FAST MODELLING OF PINNA SPECTRAL NOTCHES FROM HRTFs USING LINEAR PREDICTION RESIDUAL CEPSTRUM

Chaitanya Ahuja, and Rajesh M. Hegde

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

ABSTRACT

Developing individualized head related transfer functions (HRTF) is an essential requirement for accurate virtualization of sound. However it is time consuming and complicated for both the subject and the developer. Obtaining the spectral notches which are the most prominent features of HRTF is very important to reconstruct the head related impulse response (HRIR) accurately. In this paper, a method suitable for fast computation of the frequencies of spectral notches is proposed. The linear prediction residual cepstrum is used to compute the spectral notches with a high degree of accuracy in this work. Subsequent use of Batteaus Reflection model to overlay the spectral notches on the pinna images indicate that the proposed method is able to provide finer contours. Experiments on reconstruction of the HRIR indicates that the method performs better than other methods.

Index Terms— Individualized HRTF, Spectral Notches, Pinna Contours

1. INTRODUCTION

Spatial audio has been extensively explored in the past. After introduction of concepts involving ITD (Interaural Time Difference) and ILD (Interaural Level Difference), many issues such as Front-Back Ambiguity (FBA) and Inside-the-Head Localization (IHL) have emerged which require extensive application of spectral estimation techniques. Hypothesis proposed in [1] claims that FBA could be resolved with dynamic cues like head movement. Resolving IHL requires presence of accurate personalized HRTF along with reverberation effects of the room [2].

A popular method of customization of HRTF is creating structural models [3]. These models aim at relating spectral features to the anthropometry. Substantial work has been done to understand structural aspects in [4]. Psycho acoustical effects have been explored in [5] and [6]. Behavioral effects have been covered in [7], [8] and [6] while [9] discusses neurophysiological aspects.The observations are consistent with spectral notches being the dominant features.

Analysis of the distribution patterns of pressure nodes and anti-nodes on the pinna has been performed in [10]. Observations conclude that at the first spectral-notch frequency, one or two anti-nodes appear in the cymba and triangular fossa, and one node in the concha. Signal processing techniques involving LP residual and group delay have been used before in [3] to extract spectral notches, henceforth LP residual group delay (LPRGD) algorithm.

In this paper, a novel and fast algorithm for high-resolution extraction of spectral notches is proposed. HRTF's linearity and time-invariance has been assumed and hence we propose to use linear-

prediction residual cepstrum (LPRC). This method removes convolutional effects due to multiple reflections of sound waves in the pinna which makes it applicable to Batteau's Reflection Model [4].

The organization of the paper is as follows: Section 2 introduces spectral notches, reflection model, application of LPRC and illustrates simulation results using proposed algorithm. Section 3 demonstrates experiments to gauge performance using statistical analysis. The study is concluded in Section 4.

2. FAST MODELLING OF PINNA SPECTRAL NOTCHES

The motivation and methodology used is discussed in Section 2.1. Simulation results are shown in Section 2.2.

2.1. HRIR and Pinna Spectral Notches

Sound wave input is filtered by reflection and diffraction effects of head, torso and pinna (outer ear) before it reaches the ear-drum. These effects can be combined into a mathematical linear time-invariant system which is formally called Head-Related Impulse Response (HRIR) [11]. Fourier transform of HRIR is called Head-Related Transfer Function (HRTF) and is defined as

$$H(r, \theta, \phi, f) = \frac{\psi(r, \theta, \phi, f)}{\psi_0(f)} \qquad (1)$$

where H is HRTF of right/left ear, $\psi(r, \theta, \phi, f)$ is sound pressure on right/left ear drum and $\psi_0(f)$ is free-field sound pressure without subject. (r, θ, ϕ) are spherical coordinates corresponding to the position of the sound source and f is frequency. Various techniques to measure HRIR have been described in [11].

HRTF is modelled using an all-pole model on the lines of the work done in [3]. A good choice for the order is around 8 to 12 [12]. We use a 12th order linear-prediction (LP) analysis for all analysis in this paper.
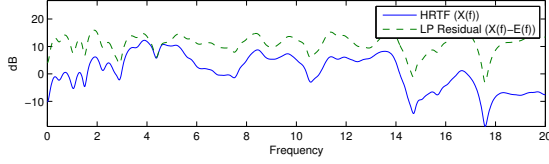
Assume that the nth point of the minimum phase, causal signal $x[n]$ is unknown. It can be predicted as a linear combination of k previous points in the signal, k being the order of the LP Residual.

$$e[n] = x[n] - \sum_{i=1}^{k} a_i x[n-i] \qquad (2)$$

$e[n]$ is error in the prediction of the signal. Minimizing expectation of the mean squared error, LP coefficients are obtained which can be used to compute an inverse filter $A(z)$

$$A(z) = 1 - \sum_{i=1}^{k} a_i z^{-i} \qquad (3)$$

Fig. 1. Illustration of HRTF and its LP Residual modelled by an all-pole transfer function (LP residual is shifted up by 10dB for clarity)

$A(z)$ is used to evaluate $E(z)$ from $H(z)$.

The choice of LP residual analysis is justified because it assumes a source filter model and estimates 3 components (1) all-pole model (2) residual, representing excitation of source of sound (3) gain, corresponding to the energy of the signal [12] [13]. A half hanning window is now applied to the obtained residual. This removes the fast changing components in frequency domain.

A slightly modified form of the real cepstrum transformation, henceforth Cepstrum, is performed on the signal $x[n]$ to give $c_x[n_q]$. The use of Cepstrum increases the resolution of the notches. Significant improvement seen in results is because of elimination of convolutional components of multiple reflections due to pinna.

$$c_x[n_q] = \mathcal{R}e \left(\text{IDCT} \left(\log_{10} \left(|\mathcal{F}\{x[n]\}| \right) \right) \right) \quad (4)$$

where $\mathcal{F}$ is discrete-fourier transform, IDCT is inverse-discrete cosine transform and $\mathcal{R}e$ is real part of the signal.

A half-rectangular lifter is now applied in the quefrency domain to remove the fast moving components in the cepstral domain. As DCT requires fewer coefficients to better approximate spectrum[14] than FFT, this gives an advantage in the time complexity of the LPRC algorithm. All the local minimas in the cepstral domain refer to the spectral notches of the given HRTF.

Reflection Model, as described in [4], [8] and [15], has been applied to overlay contour of spectral notches on the picture of pinna of a given individual. Consider a pinna being subjected to a sound wave $x[t]$. The total signal $y[t]$ received at the ear canal is the sum of direct signal $x[t]$ and reflected signal $ax[t - t_d(\theta)]$.

$$y[t] = x[t] + ax[t - t_d(\theta)] \quad (5)$$

where a is the reflection coefficient and $t_d(\theta)$ is the time delay. For destructive superposition of incident and reflected waves we have

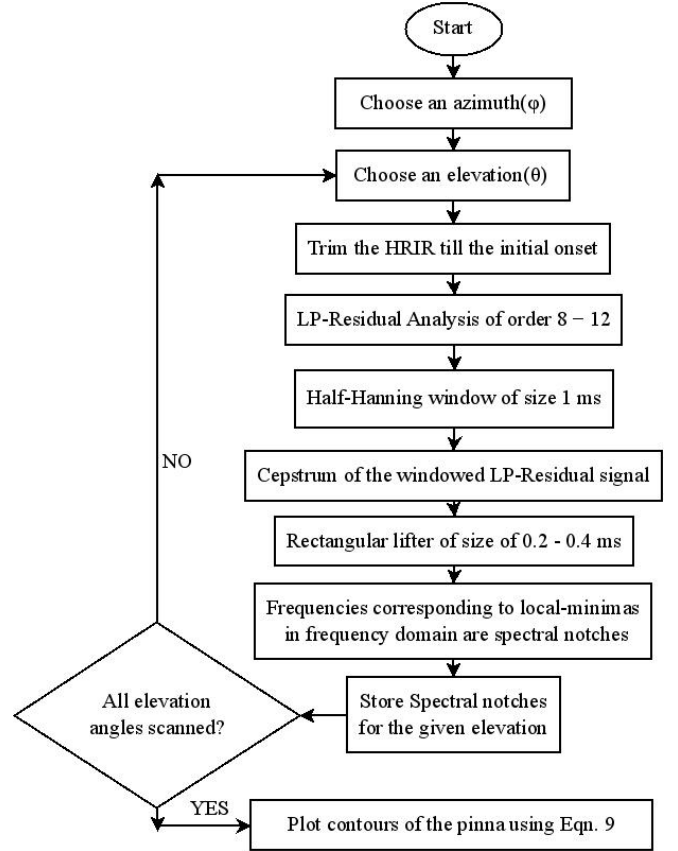$$t_d(\theta)2\pi f_n(\theta) = (2n + 1)\pi \quad \forall n = 0, 1, 2 \ldots \quad (6)$$

$$\Rightarrow f_n(\theta) = \frac{2n + 1}{2t_d(\theta)} \quad \forall n = 0, 1, 2 \ldots \quad (7)$$

$$\Rightarrow f_0(\theta) = \frac{1}{2t_d(\theta)} = \frac{c}{4d(\theta)} \quad (8)$$

Using Satarzadeh's argument[15], a factor of 2 is multiplied to Equation 8 to get

$$f_0(\theta) = \frac{c}{2d(\theta)} \quad (9)$$

where c is the speed of sound in air, $d(\theta)$ is the path difference between reflected and direct wave, $f_0(\theta)$ is the frequency of the first spectral notch and θ is the angle of elevation. Notches are overlaid on picture using points corresponding to $(d(\theta), \pi + \theta)$ with respect to the ear canal as the origin.

Fig. 2. Flow chart of LPRC algorithm for extraction and overlaying of spectral notches
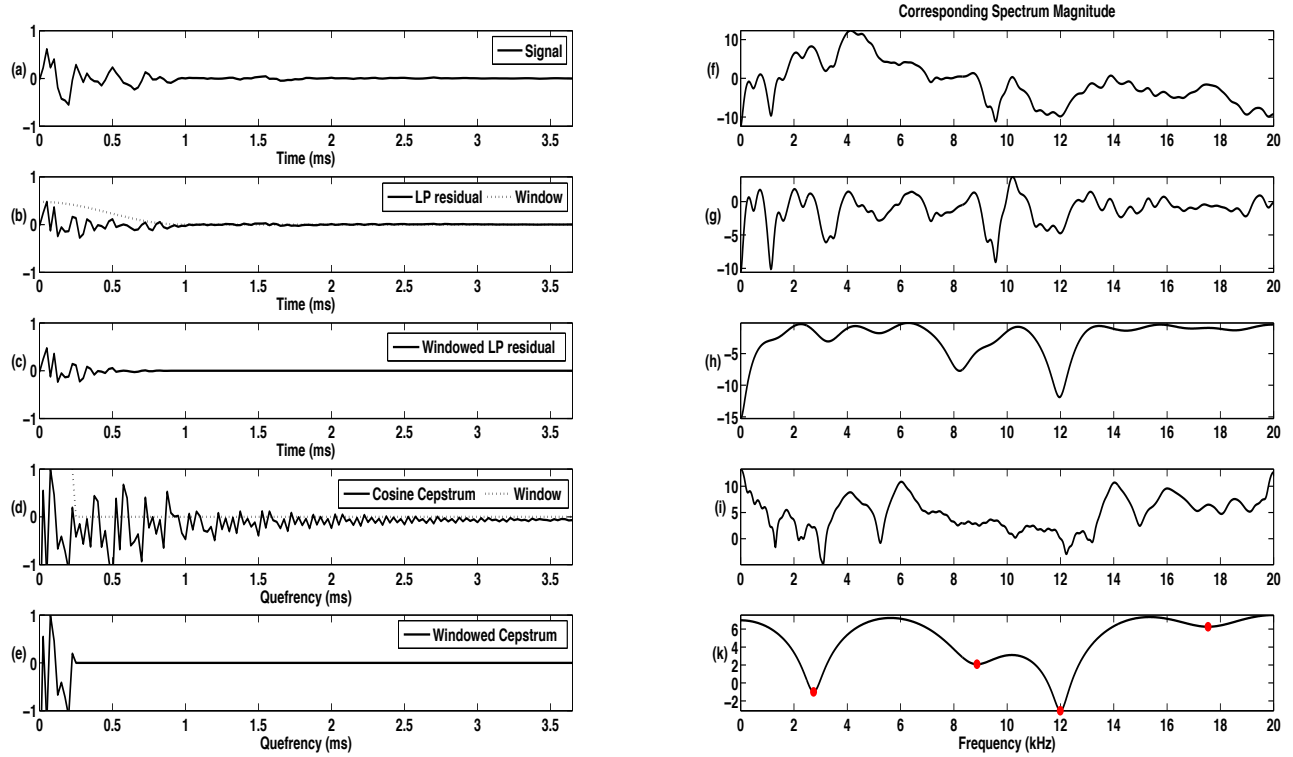
Algorithm 1, henceforth LP residual cepstrum (LPRC) algorithm, lists the steps involved in extraction and overlaying of spectral notches.

Algorithm 1 - LPRC method

1: **Initialization**: Choose a particular azimuth (ϕ) for estimation of pinna contours.
2: Choose an elevation (θ) and trim the corresponding HRIR ($x_\theta[n]$) till the initial onset.
3: Perform LP-Residual Analysis of order $8-12$ on signal obtained from step 2.
4: Apply a half-hanning window of size 1 ms on the LP-Residual signal.
5: Perform Cepstrum operation of windowed LP-Residual signal.
6: Apply a half-rectangular lifter of a size of $0.2 - 0.4$ ms on the windowed LP-Residual Cepstrum.
7: Find out frequencies corresponding to the local-minimas in the frequency domain of the liftered LP-Residual Cepstrum. These correspond to the spectral notches of the given HRIR for the given elevation (θ).
8: **Repeat**: Step 2 through 6 to obtain spectral notch frequencies for different angles of elevation.
9: **Termination**: Using Satarzadeh's method (Equation 9), overlay spectral notches on pinna picture for various values of elevation to get the spectral-notch estimated contour.

Fig. 3. Application of LPRC algorithm for extracting spectral notches for $\theta = 0$ and $\phi = 0$ of subject 119, *Courtesy: CIPIC Database*. *Figure (a):* Original signal, *Figure (b):* LP residual of original signal, *Figure (c):* Half-hann window of previous signal, *Figure (d):* Cepstrum of windowed signal, *Figure (e):* Rectangular window of previous signal. Figure (f), (g), (h), (i), (k) refer to fourier transforms of Figure (a), (b), (c), (d), (e) respectively. Local minimas in *Figure (k)* correspond to frequencies of spectral notches.

2.2. Simulation Results

LPRC algorithm was used to calculate frequency of spectral notches on given HRIRs. Simulation results are illustrated in Figure 3.
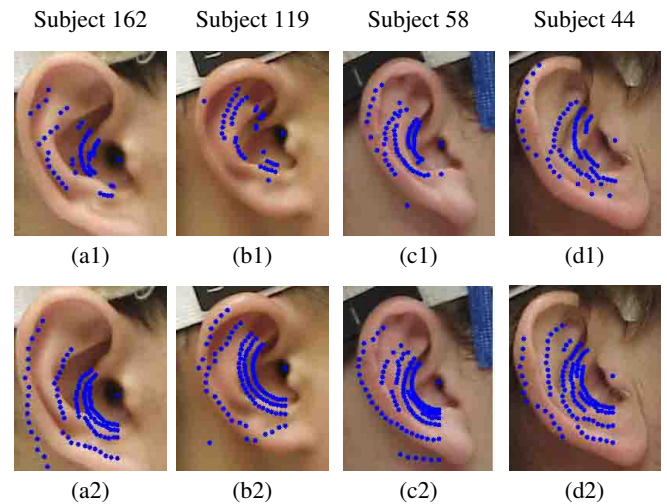
3. PERFORMANCE EVALUATION

The database used and experimental conditions are described in Section 3.1 and Section 3.2 respectively. Experiments on performance evaluation are conducted in Section 3.3 with statistical analysis in Section 3.4. Reconstruction of HRIR from spectral notches is discussed in Section 3.5.

3.1. Database

Publicly available CIPIC Database [16] has been used as the database for testing the algorithm. It provides HRIRs for 45 subjects at 50 different elevation and 25 azimuth angles. Contour overlaying on the ear was done on anthropometric measurements and ear pictures provided in the database.

HRIRs were measured [16] by placing a 1-meter radius hoop around the subject and moving the loudspeaker along the hoop at various elevation (θ) and azimuth (ϕ) angles. Ear canals of the subjects were completely blocked and microphones were placed just outside the ear canals. Finally, measured HRIRs were filtered by a modified hanning window to remove the reverberation caused due to walls of the room. This provides us with anechoic chamber Head-Related Impulse Responses.

Fig. 4. Illustration of pinna images with contours overlaid on them. (a1) through (d1) are generated using LPRGD algorithm [3]. (a2) through (d2) are using LPRC algorithm.

Table 1. Analysis of Depth-Bandwidth Ratio and Average Error Deviation in Notch Distances for two methods

	LPRGD				LPRC			
	Notch Distance		DBR		Notch Distance		DBR	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
	(cm)	(cm)	(dB kHz^{-1})	(dB kHz^{-1})	(cm)	(cm)	(dB kHz^{-1})	(dB kHz)$^{-1}$
Female	0.1496	0.1474	2.7600	8.1947	0.0511	0.0848	8.9097	1746.6
Male	0.1481	0.1375	2.8900	8.5188	0.0349	0.0701	9.9529	1507.0

3.2. Experimental Conditions

HRIRs provided by *CIPIC Database* contain 200 data points. Their sampling rate is 44100 *points per second*. An anti-aliasing FIR filter was used to downsample the HRIRs to 40000 *points per second*. LP coefficients are calculated using Levinson-Durbin recursion that arises from least-squares formulation. It is also referred to as the auto-correlation method. Cepstrum is computed efficiently using 1024 point Fast-Fourier Transform (FFT) with padded zeros at the end.

The model described in [4] is used to compute the distance of spectral notches from the ear canal. Anthropometric measurements are provided in the database which is used to calculate the distance to pixel ratio. There is currently no data of the position the ear canal so that has been done manually. Aforementioned data is used to plot spectral notches on the given ear picture which gives a visual appearance of contours of walls of a pinna.

3.3. Experiments on extracting Pinna-Contours

Frequency of spectral notches are converted to distances from the ear canal using Equation 9 and then shown on image of ears of subjects (Figure 4) in CIPIC Database. It can be seen that LPRC produces more accurate contours than its predecessor [3].

3.4. Statistical Analysis

As no reference of the original contour is available, contours on the walls of the pinna were marked manually on the image. Using Equation 9 frequency of actual spectral notches were estimated. These values were used as a reference to calculate error in distance of notches from ear canal.
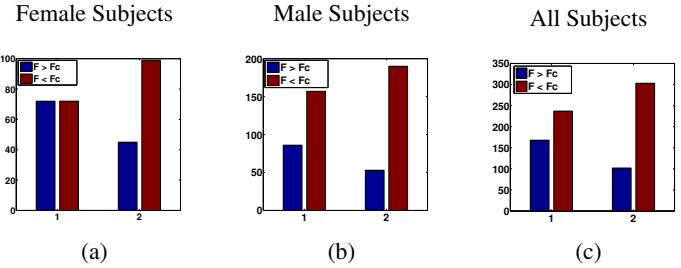
Depth-Bandwidth ratio (DBR) was also calculated for each notch. DBR is defined as

$$DBR = \frac{Depth}{3dB\ Bandwidth} \qquad (10)$$

where Depth is the distance of a notch from the closest maxima, projected onto dB Axis and 3dB Bandwidth is distance between double-power points of every notch. Mean and variance of DBR and average error deviation (AED) of the notch distances were calculated for male and female subjects separately (Table 1). LPRC method provides a statistically better mean and variance of AED for both male and female subjects. Higher values of mean of DBR are observed with LPRC. This suggests that at an average, notches generated by LPRC are more sharp than by LPRGD.

3.5. Reconstruction of HRIR using Spectral Notches

Spectral notches, being one of the primary features of any given HTRF, are used to reconstruct HRIR for rendering 3D Audio. Frequencies of extracted notches are used with a fixed pre-decided bandwidth to synthesize an all-pole filter. This filter is excited

Female Subjects Male Subjects All Subjects

Fig. 5. Comparison of LPRGD and LPRC with Original Spectrum, after reconstruction of HRIR from notches. *Y-Axis: Number of test-cases*

by an impulse train to generate the HRIR. The synthesized HRIR is compared with the original spectrum by Analysis of Variance (ANOVA).

ANOVA returns a decision, in form of a F-statistic Value (F), concerning rejection of null hypothesis based on whether the 2 populations come from the same gaussian population with same variance. Depending on the number of degrees of freedom of numerator (n_f) and denominator (d_f), and significance level, critical F-stat value (F_c) value is determined. If $F > F_c$ the null hypothesis is rejected.

With reference to Figure 5, ANOVA is performed on HRIRs with elevation = [-5.625 0 5.625] $\times$ azimuth = [-5 0 5], $n_f = 1$, $d_f = 1000$ and sensitivity = 5%, which implies that $F_c = 3.85$. Results are calculated for $P < 5\%$ with $F_c = 3.85$. Bar graph 1 and 2 of each sub-figure illustrate analysis on HRIR corresponding to spectral notches extracted by LPRGD and LPRC respectively. Rejection of null hypothesis is more prominent in reconstructed HRIR of female subjects and HRIR synthesized using LPRGD. As a whole LPRC provides spectral notches, which reconstruct the HRTF more accurately than LPRGD.

4. CONCLUSION

In this study, an algorithm for extraction of spectral notches using the LP residual cepstrum is proposed. This method is characterized by the assumption that pinna constitutes a linear time-invariant system while the HRTF is an all-pole model. This algorithm can also be applied to obtain spectral peaks due to resonance effect and constructive addition of waves at the ear canal. This work is one small step to the large body of work that is exploring methods to obtain individualized HRTF from pinna geometry and an existing database of HRTFs. Future work includes analyzing HRTFs for more significant features which can be related to ear structure. Spectral notches and peaks combined with other features would can provide enough data points for accurate reconstruction of HRIR using ear anthropometry. Multichannel techniques can also be explored to provide a robust framework for developing personalized spatial audio applications.

5. REFERENCES

[1] Hans Wallach, "The role of head movements and vestibular and visual cues in sound localization," *J. Exp. Psychol.*, vol. 27, no. 4, pp. 339 to 68, 1940.

[2] Toni Liitola, *Headphone sound externalization*, Ph.D. thesis, Helsinki University of Technology, 2006.

[3] 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.

[4] D. 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, August 1967.

[5] Brian C. J. Moore, Simon R. Oldfield, and Gary J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 khz," *The Journal of the Acoustical Society of America*, vol. 85, no. 2, pp. 820–836, 1989.

[6] Donald Wright, John H. Hebrank, and Blake Wilson, "Pinna reflections as cues for localization," *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 957–962, 1974.

[7] Mark B. Gardner and Robert S. Gardner, "Problem of localization in the median plane: effect of pinnae cavity occlusion," *The Journal of the Acoustical Society of America*, vol. 53, no. 2, pp. 400–408, 1973.

[8] Jack Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?," *The Journal of the Acoustical Society of America*, vol. 56, no. 3, pp. 935–938, 1974.

[9] Daniel J Tollin and Tom CT Yin, "Spectral cues explain illusory elevation effects with stereo sounds in cats," *Journal of neurophysiology*, vol. 90, no. 1, pp. 525–530, 2003.

[10] Hironori Takemoto, Parham Mokhtari, Hiroaki Kato, Ryouichi Nishimura, and Kazuhiro Iida, "Mechanism for generating peaks and notches of head-related transfer functions in the median plane," *The Journal of the Acoustical Society of America*, vol. 132, no. 6, pp. 3832–3841, 2012.

[11] Durand R. Begault, "3-d sound for virtual reality and multimedia," *NASA*, April 2000.

[12] Sree Hari Krishnan Parthasarathi, Herv Bourlard, and Daniel Gatica-Perez, "Lp residual features for robust, privacy-sensitive speaker diarization.," in *INTERSPEECH*. 2011, pp. 1045–1048, ISCA.

[13] Charles Lueck, "Efficient modeling of the head-related transfer functions," *International Community for Auditory Display*, 1994.

[14] Qifeng Zhu and Abeer Alwan, "An efficient and scalable 2d dct-based feature coding scheme for remote speech recognition," in *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*. IEEE, 2001, vol. 1, pp. 113–116.

[15] Patrick Satarzadeh, *A study of physical and circuit models of the human pinnae*, Ph.D. thesis, Citeseer, 2006.

[16] V.R. Algazi, R.O. Duda, D.M. Thompson, and C. Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, 2001, pp. 99–102.