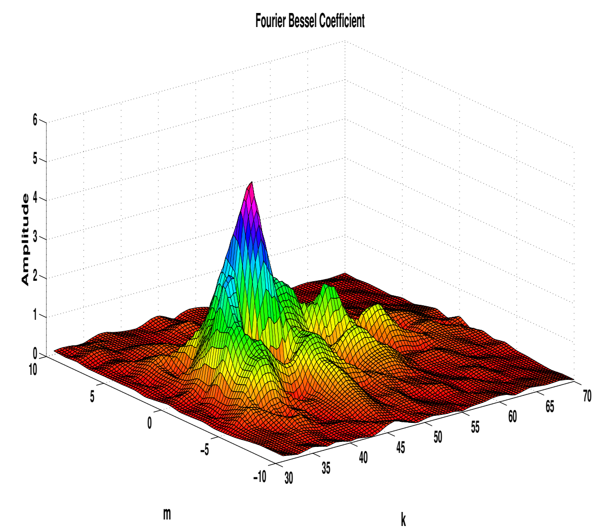

Fig. 1. Amplitude of Fourier Bessel coefficient $|C_{mk}|$ in the median plane for subject 50 left pinna in CIPIC Database.

$$C_{mk} = \frac{1}{\pi [J_{|m+1|}(\beta_k^{|m|})]^2} \sum_{f_i=0}^{f_{max}} \sum_{\phi_i=-\frac{\pi}{4}}^{\frac{5\pi}{4}} f_i H(f_i,\phi_i) J_{|m|}(\beta_k^{|m|} \cdots \cdots \frac{f_i}{f_{max}}) e^{-jm\phi_i} \tag{10}$$

2.4. Choice of Truncation Number

In practice, it is not possible to use infinite number of basis functions in Equation 7. Therefore, median plane HRTF should be reconstructed with the finite number of basis functions under minimum truncation error. It is found that the amplitude of modal parameter $|C_{mk}| \neq 0$ for $|m| \leq M$ and $K' < k < K' + K$. The modal parameters C_{mk} corresponding to first K' roots of the Bessel series are not significant. They preserve the faint initial pulses (artifacts due to probe microphone recording) arriving before the direct wave in HRIR. The roots from K' to $K' + K$ are very significant for spectral notches since they preserve the reflections due to shoulder, torso, and specially pinna cavities in HRIR. In CIPIC Database, it is observed that convergence is rapidly achieved with minimum truncation error when M = 10, K' = 30, and K = 40. Figure 1 demonstrates the dimensionality of Bessel functions as finite truncated representation of Equation 7.

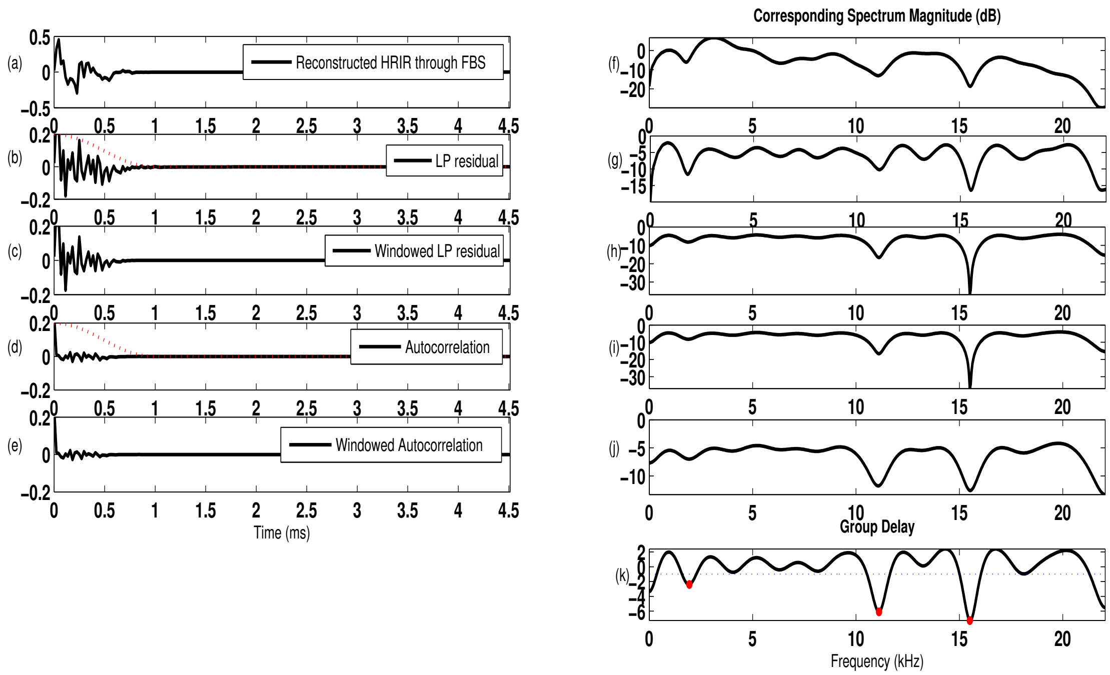

Fig. 2. The extraction of pinna spectral notch frequencies from HRIR reconstructed through FBS by the algorithm proposed in [12] at elevation angle $\phi = -22.5^\circ$ for the right pinna of subject 119.

3. EXTRACTION OF PINNA SPECTRAL NOTCHES

According to two ray reflection model [12], the resultant signal $y(t)$ due to the interference between direct wave $x(t)$ and the wave reflected by pinna wall $x(t - t(\phi))$ is given by

$$y(t) = x(t) + \rho(\phi)x(t - t(\phi)) \tag{11}$$

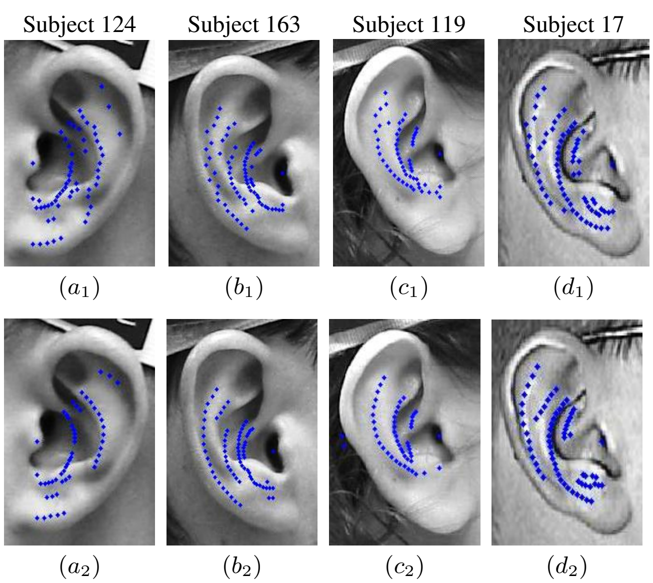

Fig. 3. The dimensions corresponding to pinna spectral notches extracted from measured HRIR (a_1, b_1, c_1, d_1) and from reconstructed HRIR (a_2, b_2, c_2, d_2) are marked on the pinna images available on CIPIC Database for subjects 124, 163, 119, and 17 respectively.

where $\rho(\phi)$ is the reflection coefficient. $t(\phi)$ is the delay between direct wave and reflected wave arriving from elevation angle ϕ. This elevation-dependent temporal delay $t(\phi)$ results in the point of reflection in the pinna image at a distance given by

$$d(\phi) = \frac{ct(\phi)}{2} \tag{12}$$

It also results in the periodic spectral notches whose frequencies are given by

$$f_n(\phi) = \frac{2n+1}{2t(\phi)} = \frac{c(2n+1)}{4d(\phi)}, \forall n = 0, 1, 2, \cdots \tag{13}$$

Note that the reflection coefficient $\rho(\phi)$ to be positive [12]. The first spectral notch frequency occurs at

$$f_0(\phi) = \frac{c}{4d(\phi)} \tag{14}$$

Under the assumption of reflection coefficient $\rho(\phi)$ to be negative as proposed in [13], the spectral notch frequency gets doubled as

$$f_0(\phi) = \frac{c}{2d(\phi)} \tag{15}$$

The spectral notch extraction process used in this work is based on the algorithm that is proposed in [12]. In this algorithm [12], robust signal processing techniques such as windowing, LP-residual analysis, auto-correlation are applied and spectral notches are extracted using the group delay function. Based on the prior research-hypothesis, we assume that pinna spectral notches appear in the frequency range from 5 kHz to 16 kHz. This is also found to be true for almost all subjects in the CIPIC Database. The process of extracting pinna spectral notches, as mentioned above, is illustrated in Figure 2.

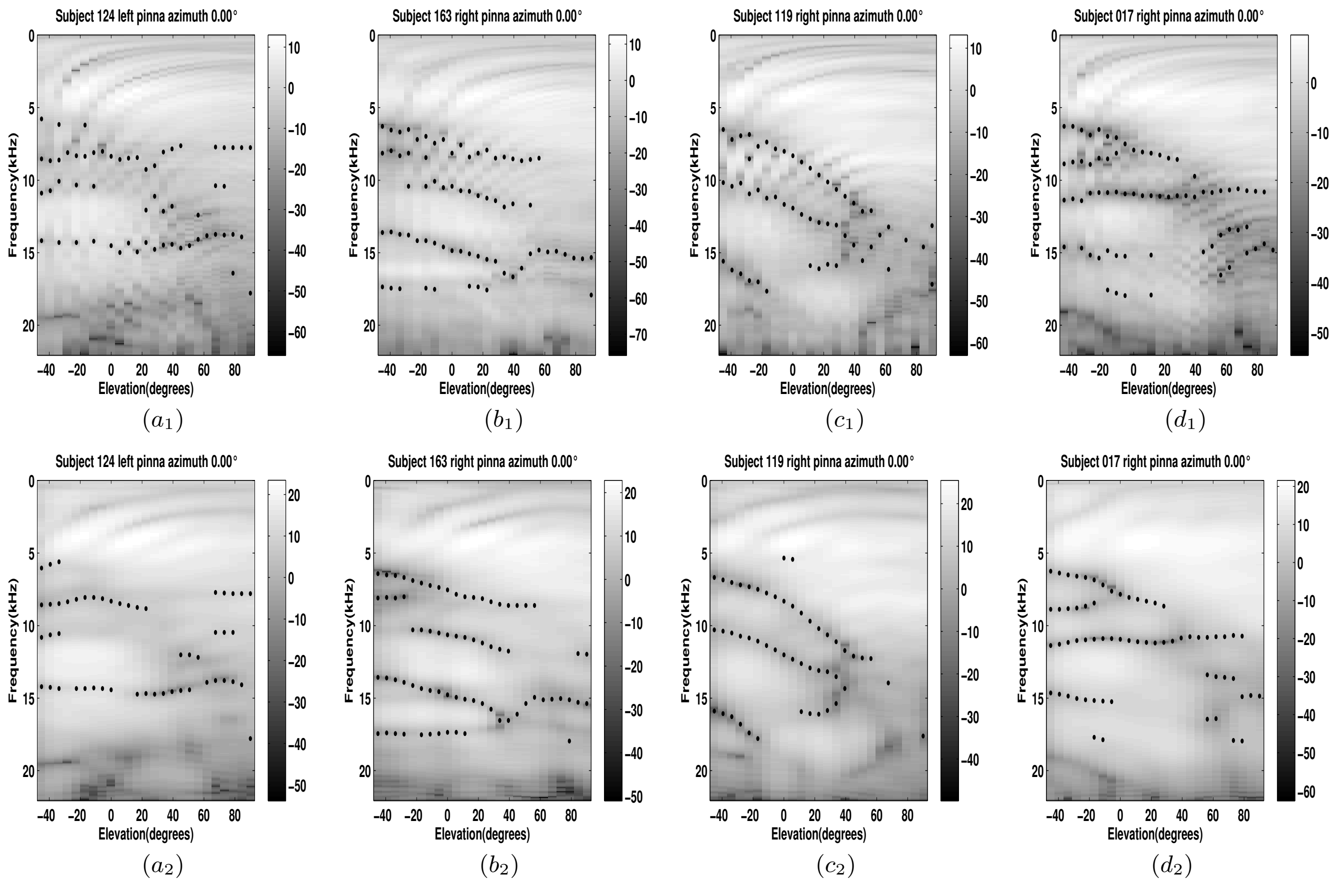

Fig. 4. The pinna spectral notch frequencies extracted from measured HRTF (a_1, b_1, c_1, d_1) and reconstructed HRTF through FBS (a_2, b_2, c_2, d_2) for different elevations from -45° to 90° and azimuth 0° of subject 124, 163, 119, and 17 respectively in CIPIC database.

4. EXPERIMENTS ON PINNA NOTCH EXTRACTION ON CIPIC DATABASE

Publicly available CIPIC database [4] has been used for testing purposes. We have used the data of 45 subjects' whose HRIRs are available with their pinna images and corresponding anthropometry parameters. Following the hypothesis of negative reflection coefficient that is proposed in [13], the relationship between spectral notches and pinna dimensions is established. It is found that the pinna of most of the subjects exhibit high degree of resemblance with this hypothesis. Pinna image of a particular subject is taken from the database and uniformly scaled in order to match with pinna parameters d_5 (pinna height) and d_6 (pinna width) [4]. The distance between pinna reflection point and the entrance of the ear canal is calculated from Equation 15. The dimensions corresponding to notch frequencies extracted from measured and reconstructed HRIR for elevation angles from -45° to 90° are marked on pinna contours of several subjects using 2D polar coordinate system. Each notch point is mapped to $(d(\phi), \pi + \phi)$ in the right pinna and $(d(\phi), -\phi)$ in the left pinna with respect to entrance of the ear canal as origin. It can be clearly observed from Figure 3 that the extracted notch frequencies from reconstructed HRIR smoothly follow the pinna edges (a_2, b_2, c_2, d_2) even for lower elevation angles. These results are compared to those obtained from measured HRIR (a_1, b_1, c_1, d_1). It can also be seen in Figure 4 that the marked spectral notches on the gray image of reconstructed HRTFs (a_2, b_2, c_2, d_2) smoothly vary with the elevation angle as compared to slightly deviated spectral notches of measured HRTFs (a_1, b_1, c_1, d_1). For the smooth and unambiguous extraction of pinna spectral notches, it is necessary to highlight only the reflections due to pinna alone while removing other anatomical reflections. HRIRs corresponding to lower elevation angles suffer from knee reflections which have slight contribution as compared to other anatomical reflections in the measured signal. The proposed method can suppress the knee reflections due to capability of preserving strong variations of pinna alone under finite truncation as seen in Figure 1. It can be noted from Figure 4 that the proposed method is robust to extract the pinna notches even if HRIR is measured over the complete hemisphere.

5. CONCLUSION

A fast method to extract accurate pinna spectral notches that follow the actual pinna wall structure is discussed in this work. The primary contributions of this work are the efficient reconstruction of Head Related Impulse Response over the median plane of a virtual spherical array simulated using the Fourier Bessel series, especially at lower elevation angles. The proposed method is computationally efficient when compared to pinna spectral notches extraction over a full spherical array. The pinna spectral notches extracted are also very accurate and smooth when compared to conventional spherical array based approach. Future scope of this work is to extend this to the complete hemisphere in the median plane.

6. REFERENCES

[1] Durand R Begault, *3-D sound for virtual reality and multimedia*, Citeseer, 2000.

[2] Elizabeth M Wenzel, Marianne Arruda, Doris J Kistler, and Frederic L Wightman, "Localization using nonindividualized head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 94, pp. 111, 1993.

[3] V Ralph Algazi, Richard O Duda, Reed P Morrison, and Dennis M Thompson, "Structural composition and decomposition of hrtfs," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 103–106.

[4] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 99–102.

[5] Lord Rayleigh, "Xii. on our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.

[6] Futoshi Asano, Yoiti Suzuki, and Toshio Sone, "Role of spectral cues in median plane localization," *The Journal of the Acoustical Society of America*, vol. 88, pp. 159, 1990.

[7] Jack Hebrank and D1 Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, pp. 1829, 1974.

[8] Dwight W Batteau, "The role of the pinna in human localization," *Proceedings of the Royal Society of London. Series B. Biological Sciences*, vol. 168, no. 1011, pp. 158–180, 1967.

[9] Donald Wright, John H Hebrank, and Blake Wilson, "Pinna reflections as cues for localization," *The Journal of the Acoustical Society of America*, vol. 56, pp. 957, 1974.

[10] Mengqiu Zhang et al., "Experimental guided spherical harmonics based head-related transfer function modeling," 2013.

[11] Thushara D. Abhayapala, Rodney A. Kennedy, and Wen Zhang, "Horizontal plane hrtf reproduction using continuous fourier-bessel functions," in *Audio Engineering Society Conference: 31st International Conference: New Directions in High Resolution Audio*, Jun 2007.

[12] Vikas C. Raykar, Ramani Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *The Journal of the Acoustical Society of America*, vol. 118, no. 1, pp. 364–374, 2005.

[13] V Ralph Algazi, Richard O Duda, and Patrick Satarzadeh, "Physical and filter pinna models based on anthropometry," in *Audio Engineering Society Convention 122*. Audio Engineering Society, 2007.