Extracting frequencies of the pinna spectral notches in measured head related impulse responses

Vikas C. Raykar[a] and Ramani Duraiswami[b]
Perceptual Interfaces and Reality Laboratory, Institute for Advance Computer Studies, Department of Computer Science, University of Maryland, College Park, MD 20742, USA

B. Yegnanarayana[c]
Department of Computer Science and Engineering, Indian Institute of Technology, Madras, Chennai-600036, Tamilnadu, India

(Dated: July 13, 2004)

Experimental data is available for the Head Related Impulse Response (HRIR) for several azimuth and elevation angles, and for several subjects. The measured HRIR depends on several factors such as reflections from body parts (torso, shoulder and knees), head diffraction, and reflection/diffraction effects due to the external ear (pinna). Due to the combined influence of these factors on the HRIR, it is difficult to isolate features thought to be perceptually important (such as the frequencies of pinna spectral notches) using standard signal processing techniques. Signal processing methods to extract the frequencies of the pinna spectral notches from the HRIR are presented. The techniques are applied to extracting features from the publicly available CIPIC[1] HRIR database. A brief discussion relating the notch frequencies to the physical dimensions and the shape of the pinna is given.

I. INTRODUCTION

Humans have an amazing ability to localize a sound source, i.e., determine the range, elevation and azimuth of the sound source relative to them.[2–4] The mechanisms responsible for the localization ability of the human hearing system have been fairly well understood though not completely. Interaural Time and Level Differences (ITD and ILD) are known to provide primary cues for localization in the horizontal plane, i.e., azimuth of the sound source.[4–7] However these differences do not account for the ability to locate sound for positions in the so called *cone of confusion*, which have the same ITD cues[63]. This can be explained in terms of the spectral filtering provided by the torso, head and the pinnae. This filtering is usually described by a complex frequency response function called the Head Related Transfer Function (HRTF). For a particular sound source location, the HRTF is defined as the ratio of the complex sound pressure level(SPL) at the eardrum to the SPL at the location of the center of the head when the listener is absent[64]. All acoustic cues resulting from the scattering of the incident sound by the torso, shoulder, head, and pinnae are captured by the HRTF. The corresponding impulse response is called the Head Related Impulse Response (HRIR).

The prominent features contributed by the pinna are the sharp spectral notches and the peaks. There is substantial psychoacoustical,[8,9] behavioral,[10–12] and neurophysiological[13–16] evidence to support the hypothesis that the spectral notches caused due to interaction of the sound with the pinna are important cues for vertical localization, i.e., elevation of the source[65].

The focus of the work presented in this paper is to automatically extract the frequencies corresponding to the spectral notches[66]. The experimentally measured HRIR includes the combined effects of the head diffraction and shoulder, torso, and as an artifact, the knee reflection. We develop signal processing techniques to extract the frequencies of the spectral notches due to the pinna, guided by our prior knowledge of the physics of the problem.

The temporal and spectral features in the HRIR/HRTF useful for localization have been previously studied in different ways. Contributions of different parts such as the pinna, head, torso and shoulder to these features have been studied using the KEMAR mannequin.[17,18] Analytical solutions were obtained using simple geometrical models for the head and torso.[19–22] While good geometrical models exist for the effects of head, torso and shoulders, a simple model for the pinna that connects pinna anthropometry to the features in the HRIR does not exist. One difficulty in developing such models is that it is difficult to automatically extract these frequencies from measured data. Once we have the frequencies of the spectral peaks and notches, a model for the pinna could be build by relating them to the shape and the anthropometry of the pinna.

Several studies were made to approximate the HRTFs by pole-zero models.[23–27] These studies fit a pole-zero model

[a] Electronic address: vikas@umiacs.umd.edu; URL: http://cvl.umiacs.umd.edu/users/vikas/
[b] Electronic address: ramani@umiacs.umd.edu; URL: http://www.umiacs.umd.edu/~ramani/
[c] Electronic address: yegna@cs.iitm.ernet.in; URL: http://speech.cs.iitm.ernet.in/~yegna/; This work was performed when the author was visiting the University of Maryland, College Park.

based on suitable error measure. The perceptual significance of the resulting poles and zeros is not obvious in most cases. Ideally, we would like to fit a model to account for the perceptually relevant features in the HRIR/HRTF. The model would typically require high orders to capture the nulls due to torso/shoulder reflection. Moreover, even with increased model order it is not guaranteed that the perceptually relevant spectral notches will be captured. Most of the above studies also do not provide measurable values for the features in the experimental data like frequencies of spectral notches.

In Section II we give a summary of the previous experimental, behavioral and neurophysiological results available on dummy models, humans and animals. In Section III we discuss signal processing techniques we will use for extracting the notch frequencies. Application of these techniques to HRIR data is illustrated in Section IV. In Section V two methods are proposed to extract the frequencies of spectral notches corresponding to pinna. The methods are based on the residual of a linear prediction model, windowed autocorrelation functions, group-delay function and all-pole modelling. In Section VI the method is applied to the HRIR data for different subjects in the public domain CIPIC[1] HRIR database. In Section VII we suggest further studies that can be carried out using the extracted frequencies of the pinna spectral notches. We also discuss methods for relating the frequencies of the spectral notches due to pinna to the shape and physical dimensions of the pinna.

II. ANALYSIS OF HRIR

The analysis of the HRIR is usually performed using experimental data obtained with real subjects,[1,28] experimental data obtained with the KEMAR or other mannequin,[1,29] theoretical data obtained with some simplified geometrical models[17–19,30–33] and data obtained from numerical modelling.[33–35] While we mainly discuss how different anatomical parts contribute to different features, we also provide a brief discussion as to whether each of these features are relevant for the perception of the direction of sound. Previous studies that discuss the relevance of different features for the perception of sound location can be classified as: controlled psychoacoustical,[8,9] behavioral[10–12] and neurophysiological[13–16,36] experiments done both on humans and animals.

The HRIRs used for analysis in our paper were taken from the CIPIC database.[1] This is a public domain database of high spatial resolution HRIR measurements along with the anthropometry for 45 different subjects. The azimuth is sampled from -80^o to 80^o and the elevation from -45^o to $+230.625^o$ in a head-centered interaural polar coordinate system. The interaural axis is the line passing through the center of the left and the right ears. The origin of this spherical coordinate system is the interaural midpoint, which is exactly the midpoint of the line joining the two ears. This point is usually somewhat below and behind the center of the head. The azimuth angle θ is the angle between a vector to the sound source and the vertical median plane or the midsagittal plane, and varies from -90^o to $+90^o$. An azimuth of $+90^o$ corresponds to the right side of the subject and -90^o corresponds to the left side of the subject, while $\theta = 0^o$ defines the midsagittal plane. The elevation φ is the angle from the horizontal plane to the projection of the source into the midsagittal plane, and varies from -90^o to 270^o. The elevation sequence -90^o,0^o,90^o,180^o and 270^o corresponds, respectively, to locations below, in front of, above, back and below the subject.

For any given azimuth, we form a 2-D array, where each column is the HRIR or the HRTF for a given elevation, and the entire array is displayed as an image[67]. This method of visualization helps identify variation of different features with elevation. Figure 1 shows the HRIR and HRTF images (for all elevations) corresponding to azimuth 0^0 for the right ear for subject 10 in the CIPIC database. In Figure 1 (a) the gray scale value represents the amplitude of HRIR, and in Figure 1 (b) it is the magnitude of the HRTF in dB. The different features, thought to correspond to different structural components are also marked by hand.

Composition of the responses in terms of head diffraction effects, head and torso reflection, pinna effects and the knee reflection artifact can be seen both in the time domain and in the frequency domain. Most of the features marked in Figure 1 were confirmed experimentally with the KEMAR mannequin, where the responses were measured by removing and adding different parts like the pinna, head and torso.[18] Consider the HRIR image plot as shown in Figure 1 (a). Three distinct ridges which are marked as 1, 2 and 3 can be seen in the HRIR image plot. The first distinct feature is due to the direct sound wave that reaches the pinna. The difference between the time of arrival for the left and the right ear is the Interaural Time Difference (ITD). It can be seen that the ITD has a slight dependence on the elevation. At the gross level the Interaural Time Difference (ITD) provides the significant binaural effect, which enables us to determine the direction in the horizontal plane, i.e, the azimuth of the source. Experimental studies, and also the correlation of the measured delay and the estimated delay from the head measurements using the spherical head model[37] or the ellipsoidal head model,[38] support this view. The Interaural Level Difference (ILD) between the left and the right ear also helps to reinforce the horizontal localization. The duplex theory[5] explains the horizontal localization of the sound in terms of the ITD and the ILD cues. It is likely that features in the interaural spectral level differences also might provide significant binaural cues.[7]

Referring to Figure 1 (a) we see that immediately after the direct wave, activity is seen in the close vicinity, which is due to diffraction of the sound around the head and the pinna. The corresponding diffraction pattern due to the head in the frequency domain can be explained by Lord Rayleigh's analytical solution for scattering from a sphere.[5,20] The diffraction pattern is not clearly visible in the HRTF image. The effect of head diffraction is more prominent in the contralateral HRTF than in the ipsilateral HRTF.[30]

The second valley shaped ridge between 1 ms and 2 ms is due to the reflected wave from the torso, reaching the pinna. The delay between the direct and the reflected sound from the torso is maximum above the head, and decreases on either side. This can be explained using simple ellipsoidal models

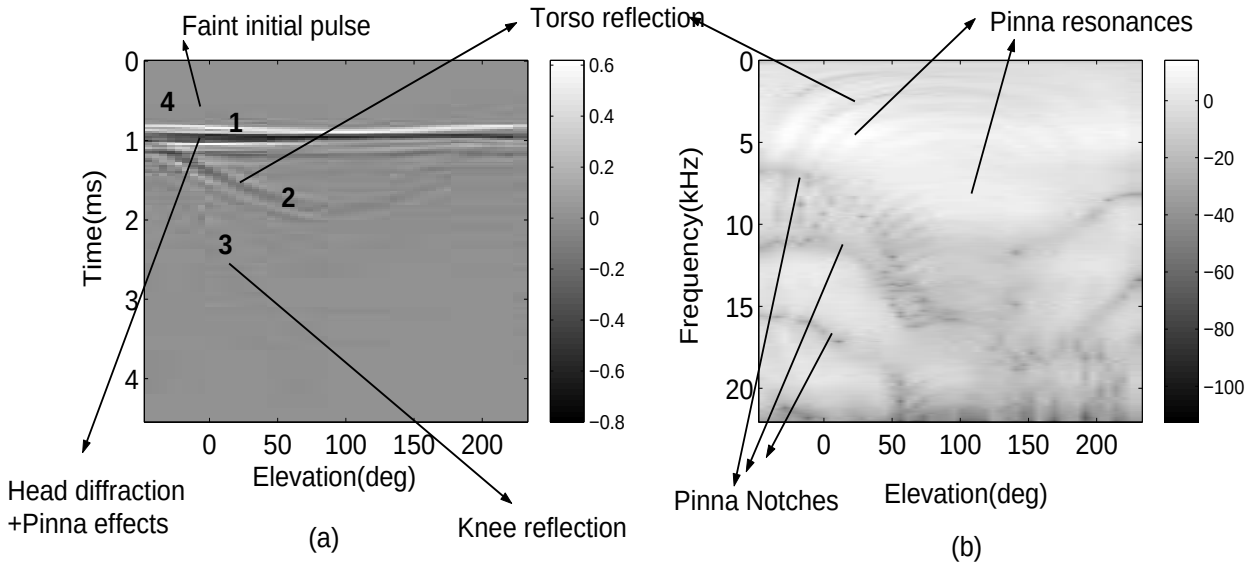

FIG. 1. a) HRIR and (b) HRTF images for the right ear for subject 10 in the CIPIC database for azimuth angle $\theta = 0^o$ for all elevations varying from -45^o to $+230.625^o$. The different features are marked in both the HRIR and the HRTF plots. In (a) the gray scale value represents the amplitude of HRIR, and in (b) the gray scale value is the magnitude of the HRTF in dB.

for the head and torso.[21] In the frequency domain the effect of this delay is the arch shaped comb-filter that can be seen throughout the spectrum (see Figure 1 (b)). The delayed component is not prominent for negative elevations due to shadowing by the torso.[21] Some studies have shown that the notches of the comb-filter in the low frequency range (< 3 kHz) could be used as a potential cue for vertical localization for low frequency sounds.[17]

The activity seen after 2 ms is due to knee reflections, since these measurements were done with the subjects seated.[1] This is confirmed by the observation that similar activity is not seen in the back ($\varphi > 90^o$). The other artifact is the faint pulse (marked 4 in Figure 1(a)) seen arriving before the main pulse. It is interesting to see that the location of the faint pulse varies with elevation. This is probably due to the nature of the ER-7C probe microphone used in the measurements.[1] The ER-7C microphone has a 76 mm silicone probe tube which conducts the acoustic wave to the microphone.[39] It is likely that the signal first hits the microphone outside before reaching the probe[68]. The maximum delay between the faint pulse and the direct pulse is around 0.25 ms which roughly corresponds to the length of the probe tube.

The other prominent features in the frequency domain, but difficult to in the time domain, are the prominent notches above 5 kHz. Three prominent notches can be seen in Figure 1(b) for elevations from -45^o to 90^o. As the elevation increases the frequency of these notches increases. Experiments with the KEMAR mannequin, in which the HRIRs were measured with and without the pinna,[18] confirm that these notches are caused due to scattering of acoustic wave by the pinna.

Various models have been proposed to explain the cause of these notches. Batteau[40] was the first to suggest that the structure of the pinna caused multiple reflections of sound, and the delay between the direct and the reflected sound varies with the direction of the sound source, providing a localization cue. These delays cause the notches in the spectrum. Hebrank and Wright[11,41] attributed the pinna spectral notches to the reflection of sound from the posterior concha wall. This idea was further refined by Lopez-Poveda and Meddis[42] who incorporated diffraction in the model, and predicted spectral notches that agreed closely with the measured ones. Their model predicts the elevation-dependent spectral features corresponding to the transverse dimensions of the concha. Specifically the first and the third nulls matched with their prediction. However the second null was not predicted. This notch could be due to the crus helias (the concha is partly divided by the crus helias), which they did not include in their model (See Figure 23(a) for a picture of the human pinna). They also showed that a simple delay model may not accurately predict the location of the notches (perhaps due to diffraction effects). The notches in the contralateral side can be explained if one assumes that the sound creeps around the head entering the contralateral concha at approximately the same elevation angle as the source was at in the ipsilateral hemisphere.[42]

Also present in the response are the resonances due to the pinna (the bright patches in the HRTF image in Figure 1 (b)), which were experimentally measured by Shaw[43] and numerically verified by Kahana and Nelson.[34] Shaw described six normal modes of resonances under the blocked meatus condition based on experimental measurements for 10 subjects.[43] The first mode is a simple quarter-wavelength depth resonance with uniform sound pressure across the base of the concha. It is strongly excited from all directions. The other modes are essentially transverse, and fall into two groups: A vertical pair (modes 2 and 3) and a horizontal triplet (modes 4, 5 and 6) (See Figure 26).

The question that naturally arises is which features of the pinna scattering contribute to the localization abilities.

Psychoacoustical experiments have demonstrated that high frequencies are necessary for localization in the vertical plane.[12,41,44] By progressively occluding the pinna cavities, it was shown that localization ability decreases with increasing occlusion.[12] Hoffman et. al[10,45] measured the localization ability of four subjects before and after the shapes of their ears were changed by inserting plastic moulds in the pinna cavity. Although localization of sound elevation was dramatically degraded immediately after the modification, accurate performance was steadily acquired again. Hebrank and Wright[11] showed that binaural and monaural subjects have similar difficulty in localizing unfamiliar sounds, and show that monaural subjects can easily be trained to localize as well as they would binaurally. These results suggest median plane localization is a monaural process and the pinna cues seem to be the dominant ones for elevation perception.

The spectral peaks and the notches are the dominant cues contributed by the pinna. Since the notch frequency varies smoothly with elevation, it is thought to be the main cue for perception of elevation. On the other hand, the spectral peaks do not show this smooth trend. However, it is likely that the presence or absence of the spectral peak could itself be a strong cue for the elevation. For example, the second normal mode identified by Shaw is excited strongly only for elevations around 90^0. This is consistent with experiments by Hebrank and Wright[41] where the one-octave peak between 7-9 kHz was identified as the *above cue*. Wright et. al.[9] present experiments to determine whether delays caused due to pinna reflections are detectable by humans. The results show that delay times of 20 μsec are easily recognizable when the amplitude ratio of the delayed signal to the leading signal is greater than 0.67. Just noticeable results agreed with the measurements of the minimum audible angle for monaural localization. Moore et al[8] describe experiments on the detection and discrimination of spectral peaks and notches. The results suggest that spectral peaks are more salient than spectral notches. However, changes in the center frequency of the notches are detectable even for rather narrow notches.

There is some neurophysiological evidence indicating that spectral notches are a dominant cue for perception of elevation. Experiments on cats suggest that single auditory nerve fibers are able to signal in their discharge rates the presence of a spectral notch embedded in bursts of noise or in continuous noise.[13] For high-spontaneous rate fibers, a moving-notch stimulus reduced the discharge below the spontaneous rate at and in the immediate vicinity of the most effective notch frequency. This suggests a mechanism by which localization ability is enhanced when there is relative motion between the sound source and the head. This vertical illusion that was observed in cats[15,36] can be explained well by a model that attributes vertical localization to recognition of the spectral shape cues. When cats were presented with simultaneous sounds located in the horizontal plane, they oriented to a location on the vertical midline about 8 degrees above the horizontal plane. The spectra resulting from the interaction of two simultaneous sound sources located in the horizontal plane is marked by a conspicuous spectral notch near 12 kHz. Examination of the cats filter function for single sound source showed that the filters for elevation near 8 degrees showed a notch near 12 kHz.

These studies suggest that the pinna spectral notches play a crucial role in the elevation perception.

III. SIGNAL PROCESSING TECHNIQUES

In this section we briefly review some important signal processing techniques which will be used in the later sections. While some of these methods are standard, we prefer to present them here, to collect all methods in one place, and to establish notation.

A. Spectrum analysis

The most obvious way of looking for features in any signal is by spectrum analysis.[46] The signals are sampled suitably, and the sampled discrete-time signals are used for further processing. The sampling frequency determines the highest frequency of interest, which, in the case of the HRIR data in the CIPIC database[1] is 22.05 kHz (sampling frequency is 44.1 kHz). Figure 2 shows the plot of 200 samples of a HRIR and the corresponding spectrum (HRTF) computed as the log squared magnitude of 1024-point Discrete Fourier Transform (DFT) of the 200 samples data with zeros padded. The N point squared magnitude spectrum of a sequence $x(n)$ is

$$X(k) = |\sum_{n=0}^{N-1} x(n)e^{-j(2\pi/N)kn}|^2.\ k = 0, 1, \ldots, N-1 \quad (1)$$

The number of points used to compute the DFT decides the spacing of the sampling points in the frequency domain. It does not add any new information, nor improve the resolution in the frequency domain, though it does change the visible fine structure in the spectrum. The resolution is determined by the number of signal samples used in the time domain, which in this case is 200. The spectral features are reflected both in the envelope of the magnitude spectrum and in the fine structure. The envelope information corresponds mostly to the dependence of the adjacent samples in the time sequence. The fine structure corresponds normally to the samples (or pulses) which do not have dependencies on adjacent samples. Periodic nulls in the fine structure of the spectrum correspond to impulses separated by a delay in the time domain. Aperiodic or isolated nulls are also possible due to noise and due to a signal with a short sequence of samples corresponding to the Finite Impulse Response (FIR) of a system. Typically features due to spectral nulls are affected significantly by additive noise to the signal.

B. Pole-zero modelling

The peaks and valleys of the spectral envelope are useful features for most signals. The frequencies of the peaks and valleys can be measured from the spectrum, but if done manually introduces errors due to human judgement. Since the spectral envelope is mostly dictated by the locations of peaks

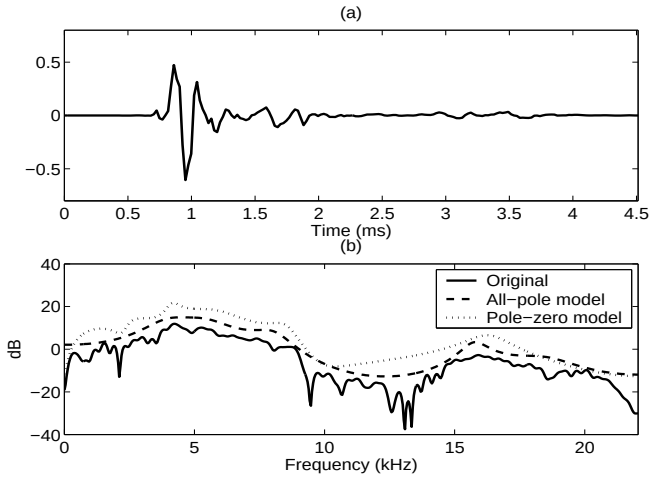

FIG. 2. (a) A typical HRIR for an elevation of 45^o and an azimuth of 0^o, (b) the log magnitude spectrum, a $(12, 12)^{th}$ order pole-zero model spectrum and a 12^{th} order all-pole model spectrum. In the plots the all-pole spectrum and the pole-zero spectrum are displaced vertically by 5 dB and 10 dB, respectively, for clarity.

and valleys, pole-zero models can be used to fit the data either in the time domain or in the frequency domain.[47,48]

If $x(n)$ is the output of a pole-zero model with input signal $u(n)$, then $x(n)$ can be predicted using linear combinations of past outputs and inputs, i.e.,

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + G\sum_{l=0}^{q} b_l u(n-l), \qquad (2)$$

where $\{a_k,\ k = 1,\ldots,p\}$, $\{b_l,\ l = 1,\ldots,q\}$ and $b_0 = 1$ are the parameters of the system, and G is the gain.[46] Taking the z-transform on both sides, the system transfer function is given by

$$H(z) = \frac{X(z)}{U(z)} = G\frac{1+\sum_{l=1}^{q} b_l z^{-l}}{1+\sum_{k=1}^{p} a_k z^{-k}}, \qquad (3)$$

where, $X(z) = 1+\sum_{l=1}^{q} b_l z^{-l}$ and $U(z) = 1+\sum_{k=1}^{p} a_k z^{-k}$ are the z-transforms of the $x(n)$ and $u(n)$, respectively and z is a complex variable. The spectrum of the pole-zero system (model) is given by evaluating $H(z)$ on the unit circle in the complex plane, i.e.,

$$|H(z)|^2_{z=e^{j\omega}} = G^2\left|\frac{1+\sum_{l=1}^{q} b_l e^{-j\omega l}}{1+\sum_{k=1}^{p} a_k e^{-j\omega k}}\right|^2. \qquad (4)$$

If all $\{a_k\}$s are zero, then the system is called an all-zero system or an FIR system. If all the $\{b_l\}$s are zero, then the system is called an all-pole system. If any of the $\{a_k\}$s are non-zero, then the system is called an IIR system.

If $x(n)$ is assumed to be the output of an all-pole system of order p, then it is given by,

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + Gu(n). \qquad (5)$$

We assume that $u(n)$ is unknown. Then the signal can be predicted approximately from a linearly weighted summation of the past p samples, i.e.,

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k). \qquad (6)$$

The error in the prediction is therefore given by,

$$e(n) = x(n) - \hat{x}(n) = x(n) + \sum_{k=1}^{p} a_k x(n-k). \qquad (7)$$

The sequence $e(n)$ is also known as the *Linear Prediction (LP) residual*. The total squared error is

$$E = \sum_n e(n)^2 = \sum_n [\, x(n) + \sum_{k=1}^{p} a_k x(n-k)\,]^2. \qquad (8)$$

Minimization of the mean squared error with respect to the coefficients $\{a_k\}$ gives the following normal equations[47]

$$\sum_{k=1}^{p} a_k R(n-k) = -R(k),\ \ k = 0, 1, \ldots, p, \qquad (9)$$

where $R(k) = \sum_n x(n)x(n-k)$ is called the autocorrelation function for a lag of k samples. Eq. 9 can be solved to get the coefficients $\{a_k\}$. Substituting the solution of the normal equations Eq. 9 into the expression for the error in Eq. 7 gives the sequence corresponding to the minimum total error, the LP residual. This is called linear prediction analysis.[47]

The system (or model) represented by the system function with linear prediction coefficients $\{a_k\}$ is given by

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k}, \qquad (10)$$

and is called the inverse (all-zero) system. The corresponding all-pole model is

$$H(z) = \frac{1}{A(z)} = \frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}}. \qquad (11)$$

From the response evaluated on the unit circle in the z-plane, we get the all-pole spectrum as

$$|H(\omega)|^2 = |H(z)_{|z=e^{j\omega}}|^2 = \left|\frac{1}{1+\sum_{k=1}^{p} a_k z^{-k}}\right|^2. \qquad (12)$$

The pole-zero model for the given data can be obtained using an iterative method.[48] Pole-zero models for the spectrum envelope can also be obtained using a pole-zero decomposition of the group-delay function derived from the cepstrum, which is the inverse DFT of the log magnitude spectrum.[49,50]

Pole-zero models are not effective in capturing the spectral nulls if there is additive noise in the signal data. All-pole modelling is also not effective for noisy signals, but all-pole models pick up the dominant spectral peaks most of the time. The order of the model determines the number of peaks in

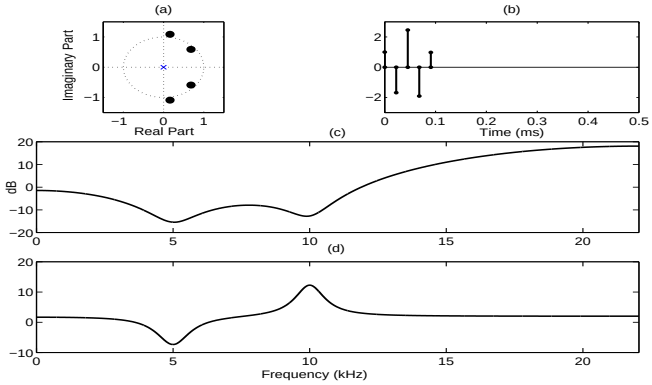

FIG. 3. Group delay function of an all-zero model. (a) Location of the zeros of an FIR system consisting of two pairs of complex conjugate zeros, (b) the corresponding FIR, (c) the log magnitude spectrum and (d) the group-delay function.

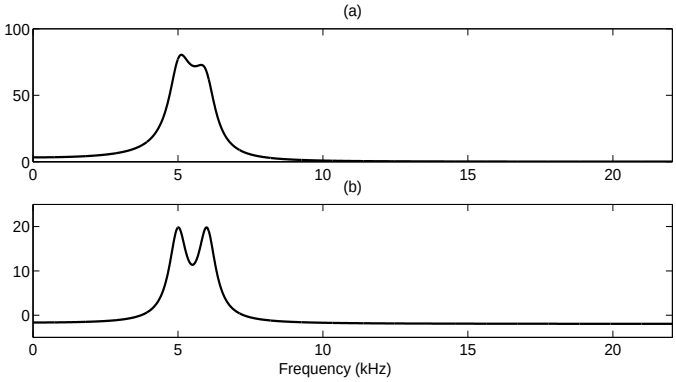

FIG. 4. Group delay function of an all-pole model. (a) The magnitude spectrum of an all-pole model with two poles and (b) the corresponding group-delay function.

the model spectrum. If the order is higher than desired, then there can be spurious peaks. Figure 2(b) shows the spectrum for a $(12, 12)^{th}$ order pole-zero model for the signal shown in Figure 2(a). Figure 2(b) also shows the spectrum for a 12^{th} order all-pole model for the same signal. In both cases, there is a good fit for the envelope of the spectrum, but important features such as nulls and peaks of the spectra are missing in the model spectra.

Pole-zero or all-pole models merely approximate the spectrum envelope, as best as they can, depending on the order of the model and the criterion used for approximation. Both the order and the criteria are independent of the nature of the signal being analyzed, and also the features expected to be highlighted. Thus these modelling techniques are unlikely to bring out the specific features one is looking in the HRIR signal. Moreover, the resolution of the peaks and the valleys in the spectral envelope depends on the order of the model. If the order is high, weak features also may be captured, but then there will be many spurious peaks and valleys, which are difficult to interpret.

C. Group-delay function

The group delay function is the negative of the derivative of the phase spectrum of a signal.[51,52] If $X(\omega)$ is the complex frequency response of a signal $x(n)$, then the group-delay function $\tau(\omega)$ is given by,

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} \tag{13}$$

where ω is the angular frequency, and $\theta(\omega)$ is the phase angle of $X(\omega)$. The group-delay function has some interesting properties which may help in extracting the spectral features better. The additive nature of the phase spectra of systems in cascade and the high frequency resolution properties of the group-delay functions help in providing better resolution of peaks and valleys in the spectral envelope even for short time segment of the data.[51] The group-delay function of a signal $x(n)$ can be computed directly using the Fourier transform of $x(n)$ and $nx(n)$ as follows.[53] Let $X(\omega)$ and $Y(\omega)$ be the Fourier transforms of $x(n)$ and $nx(n)$, respectively. Then

$$\begin{aligned} X(\omega) &= \sum_{n=0}^{N-1} x(n)e^{-j\omega n} = X_R(\omega) + jX_I(\omega), \\ Y(\omega) &= \sum_{n=0}^{N-1} nx(n)e^{-j\omega n} = Y_R(\omega) + jY_I(\omega). \end{aligned} \tag{14}$$

Since

$$\log X(\omega) = \log|X(\omega)| + j\theta(\omega), \tag{15}$$

the group-delay function can be written as

$$\begin{aligned} \tau(\omega) &= -\frac{d}{d\omega}[\theta(\omega)] = -Im(\frac{d}{d\omega}[\log X(\omega)]) \\ &= \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{X_R^2(\omega) + X_I^2(\omega)}, \end{aligned} \tag{16}$$

where $Im(z)$ corresponds to the imaginary part of z.

For an all-zero system, given by the FIR sequence shown in Figure 3(b), the group-delay function is shown in Figure 3(d). The distribution of the zeros in the z-plane is shown in Figure 3(a). The group-delay function in Figure 3(d) shows peaks for zeros outside the unit circle, and valleys for zeros inside the unit circle. The corresponding log magnitude spectrum is shown in Figure 3(c). Compared to the log magnitude spectrum, the group-delay function shows a better resolution of the zeros in the frequency domain due to the additive property.[51]

The group-delay function can also be computed for the response of an all-pole model. However, since the response is an IIR, truncation effects dominate in the group-delay function. If the coefficients of the all-pole model are known, then the group delay function can be computed for the inverse of the model (i.e., all-zero system), and then the negative of the function is used to obtain the exact group-delay function of the all-pole model. Figure 4 shows the group-delay function and the log magnitude spectrum of an all-pole model, to illustrate the high resolution property of the group-delay function.

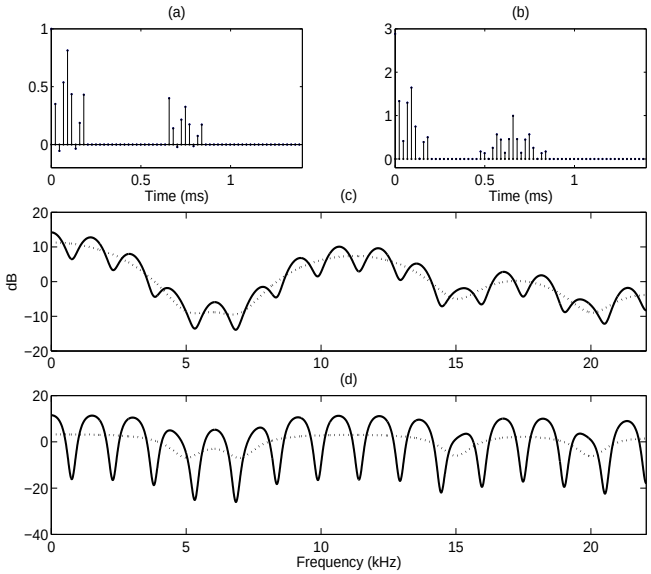

FIG. 5. (a) An FIR signal along with an attenuated delayed component, (b) the corresponding autocorrelation function, (c) the log magnitude spectrum and (d) the group-delay function. The dotted lines show the original spectrum and the group-delay function without the delayed component. The effect of the delay is to produce periodic notches in the spectrum.

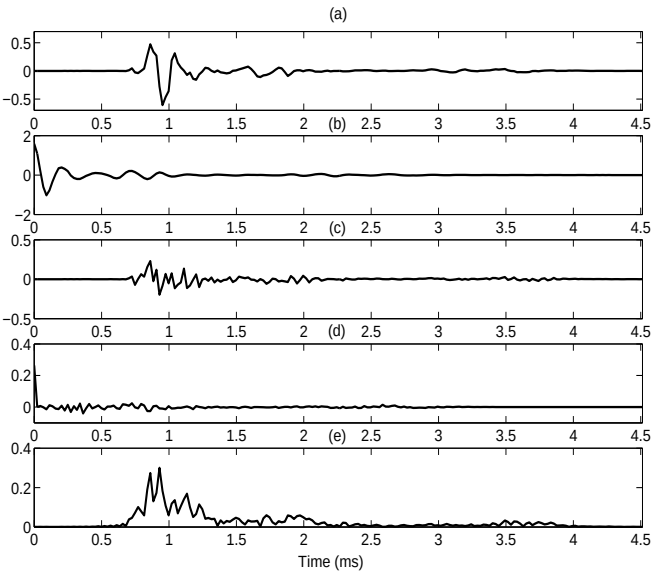

FIG. 6. (a) A HRIR signal, (b) its autocorrelation function, (c) its 12^{th} order LP residual, (d) autocorrelation function of the LP residual and (e) Hilbert envelope of the LP residual.

Since in this case all the roots are inside the unit circle in the z-plane, there are only peaks in the group-delay function, but these are extracted with higher resolution compared to the log magnitude spectrum.

The difficulty with the computation and interpretation of the group-delay function computed directly from the signal is that if there are roots (poles or zeros) lying close to or on the unit circle, the value of the group-delay function at these frequencies will be very large (positive or negative), masking the other roots, when observed at the the same scale in the plot. Moreover, windowing in the time domain produces sharp nulls in the spectrum, and hence large amplitudes in the group-delay function. Periodic nulls in the spectrum are also caused by strong impulses in the signal. The group-delay function preserves all the spectral details.

D. Autocorrelation function

The autocorrelation function $R(m)$ of a signal $x(n)$ of length N is given by

$$R(m) = \sum_{n=m}^{N-1} x(n)x(n-m), \;\; R(-m) = R(m) \qquad (17)$$

The autocorrelation function of a signal produces asymptotically reducing amplitudes, which may help in computing the group-delay function better, preserving most of the details of the spectral envelope, and mitigating the effects of windowing on the group-delay function. The autocorrelation function also helps in detecting the delays, if there is a delayed signal component in the signal. But the reflected signal should have the same characteristics of the original signal. Figure 5 shows a signal with a delayed component, and its corresponding autocorrelation function. The log magnitude spectrum and the group delay functions are also shown. The dotted line shows the original spectrum and the group-delay function without the delayed component. The effect of the delay is to produce periodic notches in the spectrum. The disadvantage of the autocorrelation function is that temporal details of the original signal are smeared, especially when the delayed signal is different from the main signal.

E. Linear Prediction Residual

As discussed in Section III B, Linear prediction analysis is used to derive an all-pole model for the given signal.[47] Linear prediction analysis can also be interpreted as the removal of redundancy in the signal samples by removing the predictable part from the signal. In this analysis the linearly weighted past samples are used to predict the sample at the current sampling instant. The linear weights or Linear Prediction coefficients (LPCs) are obtained by solving the normal equations as explained before. The LPCs determine the best all-pole filter for a given order of the model. Thus, LP analysis can be used to extract spectral peaks in the spectrum. The residual signal is obtained by subtracting the predicted value at a given instant from the actual sample value at that instant, and is also called linear prediction (LP) residual. The LP residual looks like noise, as correlation among samples is significantly reduced compared to the original signal. The autocorrelation function of the LP residual looks like an impulse at the origin (zero delay) with very small amplitudes for other lags. Figure 6 shows a signal, its LP residual and autocorrelation function. The autocorrelation of the signal is also shown for comparison.

The autocorrelation function of the LP residual enhances

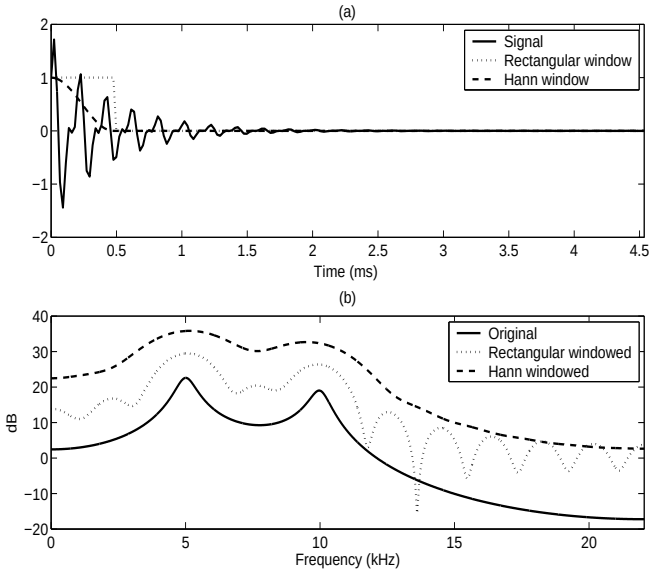

FIG. 7. Effect of windowing the response of an all-pole model.(a) The IIR of a four pole system and two windows (half Hann and rectangular, 1.5 ms) and (b) the log magnitude of the original and the windowed signals. Each of the spectra are successively shifted by 10 dB for clarity. Each pair of complex conjugate poles produce one peak in the spectrum.

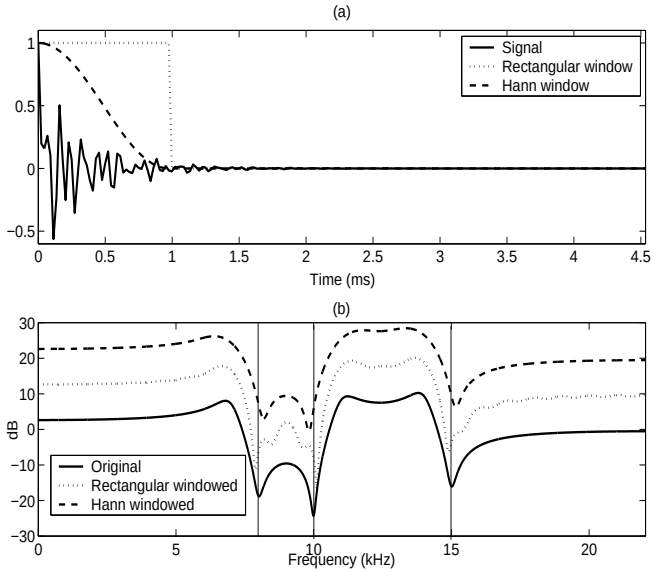

FIG. 8. Effect of windowing the response of a pole-zero model.(a) The IIR of a three pole three zero system and the two windows (half Hann and rectangular, 1.0 ms) and (b) the log magnitude of the original and the windowed signals. Each of the spectra are successively shifted by 10 dB for clarity.

the delay information if the signal has a delayed signal component. This delay is seen in the second large peak in the autocorrelation function of the LP residual. Interestingly this peak due to the delay may appear even if the delayed signal has characteristics different from the main signal component. Thus, the autocorrelation function of the LP residual can be used to determine these. However, if there is more than one delayed component, say at τ_1 and τ_2, then the autocorrelation function of the LP residual will have secondary peaks at several other instants such as $\tau_1 \pm \tau_2$, in addition to those at τ_1 and τ_2.

The instants of large error in the LP residual are also indicative of the delays, and the temporal relations of these delays are exactly the same as in the original signal. However, due to noise and also imperfect linear prediction model, there will be significant error values at many other instants, making it difficult to isolate the instants corresponding to the delays. To reduce the effects of other error samples, the Hilbert envelope of the LP residual can be used.

F. Hilbert Envelope

The Hilbert Envelope (HE) $h(n)$ of a signal $e(n)$ is computed as follows[54]:

$$h(n) = \sqrt{e(n)^2 + e_H(n)^2}, \;\; e_H(n) = IDFT[v(k)], \quad (18)$$

where $e_H(n)$ is the Hilbert transform of the signal $e(n)$, and

$$\begin{aligned} v(k) &= jE(k), \; k = 0, ..., N/2 - 1 \\ v(k) &= -jE(k), \; k = N/2, ..., N - 1 \end{aligned} \quad (19)$$

$$E(k) = DFT[e(n)], \quad (20)$$

and N is the number of DFT points. The HE of the LP residual is shown in Figure 6(d), where the large amplitude errors are more visible. Significant additional peaks are also still present in the HE. These are mostly due to noise and poor approximation of the LP analysis.

G. Windowing

Windowing of a signal is used to focus on the segment that needs to be analyzed. The size and the shape of the window can alter the features in the spectral domain significantly. When shorter time windows are used, the peaks in the spectral domain are broadened, and closely spaced peaks may merge. The spectral nulls are also affected significantly by short windows and are either modified, or lost, or new nulls are introduced due to the short window. We also note that peaks in the spectrum can significantly alter the shape of the nulls. This is because windowing in the time domain is equivalent to convolution of the spectrum of the window and the signal. Shorter windows have a broad main lobe in the spectrum, which when convolved with a spectrum with sharp peaks can produce a spectrum where the peaks are broadened and merged. Figure 7 shows the spectrum of the IIR signal of an all-pole system when the response is windowed. The use of tapered windows like the Hann window[53] may reduce the fluctuations caused by truncation of the signal due to a rectangular window, but the effects due to small size of the window still remain.

If the features that are desired to be extracted are both in the spectral peaks and nulls, then windowing affects the nulls in the resulting spectrum due to the presence of spectral peaks

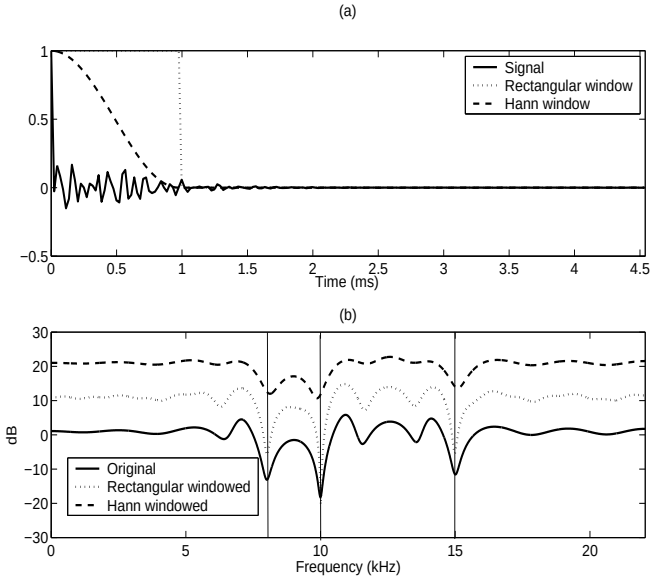

FIG. 9. Effect of windowing the LP residual of the response of the pole-zero model. (a) The 12^{th} order LP residual of the signal shown in Figure 8 and two windows (half Hann and rectangular, 1.0 ms) and (b) the log magnitude of the original and the windowed signals. Each of the spectra are successively shifted by 10 dB for clarity.

as shown in Figure 8. The effect of windowing is significantly less when an LP residual is used as shown in Figure 9. But the features of the spectrum that are retained in the model spectra depend critically on the order of the model, and also on the objective criterion used for modelling, as discussed earlier. Note that the effect of the size of window is less for an FIR signal, which is mainly due to an all-zero system, as long as the size of the window is greater than the order of the system. But in the case of noise in practical signals or when the signal is convolved with the response of the system, the desired features of the FIR system are again altered by the windowing. Windowing of the LP residual is likely to preserve many of the features of the spectral nulls, as the samples of the residual are less correlated. Windowing of the autocorrelation of the LP residual may be still better, if the temporal relations are not smeared due to delayed component signals. This is because the autocorrelation function reduces the effects of additive noise.

IV. EFFECTS OF SIGNAL PROCESSING OPERATIONS ON THE HRIR

The experimentally measured HRIRs depend on the following factors: (a) Nature of the excitation signal (the CIPIC database[1] uses Golay-code signals to measure the HRIR.), (b) position of the microphone (generally most measurements use the 'blocked meatus' configuration), (c) azimuth angle θ, (d) elevation angle φ, (e) range of the source r[69], (f) diffraction effects due to the head, (g) reflections from the torso, shoulder and knees, and (h) response of the ear canal and the pinna. The measured response is compensated for the response of the excitation signal, so that the HRIR may be assumed as the impulse response of the acoustic system from the point of excitation in the room to the point of reception at the microphone placed in the blocked ear canal. The HRIR in the CIPIC databse is 200 samples at a sampling frequency of 44.1 kHz, corresponding to 4.54 ms.

Figure 2 shows a typical HRIR (subject 10, right ear, elevation 45^o and azimuth 0^o) we consider for illustration throughout this section. This HRIR is the response of the ear due to direct and reflected components, plus noise. The log magnitude spectrum, a $(12, 12)^{th}$ order pole-zero model spectrum and a 12^{th} order all-pole model spectrum are also shown in the figure. The plots are displaced vertically to see the details of the approximation of the spectral envelope by the model spectra. As can be seen from the plots, due to the combined effects of different phenomena, it is difficult to isolate the notches due to the pinna alone. Also, in order to approximate the spectrum envelope better, the model would typically require high (> 30) orders. Even with the increased order, it is not guaranteed that the perceptually relevant notches can be captured. It is difficult to identify and extract the features in the log magnitude spectrum or in the model spectra.

We will examine how signal processing operations discussed in the previous section influence the extraction of features from the HRIR. Observing Figure 2(a) we see that the direct wave starts at around 0.8 ms. It is desirable to consider the signal from this time instant. The reason being that in the measured HRIR there is a faint pulse arriving before the main direct pulse. This behavior is likely to cause problems in analysis and hence we consider the signal from the instant of the main pulse. This instant can be found by taking the slope of the unwrapped phase spectrum or by locating the instant of the maximum amplitude in the signal and shifting back till there is a increase in the signal amplitude[70].

There are two prominent reflections one at 1.6 ms (due to the torso reflection) and another at around 3.2 ms (due to the knee reflection). To reduce the effects of the reflections on the analysis, we study two types of windows, Hann window[53] and rectangular window. The window functions and their corresponding log magnitude spectra are shown in Figure 10 for a window size of 1.5 ms. We consider the effect of widow size by considering three sizes 1.0 ms, 1.5 ms and 2.0 ms for both half Hann and rectangular window. We compute the spectrum and the group delay function in each case to see if the features of interest can be enhanced. Four cases are considered: (a) Windowing the waveform, (b) windowing the autocorrelation function of the windowed waveform, (c) windowing the LP residual, and (d) windowing the autocorrelation function of the windowed LP residual.

As mentioned earlier, windowing is used to reduce the effects of reflected components. However, windowing reduces the frequency domain resolution. The artifacts of windowing may mask or alter the frequencies of the spectral notches due to the pinna. In Figure 11 we see that windowing the waveform reduces the effect of reflection significantly. This can be seen in Figure 11(b), where the pinna notches appear prominently, and the notches due to reflection seem to be absent. The artifacts of windowing can be seen both in the log spectrum and in the group-delay function, especially for the rectangular window.

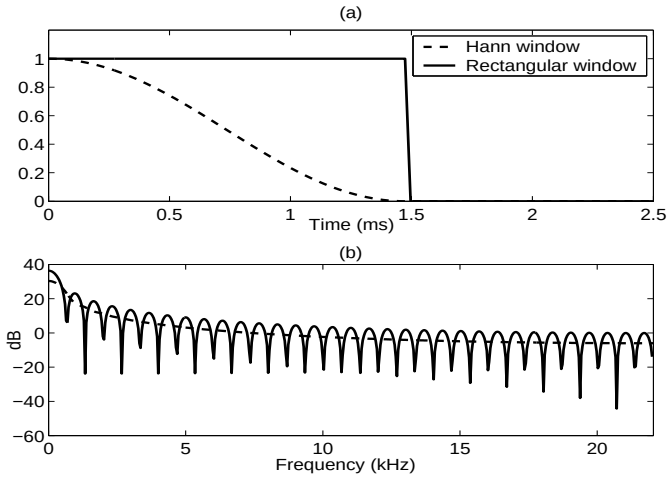

FIG. 10. Window functions used in this study. (a) Half Hann and rectangular windows of size 1.5 ms and (b)the corresponding log squared magnitude spectrum of the windows.

Autocorrelation of the windowed signal can be used to reduce the effects of truncation in the original signal. Figure 12 shows the effect of windowing the autocorrelation function for a rectangular windowed signal case. In this case, the spectral details can be seen better than in the case of direct windowing of the waveform.

Interdependence of adjacent signal samples makes the windowing less effective. Therefore to reduce this dependence, an LP residual is derived using a 12^{th} order LP analysis.[47] Effect of direct windowing of the LP residual is shown in Figure 13, where the spectral notches can be seen even better compared to the plots in Figure 11.

Effect of windowing the autocorrelation function of the windowed (rectangular) LP residual is shown in Figure 14. The influence of the spectral peaks on the notches due to windowing is significantly reduced in Figure 13 and 14. Also the effect of the size of the window is significantly less for the case of the autocorrelation function of the LP residual then for the LP residual itself.

Figure 15 shows the windowed spectrum of the HRIR for elevation -45^o and azimuth 0^o. In this case since the refection is close to the original signal, the notches caused by reflection are placed far apart in the frequency domain. As a result the original spectrum itself is quite smooth, and the notches are sharply defined. However the windowing still has an advantage in that the pinna notch which was faint in the original spectrum are highlighted now. See for example the notch between 8 and 10 kHz marked in Figure 15(b). The notch is very faint in the original spectrum, but it clearly stands out in the windowed spectrum.

Finally we examine the potential of the HE of the LP residual for determining the likely instants of the reflected sounds. Figure 16 shows the HRIR, its LP residual and the Hilbert envelope of the LP residual. The HE of the LP residual shows significant peaks at the instant of reflections. But there are many other peaks also, which may be caused by noise in the signal or poor approximation in the LP analysis. In many cases the HE of the LP residual helps determine the sizes of the windows to be used to isolate the effects of reflections. While the issue of decomposition[55] of the HRIR is not addressed in this paper, we remark that this may help decompose the HRIR into different components contributed by reflections, diffraction and pinna effects.

V. DETERMINING THE FREQUENCIES OF PINNA SPECTRAL NOTCHES

From the previous section it is obvious that windowing in the time domain helps isolate the direct component of the signal from the reflected components. The HE of the LP residual may be used to locate the first major reflection, and hence may be used to determine the window size approximately. The LP residual, rather than the original signal, is useful to reduce the effect of dependencies of the samples, thus making the windowing more effective. Autocorrelation function of the windowed LP residual helps to reduce the effects due to truncation and noise. By choosing the windowed autocorrelation function of the LP residual, one can extract the spectral components that depend on the notches caused mainly by pinna.

The features of spectral notches are highlighted in the spectra and in the group-delay functions of the windowed signals. In fact after reducing/removing the effect of reflections by windowing, some hidden spectral notches may appear prominently. Those notches were masked by the periodic notches due to reflections. The resolution of the spectral components is enhanced in the group-delay function. The presence of significant peaks and valleys (ripple effect in the spectrum and the group-delay function) is due to the windowing effect and also due to any reflection or noise component in the windowed signal. The effects of windowing can be reduced to some extent in the autocorrelation function of the windowed LP residual. The spectra exhibits smaller ripple effect, which is also reflected in the group-delay function. However still due to the high resolution property of the group delay function, there are many peaks and valleys besides the prominent valleys in the group-delay function, and it is difficult to isolate only the nulls caused by the pinna automatically.

In fact for different sizes of the windows the group-delay function does not alter significantly, at least the prominent notch regions, as shown in Figure 18. This shows that significant features of the spectrum lie within first few samples of the autocorrelation function. This is because the response is an all-zero filter (FIR). The ripple effect is reduced significantly using smaller window sizes, without affecting the notch regions. Using the zero threshold for the group delay function, all valleys below the zero value are marked as relevant notches and their frequencies are noted. In practice a slightly higher threshold of -1 was found to give good results and eliminate any spurious nulls by windowing. The algorithm for this method of extracting nulls is given in Table I and the sequence of signal processing operations is shown in Figure 17.

Another way of determining the frequencies of the notches of the spectrum is as follows. Since the spectra of the windowed LP residual or its autocorrelation function is a smooth function with nulls, the spectrum is inverted to obtain a spec-

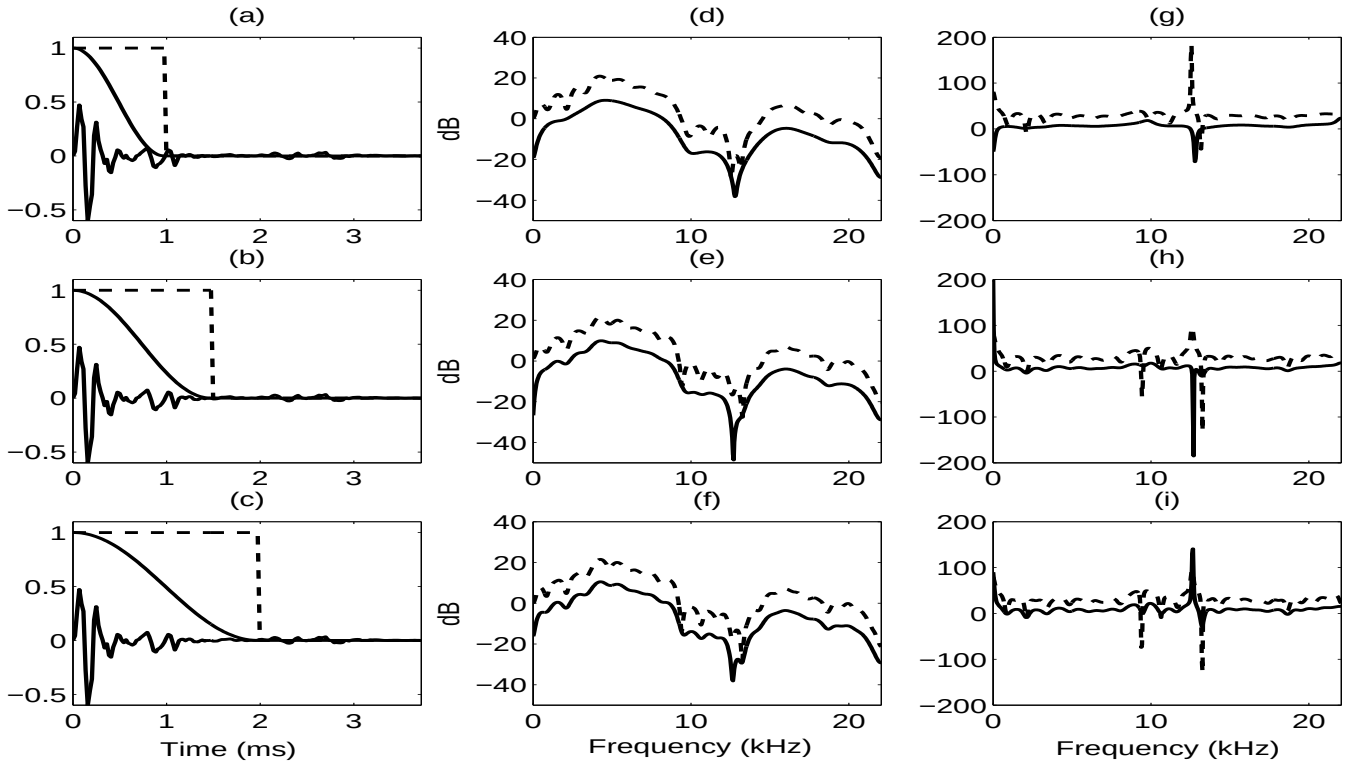

FIG. 11. Effect of windowing the HRIR. (a), (b) and (c) show the HRIR and three windows of size 1.0 ms, 1.5 ms and 2.0 ms (both Hann (solid line) and rectangular window (dotted line)), (d), (e) and (f) show the log magnitude spectra of the windowed signals (the dotted spectrum is displaced vertically by 10 dB), (g), (h) and (i) show the corresponding group-delay function (the dotted function is displaced vertically by 20). The windowing effects can be clearly seen in the group delay functions.

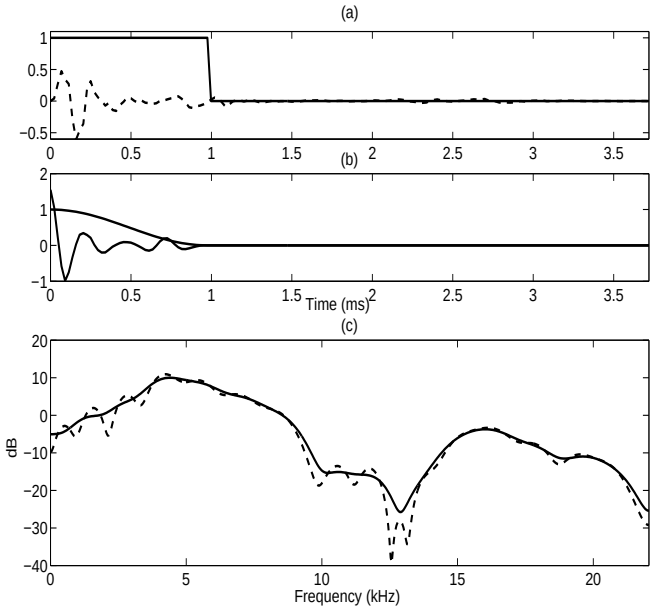

FIG. 12. Effect of windowing the autocorrelation of the windowed signal. (a) Original HRIR and the rectangular window of 1.0 ms, (b) Autocorrelation function of the windowed HRIR and a Hann window of 1.0 ms and (c) log magnitude spectrum of the windowed autocorrelation function (solid) and of the windowed signal (dotted).

TABLE I. Algorithm 1: Extracting the spectral nulls using zero thresholded group-delay function of the windowed autocorrelation function of the windowed LP residual.

- Determine the initial onset of the HRIR and use the HRIR from that instant.
- Derive the p^{th} ($p = 10 to 12$) order LP residual from the given HRIR (Figure 17(b)).
- Window the LP residual using a half Hann window of around 1.0 ms (Figure 17(c)).
- Compute the autocorrelation function of the windowed LP residual (Figure 17(d)).
- Window the autocorrelation function using a Hann window of around 1.0 ms (Figure 17(e)).
- Compute the group-delay function of the windowed autocorrelation function (Figure 17(k)).
- Zero threshold the group-delay function and locate the local minima.

trum with prominent peaks as shown in Fig. 19. An all-pole model can be fit to this spectrum by computing the autocorrelation function, and then applying the Levinson-Durbin method[47] for the first few (10) autocorrelation coefficients. The autocorrelation function is obtained by taking the inverse DFT of the inverted spectrum. The frequencies corresponding

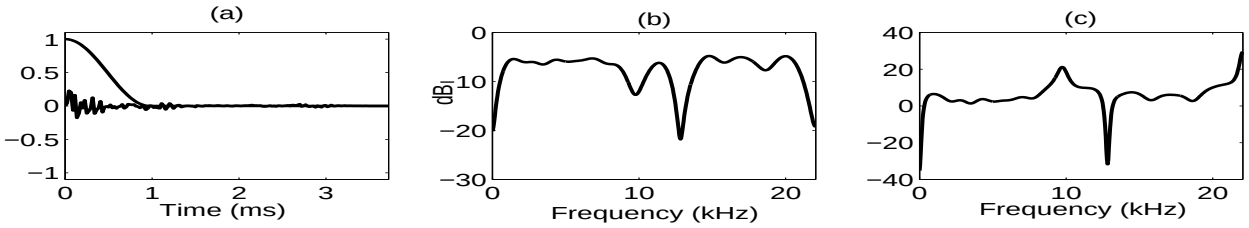

FIG. 13. Effect of windowing the LP residual of the HRIR. (a) The 12^{th} order LP residual and half Hann window of size 1.0 ms, (b) the log magnitude spectra of the windowed LP residual signal and, (c) the corresponding group-delay function.

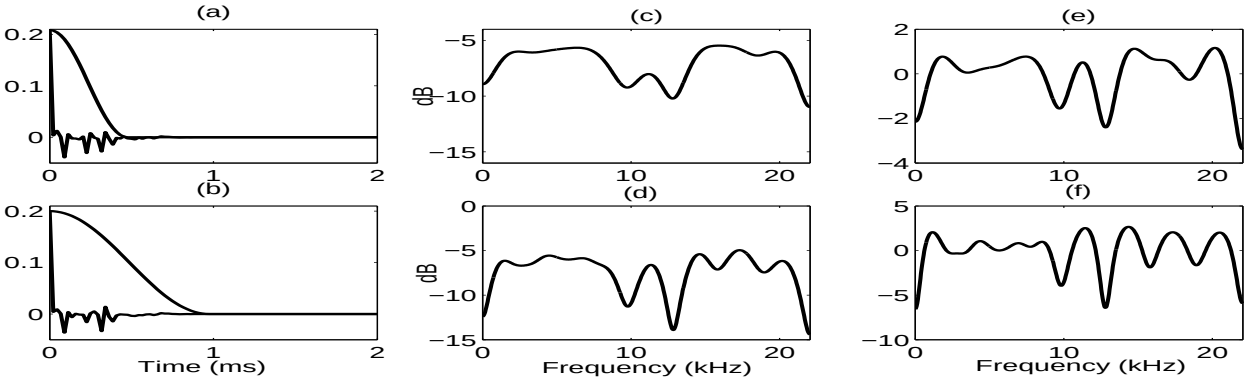

FIG. 14. Effect of windowing the autocorrelation of the windowed LP residual of the HRIR. (a) and (b) show the autocorrelation function of the 12^{th} order windowed (1.0 ms) LP residual and two windows of sizes 0.5 ms and 1.0 ms , (c) and (d) show the log magnitude spectra, (e) and (f) show the corresponding group-delay function.

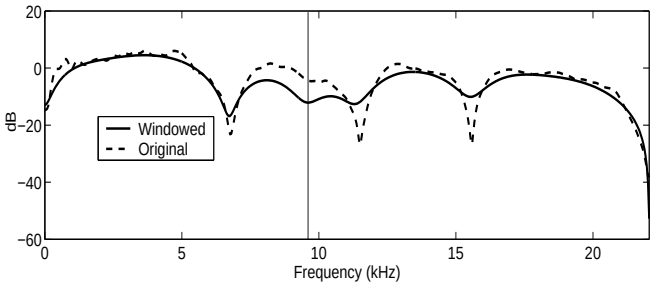

FIG. 15. The spectrum of the original signal (dashed) and the windowed signal (solid) for HRIR corresponding to elevation -45^o and azimuth 0^o and a window of 0.7 ms duration.

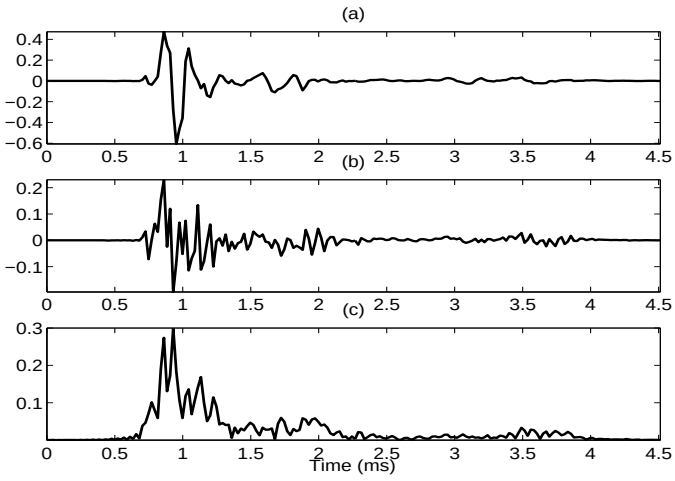

FIG. 16. (a) HRIR (b) 12^{th} order LP residual (c) Hilbert envelope of the LP residual.

to the complex roots of the all-pole model correspond to the frequencies of the prominent nulls in the spectrum of the windowed HRIR. This method of extracting the notch frequencies is summarized in the algorithm given in Table II.

The advantage of using the group-delay function (in Algorithm 1) is that we need not specify the order (number of zeros) as in Algorithm 2. Figure 20(a) shows the original log magnitude spectra for azimuth 0^o and 25 different elevations (from -45^o to 90^0). As can be seen, because of the combined effects of different phenomena, the pinna notches are not that obvious. Compare this to Figure 20(b) which shows the spectra of the windowed 12^{th} order LP residuals. The pinna notch pattern can be clearly seen. Figure 20(c) shows the group-delay functions of the windowed autocorrelation functions of the windowed LP residuals and Figure 20(d) shows the 12^{th} order all-pole model of the inverted spectrum of the windowed LP residuals.

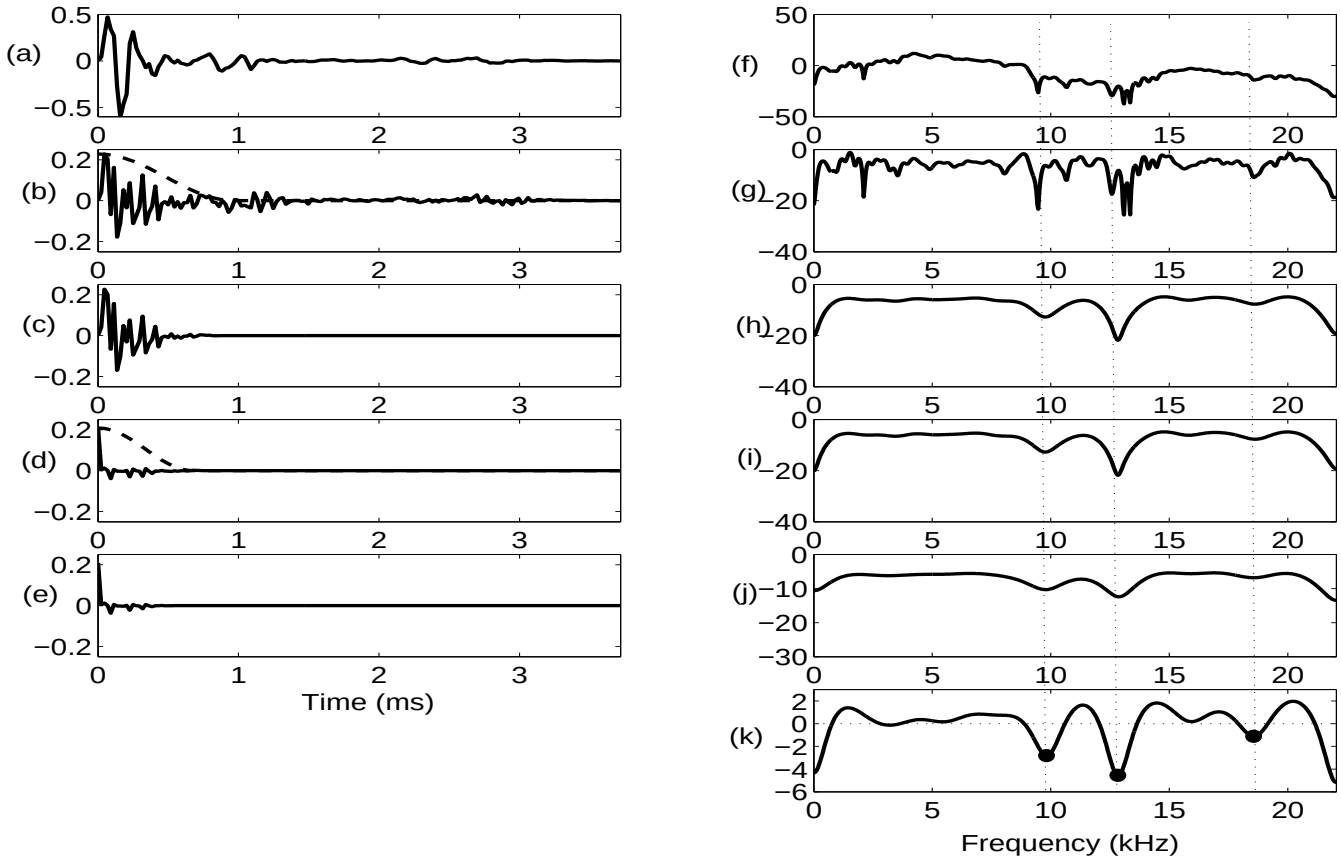

FIG. 17. Signal processing steps for extracting the pinna null frequencies (Algorithm 1). (a) Original HRIR signal, (b) 12^{th} order LP residual, (c) windowed LP residual (1.0 ms Hann window), (d) autocorrelation function of the windowed LP residual and (e) windowed autocorrelation function (0.7 ms Hann window). The plots (f), (g), (h), (i) and (j) show the log magnitude spectrum (in dB) corresponding to signals in (a), (b), (c), (d) and (e), respectively. The plot (k) shows the group-delay function of the windowed autocorrelation function. The local minima in the group-delay function (zero thresholded) are shown.

TABLE II. Algorithm 2: Extracting the spectral nulls using an all-pole model on the autocorrelation function of the inverted spectrum of the windowed LP residual.

- Determine the initial onset of the HRIR and use the HRIR from that instant.
- Derive the p^{th} ($p = 10 to 12$) order LP residual from the given HRIR (Figure 17(b)).
- Window the LP residual using a half Hann window of around 1.0 ms (Figure 17(c)).
- Compute the squared magnitude spectrum of the windowed LP residual (Figure 19(a)).
- Invert the windowed LP residual spectrum (Figure 19(b)).
- Take the IDFT of the inverted spectrum to get the corresponding autocorrelation function.
- Using Levinson-Durbin recursion[47] fit a p^{th} ($p = 10 to 12$) order all-pole model to the inverted spectrum using the first $p+1$ coeffecients of the autocorrelation function (Figure 19(c)).
- The null frequencies can abe obtained form the all-pole model by factorizing the model polynomial.

VI. SPECTRAL NOTCH FREQUENCIES FOR DIFFERENT CASES OF HRIR

The algorithms developed in the previous sections are applied for HRIRs of different subjects and for different elevations and azimuth angles. Figures 21(a)-(h) shows the spectral notch frequencies for the right ear HRTF corresponding to subject 10 in the CIPIC database. The notch frequencies are plotted as a function of elevation for different azimuths. Note that negative azimuth angles correspond to the contralateral HRTF, with the pinna in the shadow region of the head and the diffraction effects prominent. However, some pinna notches are still dominant and we were able to extract them using the same algorithm. A few notches due to head diffraction effects also appear (see Figure 21(h)). The pinna notches in the contralateral side can be explained if we assume that the sound diffracts around the head entering the contralateral concha at approximately the same elevation angle as if the source were in the ipsilateral hemisphere.[42] However, since elevation perception is essentially thought to be monaural[2] it is likely that humans use only the near ear (i.e. the ear closest to the source) for vertical localization. It is still possible that the pinna notches in the contralateral HRTF could provide extra

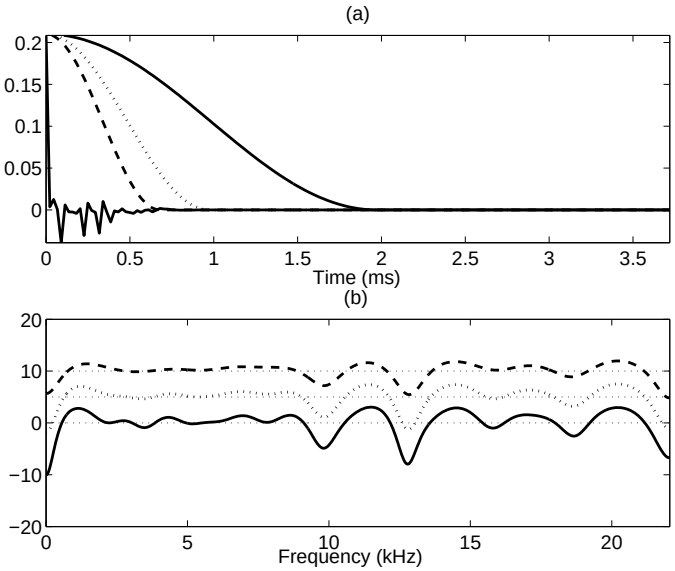

FIG. 18. (a) Autocorrelation function of the windowed LP residual and three Hann windows corresponding to 0.7, 1.0 and 2.0 ms, and (b) the corresponding group-delay functions of the windowed autocorrelation functions. The group delay-functions are shifted by a constant factor for better viewing. Note that the effect of the size of the window is less on the autocorrelation function.

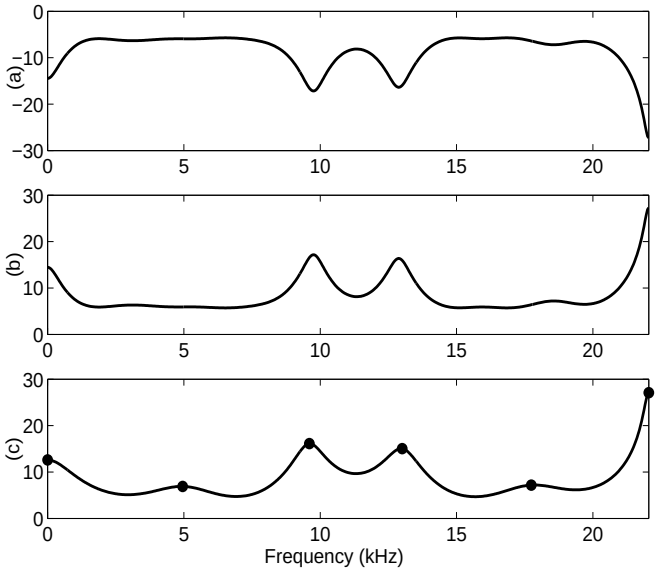

FIG. 19. (a) Log magnitude spectrum of the 12^{th} order windowed (1.0 ms Hann) LP residual, (b) the inverted spectrum and (c) a 12^{th} order all-pole fit to the inverted spectrum. The frequencies of the nulls are marked.

cues for vertical localization. Figure 21 (i) shows the same for the left pinna for subject 10 and azimuth 0^o. It was observed for most subjects that left and the right pinna do not have the same shape and dimensions.[1] The frequencies of the notches and their variation with elevation are different for the left and the right pinna. Fig. 21 (a)-(i) shows the same results for 9 different subjects and for azimuth 0^o. These results demonstrate that the proposed method is effective in determining the frequencies of the spectral notches caused by the pinna in all cases considered.

VII. DISCUSSION AND FURTHER STUDIES

We briefly discuss relating the frequencies of the pinna spectral notches to the anthropometry and shape of the pinna, and point out how these extracted pinna notch frequencies can be used for HRIR modelling, interpolation and customization. We also relate the notches with the mode patterns discussed by previous researchers. We hope this discussions and comparisons will motivate future work in modelling the pinna features from anthropometry.

A. Pinna Reflections

The structure of the pinna is fairly complicated and difficult to characterize simply. To first approximation the response can be characterized by peaks and notches observed in the spectrum. A simple model for pinna notches in terms of the time delays of reflections off the various features in the pinna (like the concha and the crus helias), was first suggested by Batteau.[40] Hebrank and Wright[11,41] attribute the pinna spectral notches to reflection of the sound from the posterior concha wall. This idea was further refined by Lopez-Poveda and Meddis[42] who incorporated diffraction in the model. Figure 23 shows the simple reflection model. The direct wave incident at an angle φ is reflected from the concha wall. If $x(t)$ is the incident wave then the total signal measured $y(t)$ is the sum of the direct and the reflected wave.

$$y(t) = x(t) + ax(t - t_d(\varphi)), \tag{21}$$

where a is the reflection coefficient and $t_d(\varphi)$ is the time delay given by

$$t_d(\varphi) = \frac{2d(\varphi)}{c}, \tag{22}$$

where $2d(\varphi)$ is the distance corresponding to the delay and c is the speed of the sound (approximately $343m/sec$). The distance $d(\varphi)$ depends on the angle φ and shape of the pinna. The delay $t_d(\varphi)$ causes periodic notches in the spectrum, whose frequencies are given by

$$f_n(\varphi) = \frac{(2n+1)}{2t_d(\varphi)} = \frac{c(2n+1)}{4d(\varphi)},\ n = 0, 1, \ldots \tag{23}$$

The frequency of the first spectral notch is given by,

$$f_0(\varphi) = \frac{c}{4d(\varphi)} \tag{24}$$

In practice there are multiple reflections occurring in the pinna. Each reflection gives rise to a spectral notch. In the previous section we extracted the frequencies of the spectral notches. From Equation 24 we can calculate the pinna distance $d(\phi)$ corresponding to the notch frequency $f_0(\phi)$. As

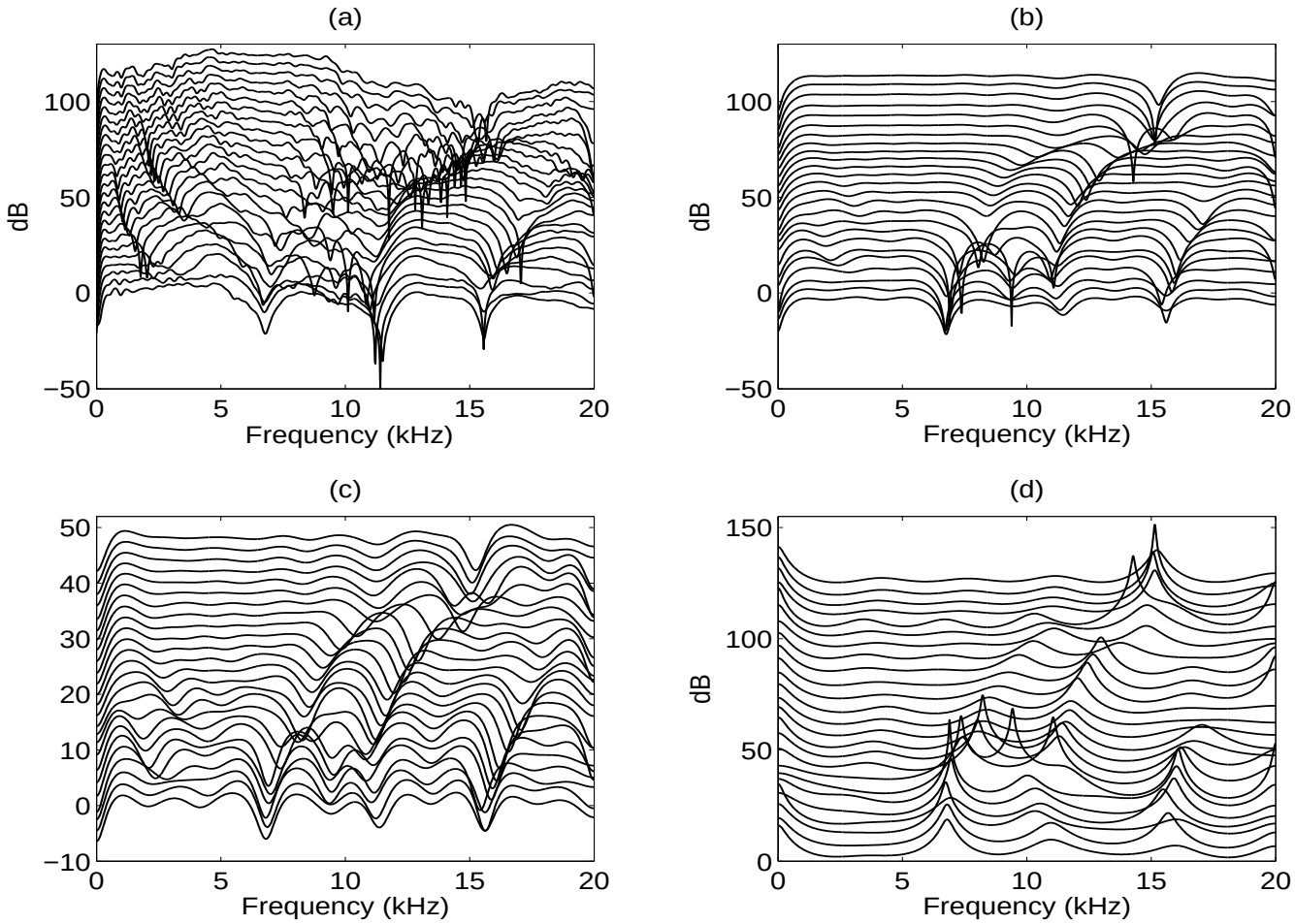

FIG. 20. (a) Log magnitude spectra for azimuth 0^o and 25 different elevations (from -45^o to 90^0.) and the corresponding (b) spectra of the windowed (1.0 ms Hann) 12^{th} order LP residuals, (c) group-delay functions of the windowed (1.0 ms Hann window) autocorrelation functions of the windowed LP residuals and (d) the 12^{th} order all-pole model of the inverted spectra of the windowed LP residuals. Each of the plots is staggered by a constant amount for better viewing.

the angle ϕ is varied, the notch frequency varies depending on the the shape of the pinna. The variation of the notch frequency reflects the shape of the pinna. The pinna images as well as the ear anthropometry are available in CIPIC[1] database. The distance can be marked on the pinna image approximately. Figure 24(a) shows the notch frequencies extracted for azimuth 0^o and elevation varying from -45^o to 90^0 (subject 10 right pinna). We consider only elevations in front of the head since, for elevations behind the ear, it is not clear what the mechanism of the spectral notches is. For each of the extracted notches the corresponding distance is plotted on the image of the pinna (Figure 24(e)) and appears consistent with this argument. It is interesting to see that the shape and dimensions of the concha are clearly seen in the extracted frequencies of the spectral nulls. The first spectral null thus appears to be caused due to reflection from the concha. As the elevation is varied it traces out the shape of the concha. The third spectral null could be due to the inner cavity in the concha caused by the crus helias dividing the concha into two. Figure 24 shows the same results for three other subjects in the CIPIC database, and exhibits the same trend. Here, we have presented results only for zero azimuth and varying elevation, and future research should complete the analysis for other different azimuths and elevations, and different subjects.

These results suggest that the shape of the different cavities in the pinna is as important as the gross dimensions. A model for the pinna should take this into consideration. Since measurement of the HRIR is a tedious process, a particulary appealing method for synthesizing the HRIR would be to take the image of an pinna and obtain the notch frequencies by analyzing the pinna anthropometry.

B. Interpolation

HRIRs are typically measured for a finite grid and some sort of interpolation is usually used. Different approaches used include direct time domain interpolation of HRIRs, direct frequency domain interpolation of HRTFs and interpolation of principal components in the PCA domain.[56] These methods do not take into account the perceptual significance of the dif-

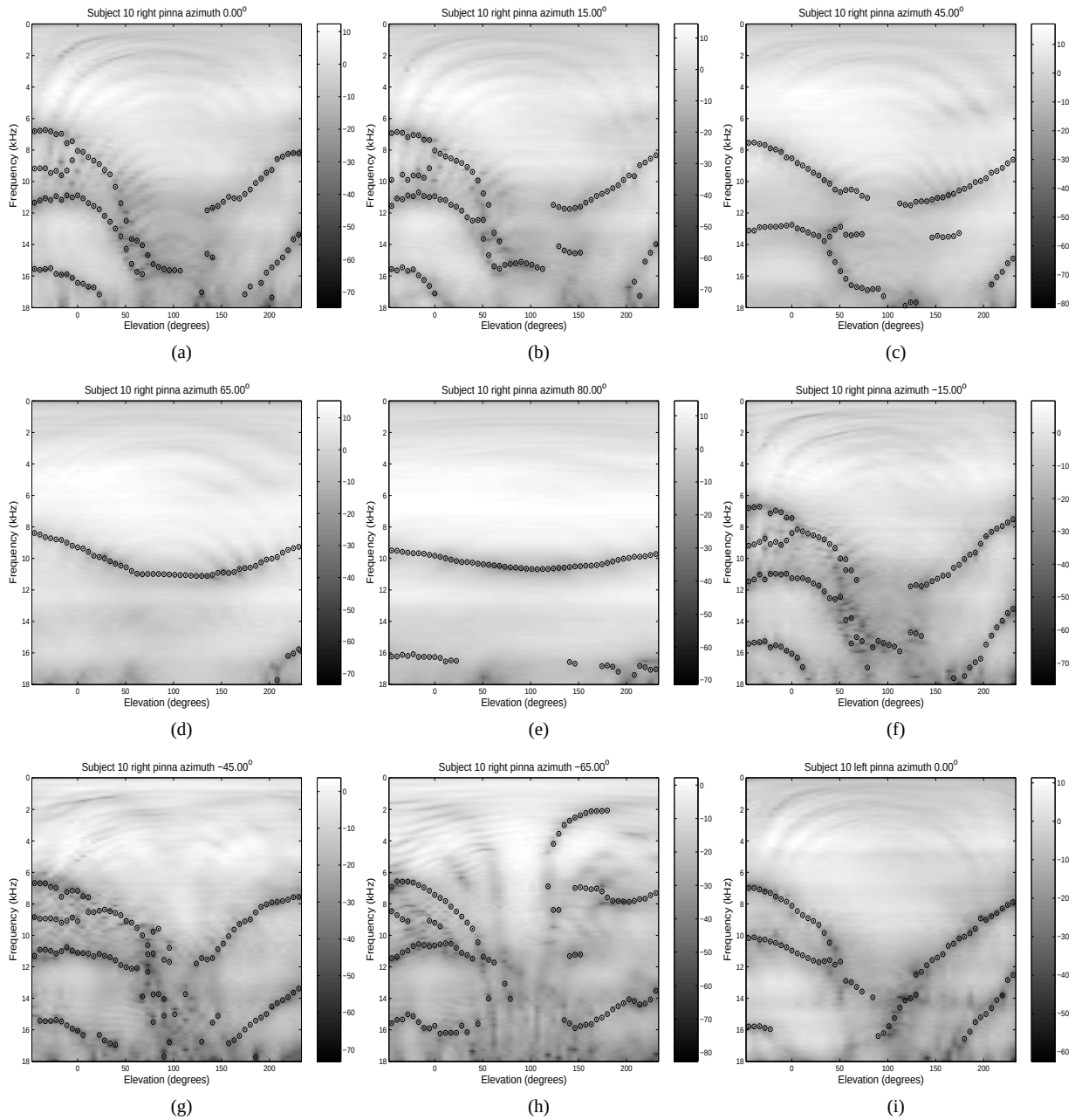

FIG. 21. The spectral notch frequencies extracted for subject 10 *right* pinna in the CIPIC database for azimuth angles (a)0^o, (b)15^o, (c)45^o, (d)65^o, (e)80^o, (f)-15^o, (g)-45^o and (h)-65^o. (i) The spectral notch frequencies corresponding to the *left* pinna of subject 10 for azimuth 0^o.

ferent features in the HRIR. If we just use linear interpolation in the frequency domain we are just averaging the magnitude of the spectrum of the closest sampling points. Figure 25 shows the HRTF for subject 10 for three elevations -45^o, -22.5^o and 0^o. The HRIR was windowed using a Hann window of 1.0 ms and the windowed spectrum is plotted. Each of the spectra is shifted by a constant amount. The dotted line shows the spectrum obtained by linearly interpolating the the spectrum corresponding to -45^o and 0^o. Note that the position of the notches does not exactly coincide with the spectrum for -22.5^o. The notches in the interpolated spectrum may correspond to those at some other elevation. One way is to interpolate the notch frequencies and then synthesize the HRTF from the notch frequencies.

C. Customization

The HRTF varies significantly between individuals due to the different sizes and shapes of different anatomical parts like the pinnae, head and torso. Applications in the creation of vir-

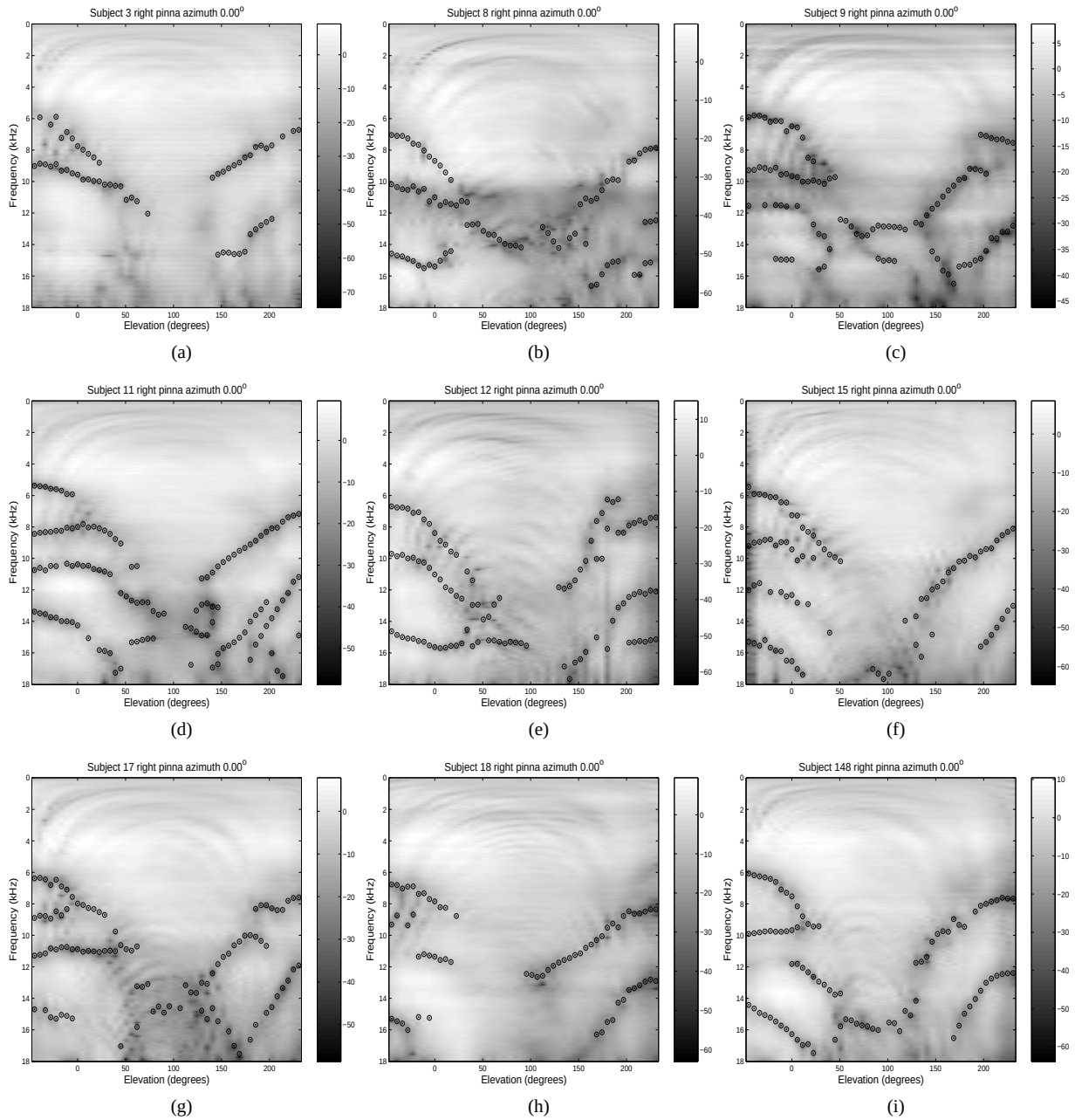

FIG. 22. The spectral notch frequencies for different elevations extracted for the right pinna for azimuth 0^o for different subjects in the CIPIC database.

tual auditory displays require individual HRTFs for perceptual fidelity. A generic HRTF would not work satisfactorily, since it has been shown that non-individual HRTFs result in poor elevation perception.[57] The usual customization method is the direct measurement of HRTFs for different subjects. Other approaches that have met with varying success include numerical modelling,[34,35] frequency scaling the non-individual HRTF to best fit the listener[58] and database matching.[59] It is possible to study the relationship between the extracted notch frequency and pinna anthropometry. Based on these observations new approaches for HRTF customization can be developed.

D. Modal Analysis

Another way of understanding the cause for the pinna spectral notches is in terms of the normal mode patterns. The pinna typically has a finite number of normal modes. Each of these modes is excited strongly only at certain frequencies and certain elevations. These modes are broadly tuned. For a given mode we can describe the mode pattern on the surface of the pinna by either measuring experimentally or numerically.

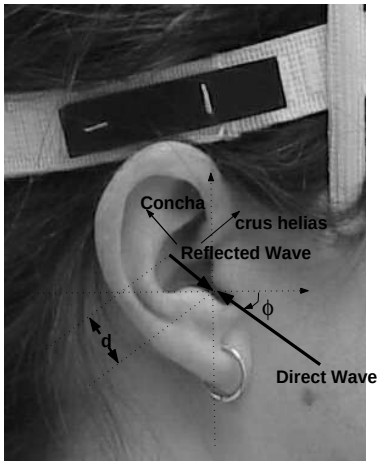

FIG. 23. A simple reflection model for the pinna spectral notches. The direct wave incident at an angle ϕ gets reflected from the concha. The time delay corresponds to a length of $2d$. The pinna image is taken from the CIPIC database.

The HRIR is this pressure measured at a particular point (the blocked meatus condition). So depending on the the particular mode, if the null pattern is formed exactly on the probe we get a drop in the pressure which are the sharp notches observed in the spectrum. When there is a resonance we get the peaks. For a given elevation and azimuth, there are a finite frequencies corresponding to the normal modes of the pinna. The overall response is obtained as a sum of these broadly tuned normal modes.

The normal modes corresponding to the peaks were first experimentally measured by Shaw[43,60,61] and later numerically verified by Kahana et. al..[35] Shaw described six modes of resonances under blocked meatus condition based on experimental measurements for 10 subjects. The six modes observed by Shaw are summarized in Figure 26.[43] The normal modes were identified by searching for the response maxima as the sound frequency and the source position were varied. The first mode is a simple quarter-wavelength depth resonance with uniform sound pressure across the base of the concha. It is strongly excited from all directions. The other modes are essentially transverse and fall into two groups: a vertical pair(modes 2 and 3) and a horizontal triplet (modes 4, 5 and 6). The poles extracted by LP analysis correspond to the resonances of the pinna reported by Shaw. Figure 27 shows the frequency response of the 12^{th} order all-pole model for the subject 10 for azimuth 0^o as a function of different elevations as a mesh plot. These six modes are marked in the plot. A similar study can be done with respect to the spectral notches. One advantage of measuring these mode patterns is that we can investigate which part of the pinna is contributing to the particular mode. For a given elevation and azimuth we can extract the nulls and measure the pressure distribution on the surface of the pinna due to a source at the corresponding direction and the particular frequency. This helps in isolating the pinna part contributing to that particular peak or a notch.

VIII. SUMMARY AND CONCLUSIONS

Automatic methods for the analysis of HRIRs and interpretation of results of the analysis in terms of the physical characteristics of the human hearing mechanism are presented. Two issues in the analysis of HRIR are: (a) the decomposition of HRIR into physical components and (b) the determination of the frequency of the spectral null caused by pinna, which determine the vertical localization. The difficulties in the analysis of HRIR due to combined effects of several components are discussed, and windowing in the time domain was proposed to reduce the effects of the reflected components. The possible artifacts in the extraction due to signal processing, especially the windowing, are addressed in detail. Finally, two methods based on the residual of a linear prediction model, windowed autocorrelation functions, group-delay function and all-pole modelling are developed. The effectiveness of the methods in isolating and determining the frequencies of the spectral nulls due to pinna has been demonstrated using the CIPIC database.

The extracted features were shown to be related to the features previous authors have shown to be important in elevation perception. The relation between the extracted features and the physical dimensions of the pinna can now be established. Some preliminary results were obtained. Once such relations are available, it should be possible to synthesize a customized individualized customized HRIR, that incorporate the relevant features of the individuals pinna in the response.

IX. ACKNOWLEDGEMENTS

The support of NSF award ITR-0086075 is gratefully acknowledged.

[1]V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," Proc.2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY,99–102 (2001).

[2]J. C. Middlebrooks and D. M. Green, "Sound localization by human listeners," Annual Review of Psychology **42**, 135–159 (1991).

[3]W. M. Hartmann, "How we localize sound," Physics Today,24–29 (1999).

[4]J. Blauert, *Spatial Hearing (Revised Edition)* (MIT Press, Cambridge, MA, 1997).

[5]J. W. Strutt (Lord Rayleigh), "On our perception of sound direction," Phil. Mag. **13**, 214–232 (1907).

[6]G.F.Kuhn, "Model for interaural time differences in the azimuthal plane," Journal of the Acoustical Society of America **62**, 157–167 (1977).

[7]F. L. Wightman and D. J. Kistler, "Factors affecting the relative salience of sound localization cues," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and A. T.B. (Lawrence Erlbaum Associates, Mahwah, NJ, 1997).

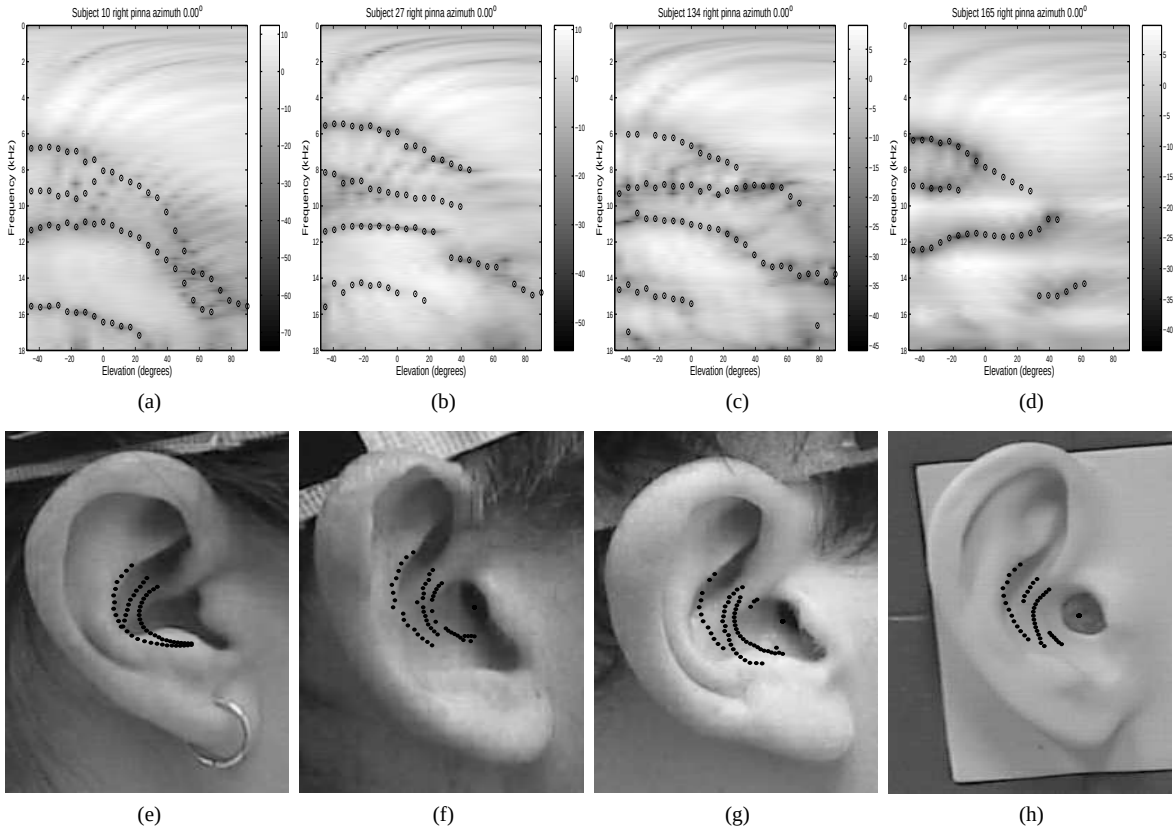

FIG. 24. The spectral notch frequencies for different elevations (from -45^o to 90^o) extracted for the right pinna for azimuth 0^o (a) subject 10, (b) subject 27, (c) subject 134 and (d) subject 165 in the CIPIC database. The dimensions corresponding to the spectral notches marked on the pinna image for (e) subject 10, (f) subject 27, (g) subject 134, and, (h) subject 165 respectively. The pinna images is taken from the CIPIC database.

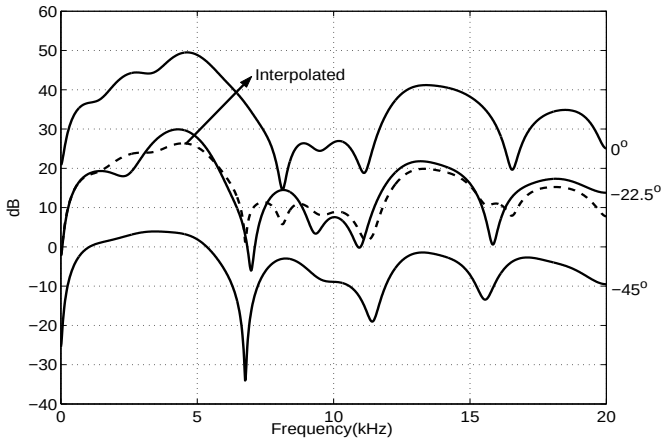

FIG. 25. Windowed spectrum (1.0 ms) for subject 10 for three elevations -45^o, -22.5^o and 0^o. Each of the spectra is shifted by a constant amount. The dotted line shows the spectrum obtained by linearly interpolating the the spectrum corresponding to -45^o and 0^o.

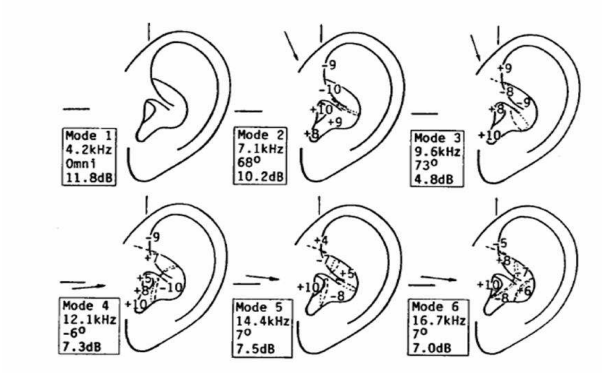

FIG. 26. The six normal modes identified by Shaw. The mode number, resonance frequencies, direction of maximum response and the response level are shown. The values are based on average value of 10 subjects. (Figure reproduced from Shaw's paper.)

[8]B. C. J. Moore, S. R. Oldfield, and G. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 khz," Journal of the Acoustical Society of America **85**(2), 820–836 (1989).

[9]D. Wright, J. H. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," Journal of the Acoustical Society of America **56**(3), 957–962 (1974).

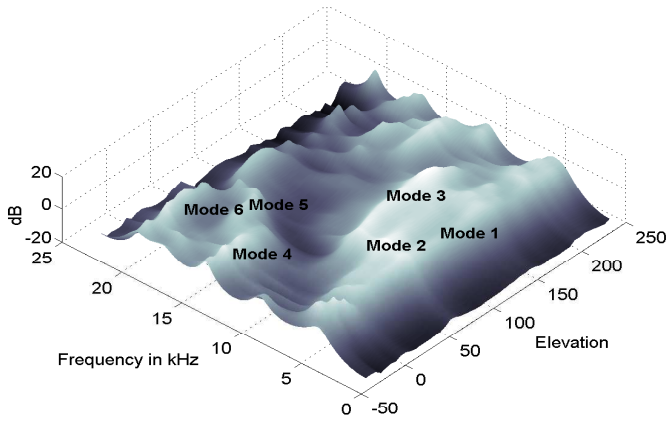

FIG. 27. Frequency response of the 12^{th} order all pole model for azimuth 0 as a function of different elevations. The six modes are approximately marked.

[10]P. Hofman, J. Van Riswick, and A. Van Opstal, "Relearning sound localization with new ears," Nature Neuroscience **1**, 417–421 (1998).

[11]J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?," Journal of the Acoustical Society of America **56**(3), 935–938 (1974).

[12]M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: effect of pinna cavity occlusion," Journal of the Acoustical Society of America **53**(2), 400–408 (1974).

[13]P. Poon and J. F. Brugge, "Sensitivity of auditory nerve fibers to spectral notches," Journal of Neurophysiology **70**(2), 655–666 (1993).

[14]P. Poon and J. F. Brugge, "Virtual-space receptive fields of single auditory nerve fibers," Journal of Neurophysiology **70**(2), 667–676 (1993).

[15]D. J. Tollin and T. C. T. Yin, "Spectral cues explain illusory elevation effects with stereo sounds in cats," Journal of Neurophysiology **90**(1), 525–530 (2003).

[16]E. Young, G. Spirou, J. Rice, and H. Voigt, "Neural organization and responses to complex stimuli in the dorsal cochlear nucleus," Phil. Trans. R. Soc. Lond. **336 (series B)**, 407–413 (1967).

[17]V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," Journal of the Acoustical Society of America **109**(3), 1110–1122 (2001).

[18]V. R. Algazi, R. O. Duda, R. P. Morrison, and D. M. Thompson, "Structural composition and decomposition of HRTF's," IEEE ASSP Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY, (2001).

[19]J. W. Strutt (Lord Rayleigh), "On the acoustic shadow of a sphere," Philos. Trans. R. Soc London **203**(Series. A), 87–110 (1904).

[20]R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," Journal of the Acoustical Society of America **104**(5), 3048–3058 (1998).

[21]V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," Journal of the Acoustical Society of America **112**(5), 2053–2064 (2002).

[22]V. Algazi, R. Duda, and D. Thompson, "The use of head-and-torso models for improved spatial sound synthesis," AES113 convention,Los Angeles,CA, USA, (2002).

[23]A. Kulkarni and H. S. Colburn, "Infinite-impulse-response models of the head-related transfer function," Journal of the Acoustical Society of America **115**(4), 1714–1728 (2004).

[24]M. A. Blommer and G. H. Wakefield, "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," IEEE Trans. Speech Audio Processing **5**, (1997).

[25]E. A. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: Pole-zero approximations of HRTFs," IEEE Trans. Speech Audio Processing **10**(1), 18–27 (2002).

[26]Y. Haneda, S. Makino, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and zero modeling of head-related transfer functions," IEEE Trans. Speech Audio Processing **7**(2), 188–196 (1999).

[27]F. Asano, Y. Suzuki, and T. Sone, "Role of spectrum cues in median plane localization," Journal of the Acoustical Society of America **88**(1), 159–168 (1990).

[28]http://www.ircam.fr/equipes/salles/listen/index.html.

[29]K. Gardner, B. amd Martin, "HRTF measurements of a KEMAR dummy-head microphone," Technical Report No. 280, MIT MEdia Lab Perceptual Computing (unpublished).

[30]C. Avendano, R. O. Duda, and V. R. Algazi, "Modeling the contralateral HRTF," AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finland,313–318 (1999).

[31]C. Avendano, V. R. Algazi, and R. O. Duda, "A head-and-torso model for low-frequency binaural elevation effects," Proc.1999 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, Mohonk Mountain House, New Paltz, NY,17–20 (1999).

[32]C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," IEEE Trans. Speech and Audio Processing **6**(5), 476–488 (1998).

[33]V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," Journal of the Acoustical Society of America **112**(5), 2053–2064 (2002).

[34]Y. Kahana, P. A. Nelson, M. Petyt, and S. Choi, "Numerical modelling of the transfer functions of a dummy-head and of the external ear," AES 16th Int. Conf. Spatial Sound Reproduction, (1999).

[35]Y. Kahana and P. A. Nelson, "Spatial acoustic mode shapes of the human pinna," Audio Engineering Society, 109th convention, (2000).

[36]J. C. Middlebrooks, "Hearing phantoms," Journal of Neurophysi-

ology **90**(1), 1–1 (2003).

[37]R. Woodworth and G. Schlosberg, *Experimental Psychology* (Holt, Rinehard and Winston,NY, ADDRESS, 1962), pp. 349–361.

[38]R. O. Duda, C. Avendano, and V. R. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," Proc. IEEE International Conference on Acoustics Speech and Signal Processing **II**, 965–968 (1999).

[39]http://www.etymotic.com/pro/er7c.asp.

[40]D. W. Batteau, "The role of the pinna in human localization," Proc. Royal Society London **168 (series B)**, 158–180 (1967).

[41]J. Hebrank and D. Wright, "Spectral cues used in the location of sound sources on the median plane," Journal of the Acoustical Society of America **56**(6), 1829–1834 (1974).

[42]E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," Journal of the Acoustical Society of America **100**(5), 3248–3259 (1996).

[43]E. A. G. Shaw, "Acoustical features of the human ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. H. Gilkey and A. T.B. (Lawrence Erlbaum Associates, Mahwah, NJ, 1997), pp. 25–47.

[44]A. Musicant and R. Butler, "The influence of pinnae-based spectral cues on sound localization," Journal of the Acoustical Society of America **75**(4), 1195–1200 (1984).

[45]F. L. Wightman and D. J. Kistler, "Of vulcan ears, human ears and 'earprints'," Nature Neuroscience **1**, 337–339 (1998).

[46]S. L. Marple Jr, "A tutorial overview of modern spectral estimation," International Conference on Acoustics, Speech, and Signal Processing **4**, 2152–2157 (1989).

[47]J. Makhoul, "Linear prediction: A tutorial review," Proceedings of the IEEE **63**(4), 561–580 (1975).

[48]K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," IEEE Transactions on Automatic Control **10**, 461–464 (1965).

[49]B. Yegnanarayana, "Speech analysis by pole-zero decomposition of short-time spectra," Signal Processing **3**(1), 5–17 (1981).

[50]B. Yegnanarayana, "Design of ARMA digital filters by pole-zero decomposition," IEEE Trans. Acoustics, Speech and Signal Processing **29**(3), 433–439 (1981).

[51]B. Yegnanarayana, "Formant extraction from linear prediction phase spectra," Journal of the Acoustical Society of America **63**(5), 1638–1640 (1978).

[52]B. Yegnanarayana, D. K. Saikia, and T. R. Krishnan, "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," IEEE Trans. Acoustics, Speech and Signal Processing **ASSP-32**(3), 610–623 (1984).

[53]A. Oppenheim and S. R. W., *Discrete-Time Signal Processing* (Prentice Hall, Englewood Cliffs, NJ, 1989).

[54]T. V. Ananthapadmanabha and B. Yegnanarayana, "Epoch extraction of voiced speech," IEEE Trans. Acoustics, Speech and Signal Processing **ASSP-23**(6), 562–570 (1975).

[55]V. C. Raykar, B. Yegnanarayana, R. Duraiswami, and L. Davis, "Extracting significant features from the HRTF," Proceedings of the 2003 International Conference on Auditory Display,115–117 (2003).

[56]T. V. Sreenivas, V. C. Raykar, and R. Raman, "Head related impulse response interpolation for dynamic spatialization," Texas Instruments DSPS Fest-2k, (2000).

[57]E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," Journal of the Acoustical Society of America **94**(1), 111–123 (1993).

[58]J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear functions in frequency," Journal of the Acoustical Society of America **106(3)**, 1493–1509 (1999).

[59]D. N. Zotkin, R. Duraiswami, L. Davis, A. Mohan, and V. C. Raykar, "Virtual audio system customization using visual matching of ear parameters," Proceedings of the 2002 International Conference on Pattern Recognition,1003–1006 (2002).

[60]E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," Journal of the Acoustical Society of America **44**(1), 240–249 (1968).

[61]E. A. G. Shaw, "Transformation of sound pressure level from the free field to the eardrum in the horizontal plane," Journal of the Acoustical Society of America **56**(6), 1848–1861 (1974).

[62]B. G. Shinn-Cunningham, S. G. Santarelli, and N. Kopco, "Tori of confusion: Binaural cues for sources within reach of a listener," Journal of the Acoustical Society of America **107**(3), 1627–1636 (2000).

[63]Same ILD cues leads to the *tori of confusion*[62]

[64]The listener is assumed to be in an anechoic chamber. The surrounding environment also has a major effect on the sound field and these must also be taken into account to render proper spatial audio. However room modelling can be done separately and can be combined with the HRTF to produce the desired effect.

[65]The spectral features due to the pinna are dominant only in the high frequency range. Some studies show that there are features in the low frequency range which seem to provide some cues for vertical localization.[17] These cues are attributed to the contributions of the head diffraction and torso reflections.[17]

[66]A small part of this work has appeared as a conference paper.[55]

[67]This method of visualization was first used by R. O. Duda and his colleagues.[17]

[68]Prof. Richard O. Duda, personal communication

[69]The range dependence is not taken into consideration. Even though the range of the sound source affects the HRIR for all practical purposes this dependence can be ignored. Most of the HRIR measurements are done with the source at a fixed distance from the center of the head.

[70]The onset times are also available in the CIPIC database itself.