Extraction of Pinna Spectral Notches in the Median Plane of a Virtual Spherical Microphone Array

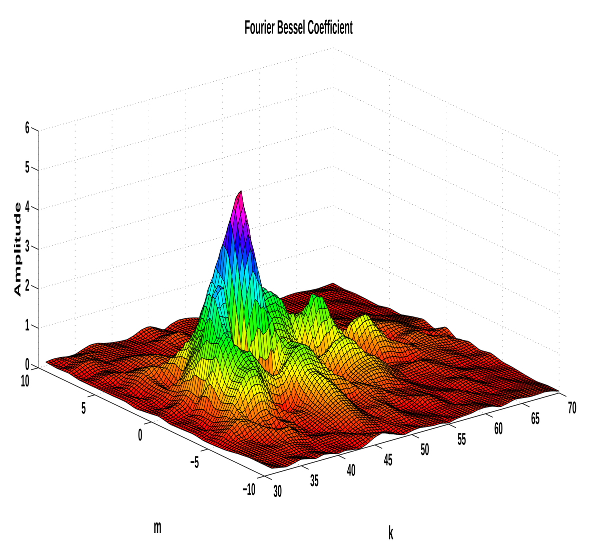
Ankit Sohni, Chaitanya Ahuja, and Rajesh M Hegde

Indian Institute of Technology Kanpur, India

Introduction

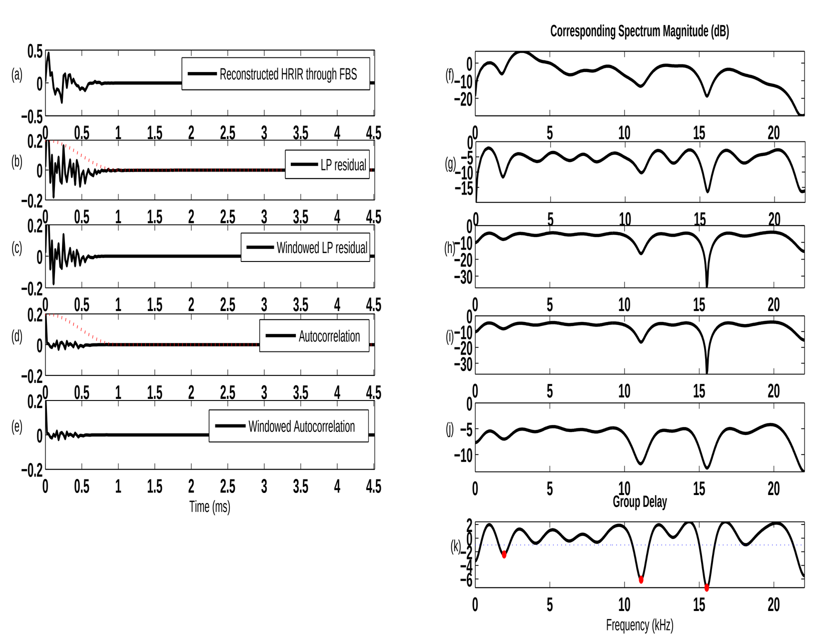
- Head Related Impulse Response (HRIR) captures the effects of interaction of sound with human anatomy.
- Head diffraction causes ITD and ILD between sound waves arriving at both ears which are the primary binaural cues in horizontal plane localization.
- The effect of head is invariant in the median plane as both the binaural cues (ITD and ILD) are nearly equal to zero.
- Pinna geometry causes multiple reflections of sound wave, and the delay between direct wave and the wave reflected by pinna wall results in periodic spectral notches.
- Head Related Transfer Function (HRTF) corresponding to measured HRIR are simulated by FBS over the median plane, and spectral notches are extracted from reconstructed HRTF.
- These spectral notches smoothly vary with elevation angles, and are highly dependent on pinna dimensions.

Choice of Truncation number



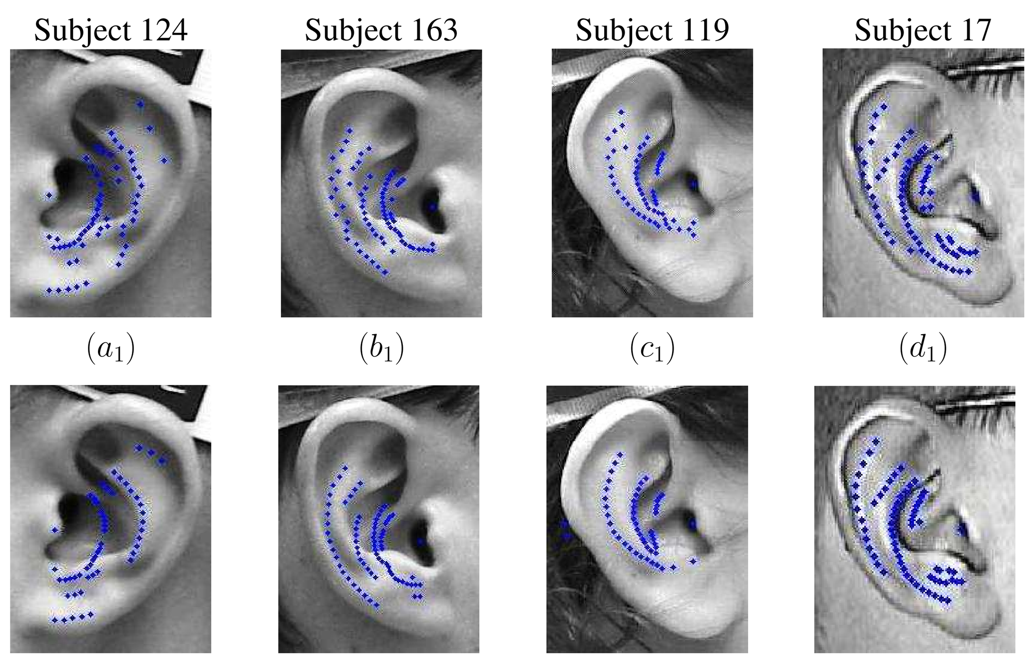
- The modal parameter C_{mk} are band limited and preserve negligible energy after some truncated value $|m| > M$ and $k > K + K'$.
- C_{mk} corresponding to first K' roots of Bessel function preserve faint initial pulse which do not contribute any structural feature of HRIR.
- C_{mk} corresponding to next K roots preserve much of variations due to pinna alone, and are very significant for pinna spectral notches.
- In CIPIC database, It is found that convergence is achieved for $M = 10, K' = 30$ and $K = 40$.

Extraction of Pinna Spectral Notches



- HRIR reconstructed through Fourier Bessel Series only highlights the effects of pinna resonances and notches.
- LP residual of reconstructed HRIR removes the pinna resonances while retains the pinna spectral nulls.
- Windowing the LP Residual of reconstructed HRIR makes the spectrum smoothen while preserving the pinna spectral notches.
- Auto-correlation of windowed LP residual preserves most of the details of spectral envelop such as notch depth and bandwidth.
- Due to high frequency resolution property of group delay function, pinna spectral notches are extracted from the group delay of the windowed auto-correlation function.
- Threshold of -1 is empirically chosen in order to avoid spurious nulls caused by windowing.

Pinna Notches marked on ear contour



(a_1) (b_1) (c_1) (d_1)

(a_2) (b_2) (c_2) (d_2)

Plane wave Decomposition

- HRTF recorded by spherical array of microphones due to source located at the entrance of ear canal can be decomposed into spherical harmonics as

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

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

$$0 \le \theta \le \pi, 0 \le \phi < 2\pi$$

- Under the far field assumption ($r > 1m$), HRTF will be independent of range r and can be represented as

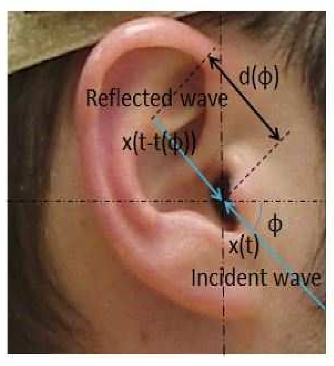
$$H(f; \theta, \phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} H_n^m(f) Y_n^m(\theta, \phi) \tag{3}$$

where $H_n^m(f)$ is Spherical Fourier Transform (SFT).

- Alternatively, the far field HRTF can be decomposed into its corresponding Legendre polynomial and complex exponential as

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

Pinna Reflection Model



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

$$y(t) = x(t) + \rho(\phi) x(t - t(\phi)) \tag{9}$$

$$\text{or } Y(e^{j\omega}) = (1 + \rho(\phi) e^{-j\omega t(\phi)}) X(e^{j\omega}) \tag{10}$$

- The elevation dependent temporal delay $t(\phi)$ results the point of reflection in the pinna image at a distance given by

$$d(\phi) = \frac{ct(\phi)}{2} \tag{11}$$

- It also results in the periodic spectral notches whose frequencies (assuming $\rho(\phi) > 0$) are given by

$$f_n(\phi) = \frac{2n+1}{2t(\phi)} = \frac{c(2n+1)}{4d(\phi)}, \forall n = 0, 1, 2, \cdots \tag{12}$$

- The first spectral notch frequency occurs at $f_0(\phi) = \frac{c}{4d(\phi)}$
- Assuming Satarzadeh's hypothesis of negative reflection coefficient ($\rho(\phi) < 0$), the spectral notch frequency gets doubled as

$$f_0(\phi) = \frac{c}{2d(\phi)} \tag{13}$$

Experiments on Pinna Spectral Notches

- Publicly available CIPIC database is used where the data-set of several subjects with their pinna images and corresponding anthropometry parameters are available.
- HRIRs are measured using head-centered interaural polar coordinate system with elevation uniformly sampled from $-45°$ to $230.625°$ in the median plane.
- Based on prior researches, Pinna spectral notch frequencies are assumed to appear in frequency range from 5 kHz to 16 kHz, and are extracted from robust signal processing techniques.
- Pinna image of particular subject is uniformly scaled in order to match with pinna parameters such as d_5 (pinna height) and d_6 (pinna width).
- The distance $d(\phi)$ between pinna reflection point and the entrance of the ear canal is calculated from Equation 13 for frontal median plane $\phi \in [-45° \; 90°]$.
- 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.

Conclusion

- A fast method to extract accurate pinna spectral notches that follow the actual pinna wall structure is proposed.
- The main novelty of the proposed work is the efficient reconstruction of HRIR over the median plane of a virtual spherical array simulated using the Fourier Bessel series, especially at lower elevation angles.
- HRIRs corresponding to lower elevation angles suffer from knee reflections which have slight contribution as compared to other anatomical reflections in the measured HRIR.
- The proposed method can suppress the knee reflections due to capability of preserving strong variations of pinna alone under finite truncation.
- The pinna spectral notches extracted are also very accurate and smooth when compared to conventional spherical array based approach.
- The proposed method is robust to extract the pinna spectral notches even if HRIR is measured over the complete hemisphere.

HRTF Modeling over Median Plane

- In terms of convergence and computational complexity, complex exponents are better choice as compared to associated Legendre polynomial to represent HRTF over the median plane.
- Using head-centered interaural polar coordinate system, 3 dimensional HRTF in Equation 4 can be represented over the median plane ($\theta = \frac{\pi}{2}$) as

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

- The spectral component $C_m(f)$ can be modeled by the family of Bessel functions of first kind as

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

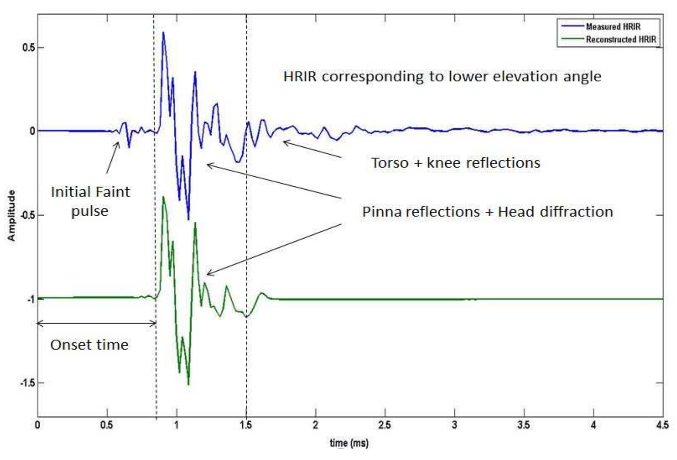
- Combining Equations 5 and 6, median plane HRTF can be decomposed into Fourier Bessel Series as

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

where C_{mk} represent Fourier Bessel Coefficient, and are calculated as

$$C_{mk} = \frac{1}{\pi [J_{|m+1|}(\beta_k^{|m|})]^2} \int_0^{f_{max}} \int_{-\pi}^{\pi} f H(f, \phi) J_{|m|}(\beta_k^{|m|} \frac{f}{f_{max}}) \cdots e^{-jm\phi} df \, d\phi \tag{8}$$

Reconstructed HRIR



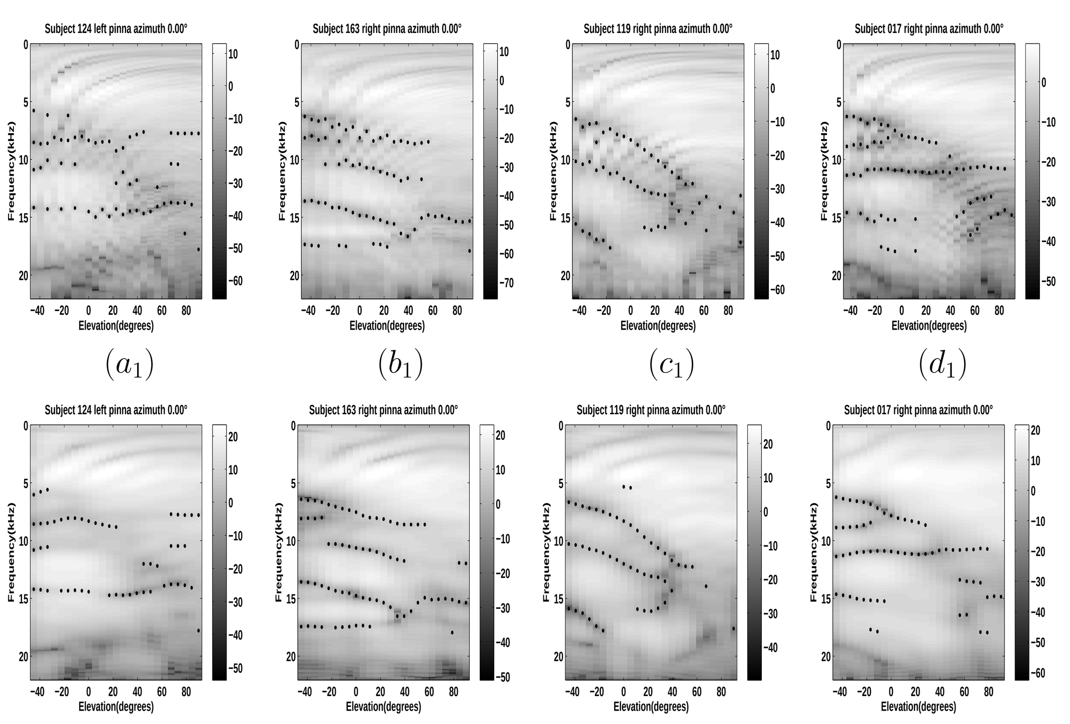
- The Fourier Bessel Coefficients in Equation 7 are calculated from discrete spatial and spectral HRTF measured over the hemispherical median plane as

$$C_{mk} = \frac{1}{\pi [J_{|m+1|}(\beta_k^{|m|})]^2} \sum_{f_i=0}^{f_{max}} \sum_{\phi_i=-\frac{\pi}{4}}^{\frac{5\pi}{4}} f_i H(f_i, \phi_i) J_{|m|}(\beta_k^{|m|} \frac{f_i}{f_{max}}) e^{-jm\phi_i} \tag{14}$$

$$|m| \le M, K' < k < K' + K$$

- Measured HRIR is composed of head diffraction, pinna and torso reflections, and as an artifact, knee reflection.
- In the lower elevation angles, this knee reflection appears within 1 ms time window along with pinna reflections.
- HRIR reconstructed through Fourier Bessel Series only preserves the pinna reflections that appear within 0.5 ms window range.

Pinna Spectral Notches overlaid on HRTF



(a_1) (b_1) (c_1) (d_1)

(a_2) (b_2) (c_2) (d_2)

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.