Extracting the 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 Advanced Computer Studies, Department of Computer Science, University of Maryland, College Park, Maryland 20742

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

(Received 13 July 2004; revised 6 April 2005; accepted 6 April 2005)

The head related impulse response (HRIR) characterizes the auditory cues created by scattering of sound off a person's anatomy. The experimentally 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 pinna. Structural models (Algazi *et al.*, 2002; Brown and Duda, 1998) seek to establish direct relationships between the features in the HRIR and the anatomy. While there is evidence that particular features in the HRIR can be explained by anthropometry, the creation of such models from experimental data is hampered by the fact that the extraction of the features in the HRIR is not automatic. One of the prominent features observed in the HRIR, and one that has been shown to be important for elevation perception, are the deep spectral notches attributed to the pinna. In this paper we propose a method to robustly extract the frequencies of the pinna spectral notches from the measured HRIR, distinguishing them from other confounding features. The method also extracts the resonances described by Shaw (1997). The techniques are applied to the publicly available CIPIC HRIR database (Algazi *et al.*, 2001c). The extracted notch frequencies are related to the physical dimensions and shape of the pinna. © *2005 Acoustical Society of America.* [DOI: 10.1121/1.1923368]

PACS number(s): 43.66.Qp, 43.64.Ha, 43.66.Pn [AK] Pages: 364–374

I. INTRODUCTION

Humans have an amazing ability to determine the elevation and azimuth of the sound source relative to them (Blauert, 1996; Middlebrooks and Green, 1991). 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 (Blauert, 1996; Kuhn, 1977; Strutt, 1907; Wightman and Kistler, 1997). 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, or the *torus of confusion* which have the same ILD cues (Shinn-Cunningham *et al.*, 2000). Additional cues are provided by the distinctive location specific features in the received sound arising due to interactions with the torso, head, and pinna. This filtering process can be described using 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. The corresponding impulse response is called the head related impulse response (HRIR).

The HRTF varies significantly between different individuals due to differences in the sizes and shapes of different anatomical parts like the pinna, head, and torso. Applications in the creation of virtual auditory displays require individual HRTFs for perceptual fidelity. A generic HRTF would not work satisfactorily since it has been shown that nonindividual HRTF results in poor elevation perception (Wenzel *et al.*, 1993). The usual customization method is the direct measurement of HRTFs, which is a time-consuming and laborious process. Other approaches that have met with varying success include numerical modeling (Kahana *et al.*, 1999), frequency scaling the nonindividual HRTF to best fit the listeners one (Middlebrooks, 1999), and database matching (Zotkin *et al.*, 2002).

A promising approach for HRTF customization is based on building structural models (Algazi *et al.*, 2002; Brown and Duda, 1998; Raykar *et al.*, 2003) for the HRTF. Different anatomical parts contribute to different temporal and spectral features in the HRIR and HRTF, respectively. Structural models aim to study the relationship between the features and anthropometry. While good geometrical models (Algazi *et al.*, 2002; Duda and Martens, 1998) 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.

The prominent features contributed by the pinna are the

[a]Electronic mail: vikas@umiacs.umd.edu
[b]Electronic mail: ramani@umiacs.umd.edu
[c]Electronic mail: yegna@cs.iitm.ernet.in. This work was performed when the author was visiting the University of Maryland, College Park.

 0001-4966/2005/118(1)/364/11/$22.50

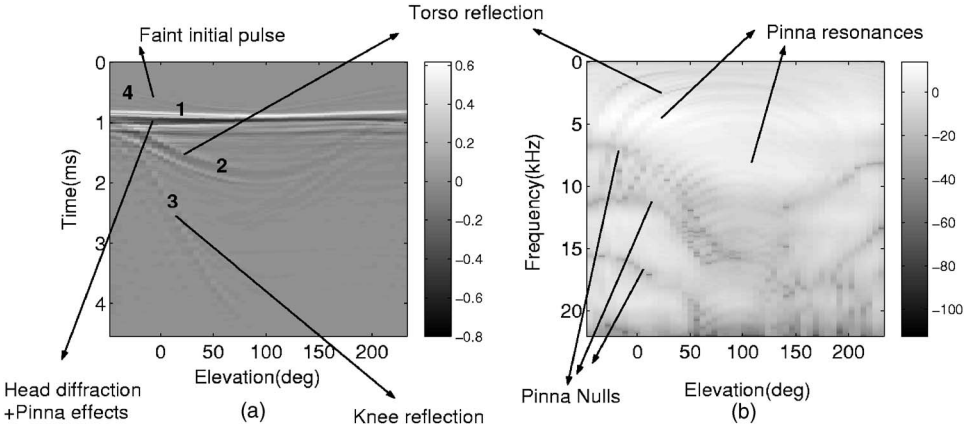

FIG. 1. (a) HRIR and (b) HRTF displayed as images for the right ear for subject 10 in the CIPIC database for azimuth angle $\theta=0°$ for all elevations varying from −45° to +230.625°. The different features are marked in both the HRIR and the HRTF plots. In (a) the gray scale value represents the amplitude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

sharp notches in the spectrum, commonly called the *pinna spectral notches*. There is substantial psychoacoustical (Moore *et al.*, 1989; Wright *et al.*, 1974), behavioral (Gardner and Gardner, 1974; Hebrank and Wright, 1974a; Hofman *et al.*, 1998), and neurophysiological (Poon and Brugge, 1993a,b; Tollin and Yin, 2003) evidence to support the hypothesis that the pinna spectral notches are important cues for vertical localization, i.e., determining the elevation of the source.

One difficulty in developing structural models for the pinna is that it is difficult to automatically extract these frequencies from measured data. Once we have quantitative values for the frequencies of the spectral peaks and notches, a model could be built relating them to the shape and the anthropometry of the pinna. Based on these, new approaches for HRTF customization using these features could be developed, and the role of the pinna in spatial localization better understood. Various psychoacoustical and neurophysiological experiments which explore the significance of the pinna spectral notches can benefit from a procedure that automatically extracts the pinna spectral notches from the measured impulse responses.

The focus of the work presented in this paper is to automatically extract the frequencies corresponding to the spectral notches. A major difficulty is that the experimentally measured HRIR includes the combined effects of the head diffraction and shoulder, torso, and, as an artifact, the knee reflection. Robust signal processing techniques need to be developed to extract the frequencies of the spectral notches due to the pinna alone, in the presence of these confounding features. The methods proposed are based on the residual of a linear prediction model, windowed autocorrelation functions, group-delay function, and all-pole modeling, guided by our prior knowledge of the physics of the problem. Our method also extracts the normal modes first described by Shaw (1997).

Several studies were made to approximate HRTFs by pole-zero models (Asano *et al.*, 1990; Blommer and Wakefield, 1997; Durant and Wakefield, 2002; Haneda *et al.*, 1999; Kulkarni and Colburn, 2004). These studies fit a pole-zero model based on a suitable error measure. However since the spectral notches extracted by them are caused due to various phenomena it is not obvious which of the spectral notches extracted by the model are due to the pinna.

II. STRUCTURAL COMPOSITION OF THE HRIR

Following the work of Algazi *et al.* (2001b) we illustrate the potential of explaining and eventually synthesizing HRTFs from the anthropometry. To this end we will consider measured HRIRs from the CIPIC database (Algazi *et al.*, 2001c). 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° to 80° and the elevation from −45° to +231° in a head-centered interaural polar coordinate system. For any given azimuth, we form a two-dimensional array, where each column is the HRIR or the HRTF for a given elevation, and the entire array is displayed as an image. This method of visualization helps to identify variation of different features with elevation (Algazi *et al.*, 2001a). Figure 1 shows the HRIR and HRTF images (for all elevations) corresponding to azimuth 0° for the right ear for subject 10 in the CIPIC database. In Fig. 1(a) the gray scale value represents amplitude of the HRIR, and in Fig. 1(b) it is the magnitude of the HRTF in dB. The different features corresponding to different structural components are also marked by hand. Figure 2 shows the HRIR as a mesh plot so that the features can be clearly seen.

Composition of the responses in terms of head diffraction, 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 features marked in Fig. 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 (Algazi *et al.*, 2001b).

Three distinct ridges, which are marked as 1, 2, and 3, can be seen in the HRIR image plot [Fig. 1(a) and Fig. 2]. The first distinct ridge is due to the direct sound wave that reaches the pinna. We see that immediately after the direct wave, activity is seen in the close vicinity (within 1.2 ms), 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 (Duda and Martens, 1998; Strutt, 1907).

The second valley shaped ridge between 1 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

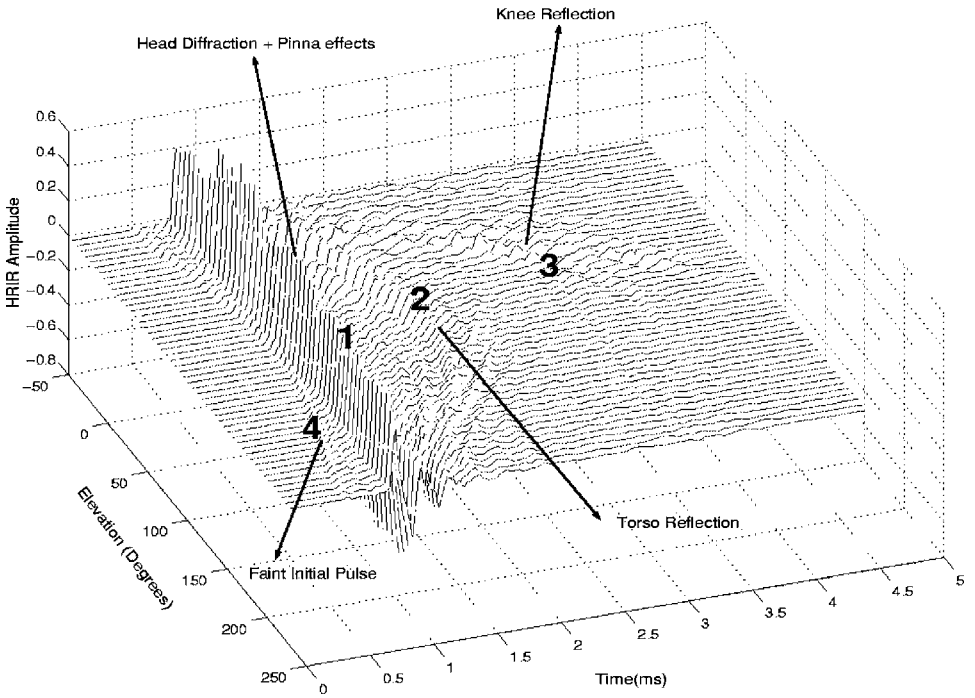

FIG. 2. The HRIR shown as a mesh plot for the right ear for subject 10 in the CIPIC database for azimuth angle $\theta=0°$ for all elevations varying from −45° to +230.625°. The faint initial pulse, the torso reflection, and the knee reflection can be clearly seen in this plot.

the torso is maximum above the head, and decreases on either side. This can be explained using simple ellipsoidal models for the head and torso (Algazi *et al.*, 2002). In the frequency domain the effect of this delay is the arch shaped comb-filter notches that can be seen throughout the spectrum [see Fig. 1(b)]. 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 (Algazi *et al.*, 2001a).

The activity seen after 2 ms is due to knee reflections, since these measurements were done with the subjects seated (Algazi *et al.*, 2001c). This is confirmed by the observation that similar activity is not seen in the back ($\varphi>90°$) and absent in the KEMAR. The other artifact is the faint pulse [marked 4 in Figs. 1(a) and 2] seen arriving before the main pulse. This is probably due to the nature of the probe microphone used in the measurements (Algazi *et al.*, 2001c). The probe microphone has a 76 mm silicone probe tube which conducts the acoustic wave to the microphone. It is likely that the signal first hits the microphone outside before reaching the probe.

The other prominent features in the frequency domain, but difficult to see in the time domain, are the prominent notches above 5 kHz. Three prominent notches can be seen in Fig. 1(b) for elevations from −45° to 90°. 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 (Algazi *et al.*, 2001b), confirm that these notches are caused due to the pinna. Also present in the response are the resonances due to the pinna [the bright patches in the HRTF image in Fig. 1(b)]. The resonances correspond to the six normal modes which were experimentally measured by Shaw (1997) and numerically verified by Kahana *et al.* (1999).

Batteau (1967) suggested 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 (1974a,b) 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 (1996) who incorporated diffraction in the model.

Previous studies done both on humans and animals that discuss the pinna features can be classified as: psychoacoustical (Langendijk and Bronkhorst, 2002; Moore *et al.*, 1989; Wright *et al.*, 1974), behavioral (Gardner and Gardner, 1974; Hebrank and Wright, 1974a; Hofman *et al.*, 1998) and neurophysiological (Poon and Brugge, 1993a,b; Tollin and Yin, 2003). Psychoacoustical experiments have demonstrated that high frequencies are necessary for localization in the vertical plane (Gardner and Gardner, 1974; Hebrank and Wright, 1974b; Musicant and Butler, 1984). By progressively occluding the pinna cavities, it was shown that localization ability decreases with increasing occlusion (Gardner and Gardner, 1974). Hofman *et al.* (1998) 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.

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°. Wright *et al.* (1974) present experiments to determine whether delays caused due to pinna reflections are detectable by humans. The results show

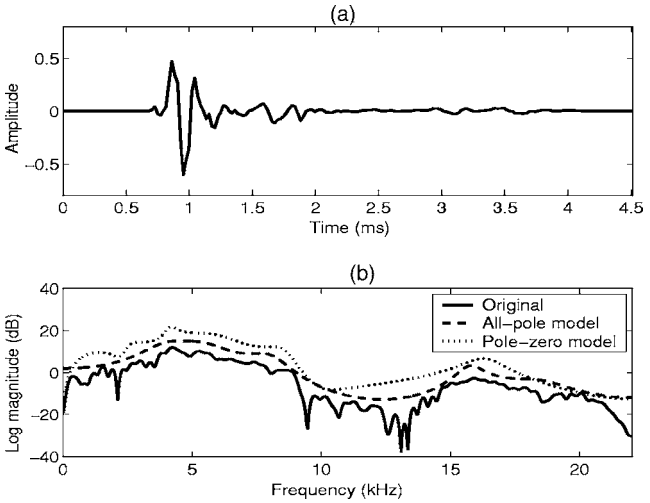

FIG. 3. (a) A typical HRIR for an elevation of 45° and an azimuth of 0°, (b) the log magnitude spectrum, a (12, 12)th order pole-zero model spectrum and a 12th order all-pole model spectrum. In the plots the all-pole spectrum and the pole-zero spectrum are displaced vertically by 5 and 10 dB, respectively, for clarity.

that delay times of 20 μs are easily recognizable when the level of the delayed signal is at least −3.5 dB with respect to the leading signal. Experiments by Moore *et al.* (1989) show that changes in the center frequency of the notches are detectable even for rather narrow notches. 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 (Poon and Brugge, 1993a). A vertical illusion that was observed in cats by Tollin and Yin (2003) can be explained well by a model that attributes vertical localization to recognition of the spectral shape cues.

Thus many studies have clearly established the importance of the pinna spectral notches in the ability to localize sounds. However we must reiterate that these studies are not able to relate the location of the notch to the pinna anthropometry, something that may be of importance in applications that seek to create personalized HRTFs without the measurements.

While previous studies address the issue of how the HRTF is *composed*, there is no attempt to *decompose* the measured HRTF of a real subject into different components. Structural models aim to decompose the HRIR into different components and then build a model for each component.

III. EXTRACTING THE FREQUENCIES OF PINNA SPECTRAL NOTCHES

One obvious way to extract the spectral notches and peaks is through pole-zero modeling (Makhoul, 1975; Steiglitz and McBride, 1965). Several studies were made to approximate the HRTFs by pole-zero models (Asano *et al.*, 1990; Blommer and Wakefield, 1997; Durant and Wakefield, 2002; Haneda *et al.*, 1999; Kulkarni and Colburn, 2004). Figure 3 shows a typical HRIR (subject 10, right ear, elevation 45° and azimuth 0°) we consider for illustration throughout this section. The HRIR is 200 samples long at a sampling frequency of 44.1 kHz, corresponding to 4.54 ms. The log magnitude spectrum, a (12, 12)th order pole-zero model spectrum and a 12th order all-pole model spectrum are also shown in the figure. 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 need to be of high order (>30). Even with the increased order, it is not guaranteed that the relevant notches can be captured. 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 modeling techniques are unlikely to bring out the specific features one is looking for in the HRIR signal. Our proposed methods do not rely on any models. We apply several signal processing techniques, including windowing, linear prediction residual analysis, group-delay function, and autocorrelation. We will motivate these in the following discussions and present the complete algorithm at the end.

In the measured HRIR there is a very faint pulse arriving before the main direct pulse, due to the nature of the measurement setup. This behavior is likely to cause problems in analysis and hence we consider the signal from the instant of the main pulse [around 0.8 ms in Fig. 3(a)]. This instant is 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 until there is an increase in the signal amplitude.

The spectral notches are caused due to multiple reflections from different parts like the head, torso, knees, and pinna cavities. In order to highlight the effects due to pinna alone, the HRIR signal is first windowed using a half Hann window (Oppenheim and Schafer, 1989). Windowing in the time domain helps isolate the direct component of the signal from the reflected components. A window of size 1.0 ms is used in order to eliminate the torso reflection [at around 1.6 ms in Fig. 3(a)] and the knee reflection [at around 3.2 ms in Fig. 3(a)]. Figure 4(b) shows the log magnitude spectrum of the windowed signal. We see that windowing the wave form reduces the effect of reflection significantly compared to the log magnitude spectrum in Fig. 3(b). We would like to point

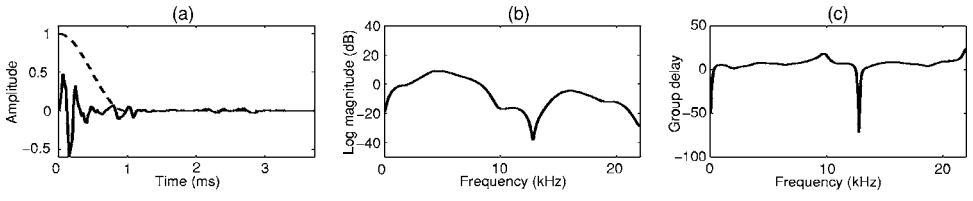

FIG. 4. Effect of windowing the HRIR. (a) HRIR (solid line) and half-Hann window (dotted line) of size 1.0 ms, (b) log magnitude spectra of the windowed signal, and (c) the corresponding group-delay function.

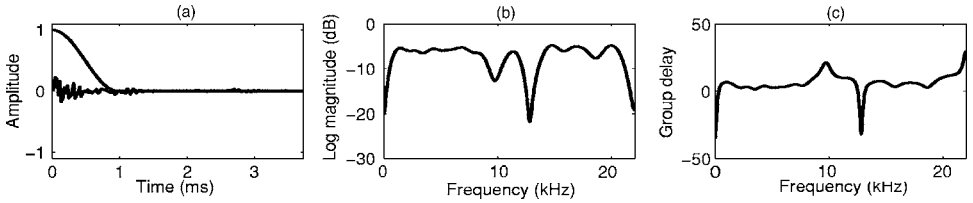

FIG. 5. Effect of windowing the LP residual of the HRIR. (a) The 12th order LP residual and half Hann window of size 1.0 ms, (b) the log magnitude spectra of the windowed LP residual signal, and (c) the corresponding group-delay function.

out that the exact size of the window is not crucial. A window size of 1.0 ms should be sufficient to suppress the reflections due to the torso and the knees.

We extract the spectral notches from the group-delay function rather than the magnitude spectrum. 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 even for a short time segment of the data (Yegnanarayana, 1978; Yegnanarayana *et al.*, 1984). The group delay function is the negative of the derivative of the phase spectrum of a signal. If $X(\omega)$ is the complex frequency response of a signal $x(n)$, then the group-delay function $\tau(\omega)$ is given by

$$\tau(\omega) = -\frac{d\theta(\omega)}{d\omega} \tag{1}$$

where ω is the angular frequency, and $\theta(\omega)$ is the phase angle of $X(\omega)$. The group-delay function can be computed directly using the Fourier transform of $x(n)$ and $nx(n)$, as follows (Oppenheim and Schafer, 1989). Let $X(\omega)$ and $Y(\omega)$ be the Fourier transforms of $x(n)$ and $nx(n)$, respectively,

$$\begin{aligned} X(\omega) &= \sum_{n=0}^{N-1} x(n)e^{-j\omega n} = X_R(\omega) + jX_I(\omega), \\ Y(\omega) &= \sum_{n=0}^{N-1} nx(n)e^{-j\omega n} = Y_R(\omega) + jY_I(\omega). \end{aligned} \tag{2}$$

Since

$$\log X(\omega) = \log|X(\omega)| + j\theta(\omega), \tag{3}$$

the group-delay function can be written as

$$\begin{aligned} \tau(\omega) &= -\frac{d}{d\omega}[\theta(\omega)] = -\operatorname{Im}\left(\frac{d}{d\omega}[\log X(\omega)]\right) \\ &= \frac{X_R(\omega)Y_R(\omega) + X_I(\omega)Y_I(\omega)}{X_R^2(\omega) + X_I^2(\omega)}, \end{aligned} \tag{4}$$

where $\operatorname{Im}(z)$ corresponds to the imaginary part of z. Figure 4(c) shows the group-delay function of the windowed signal (window size 1.0 ms). Compared to the log magnitude spectrum in Fig. 4(b), the group-delay function shows a better resolution of the notches.

However, windowing reduces the frequency domain resolution and also introduces artifacts. The artifacts of windowing may also mask or alter the frequencies of the spectral notches due to the pinna. One way to reduce the artifacts due to windowing is to remove the interdependence among adjacent signal samples by using the linear prediction (LP) residual of the original HRIR and then windowing the residual. This corresponds to removing the resonances from the signal.

The residual signal is derived using a 12th order LP analysis (Makhoul, 1975). LP analysis basically fits an all-pole model to the given signal. In LP analysis the signal $x(n)$ is predicted approximately from a linearly weighted summation of the past p samples, i.e.,

$$\hat{x}(n) = -\sum_{k=1}^{p} a_k x(n-k). \tag{5}$$

The error in the prediction, i.e., the LP residual, is therefore given by

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

The total squared error is

$$E = \sum_n e(n)^2 = \sum_n \left[x(n) + \sum_{k=1}^{p} a_k x(n-k) \right]^2. \tag{7}$$

Minimization of the mean squared error with respect to the coefficients $\{a_k\}$ gives the following normal equations (Makhoul, 1975):

$$\sum_{k=1}^{p} a_k R(n-k) = -R(k), \quad k = 0,1,\ldots,p, \tag{8}$$

where $R(k) = \Sigma_n x(n)x(n-k)$ is called the autocorrelation function for a lag of k samples. Equation (8) can be solved to get the coefficients $\{a_k\}$. Substituting the solution of the normal equations Eq. (8) into the expression for the error in Eq. (6) gives the sequence corresponding to the minimum total error, the LP residual.

LP analysis can be interpreted as the removal of redundancy in the signal samples by removing the predictable part from the signal. The linearly weighted past samples are used to predict the sample at the current sampling instant. 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. Effect of direct windowing of the LP residual is shown in Fig. 5, where the spectral notches can be seen to appear more prominently compared to the plots in Fig. 4.

The autocorrelation function of the windowed LP residual helps to reduce the effects due to truncation and noise. The autocorrelation function $R(m)$ of a signal $x(n)$ of length N is given by

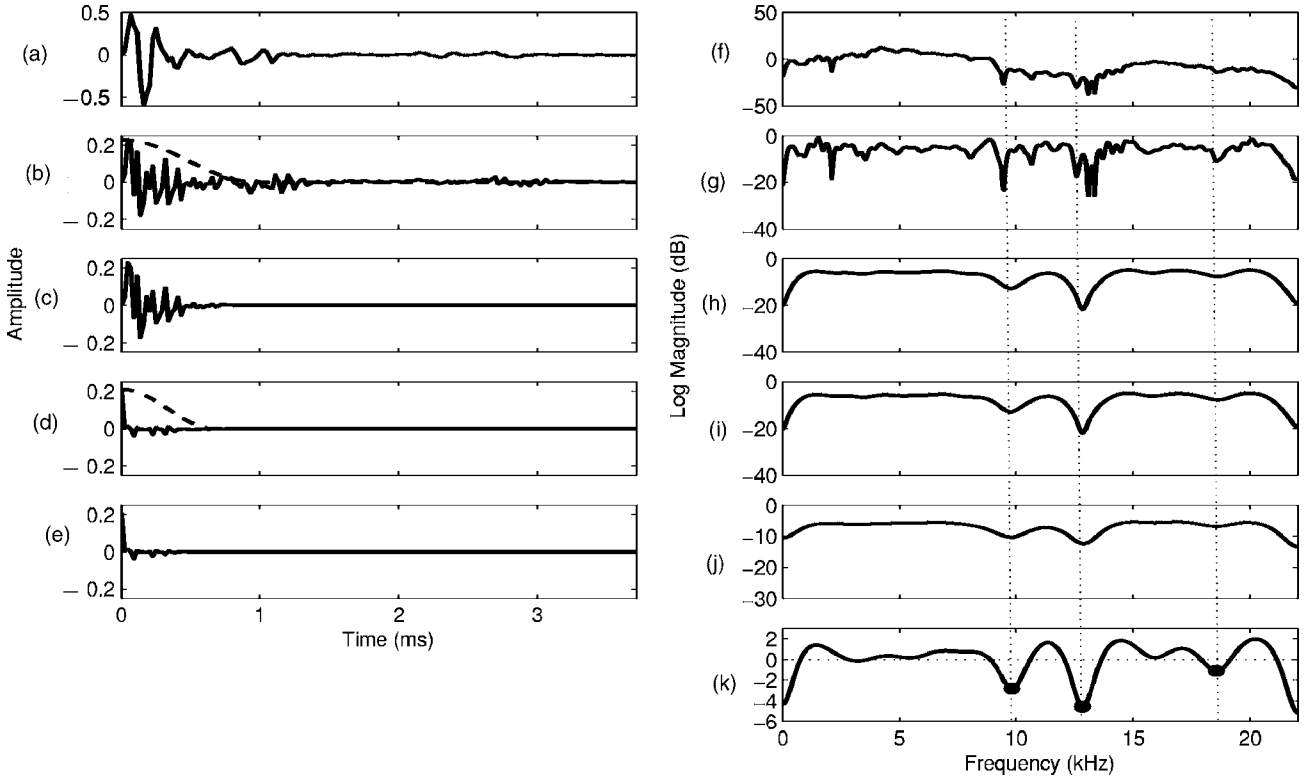

FIG. 6. Signal processing steps for extracting the pinna spectral notch frequencies. (a) Original HRIR signal, (b) 12th order LP residual, (c) windowed LP residual (1.0 ms half Hann window), (d) autocorrelation function of the windowed LP residual, and (e) windowed autocorrelation function (0.7 ms half 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.

$$R(m) = \sum_{n=m}^{N-1} x(n)x(n-m). \tag{9}$$

The autocorrelation function of a signal produces decreasing amplitudes away from its peak, which helps in computing the group-delay function better, while at the same time preserving most of the details of the spectral envelope. The resolution of the spectral components is enhanced in the group-delay function of the autocorrelation function of the windowed LP residual. 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 lower threshold of −1 was found to give better results and eliminated any spurious nulls caused due to windowing.

The sequence of signal processing operations is summarized in the following and the effect of each step on the HRIR and HRTF is shown in Fig. 6.

(1) Determine the initial onset of the HRIR and use the HRIR from that instant.
(2) Derive the pth ($p=10-12$) order LP residual from the given HRIR [Fig. 6(b)].
(3) Window the LP residual using a half Hann window of around 1.0 ms [Fig. 6(c)].
(4) Compute the autocorrelation function of the windowed LP residual [Fig. 6(d)].
(5) Window the autocorrelation function using a half Hann window of around 1.0 ms [Fig. 6(e)].
(6) Compute the group-delay function of the windowed autocorrelation function [Fig. 6(k)].
(7) Threshold the group-delay function and locate the local minima.

Since the spectra of the windowed LP residual is a smooth function with nulls, the spectrum can be inverted to obtain a spectrum with prominent peaks. An all-pole model can be fit to this spectrum by computing the autocorrelation function, and then applying the Levinson–Durbin method (Makhoul, 1975) for the first few (10) autocorrelation coefficients. The frequencies corresponding 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 also helps to get the depth and the width of the spectral notches. However, in order to extract the frequencies of the spectral notches the method based on group-delay is preferred since we do not need to specify the model order.

IV. RESULTS

The developed algorithm is applied on the measured HRIRs of different subjects and for different elevations and azimuth angles in the CIPIC database. As an example, Figs. 7(a)–7(h) show 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

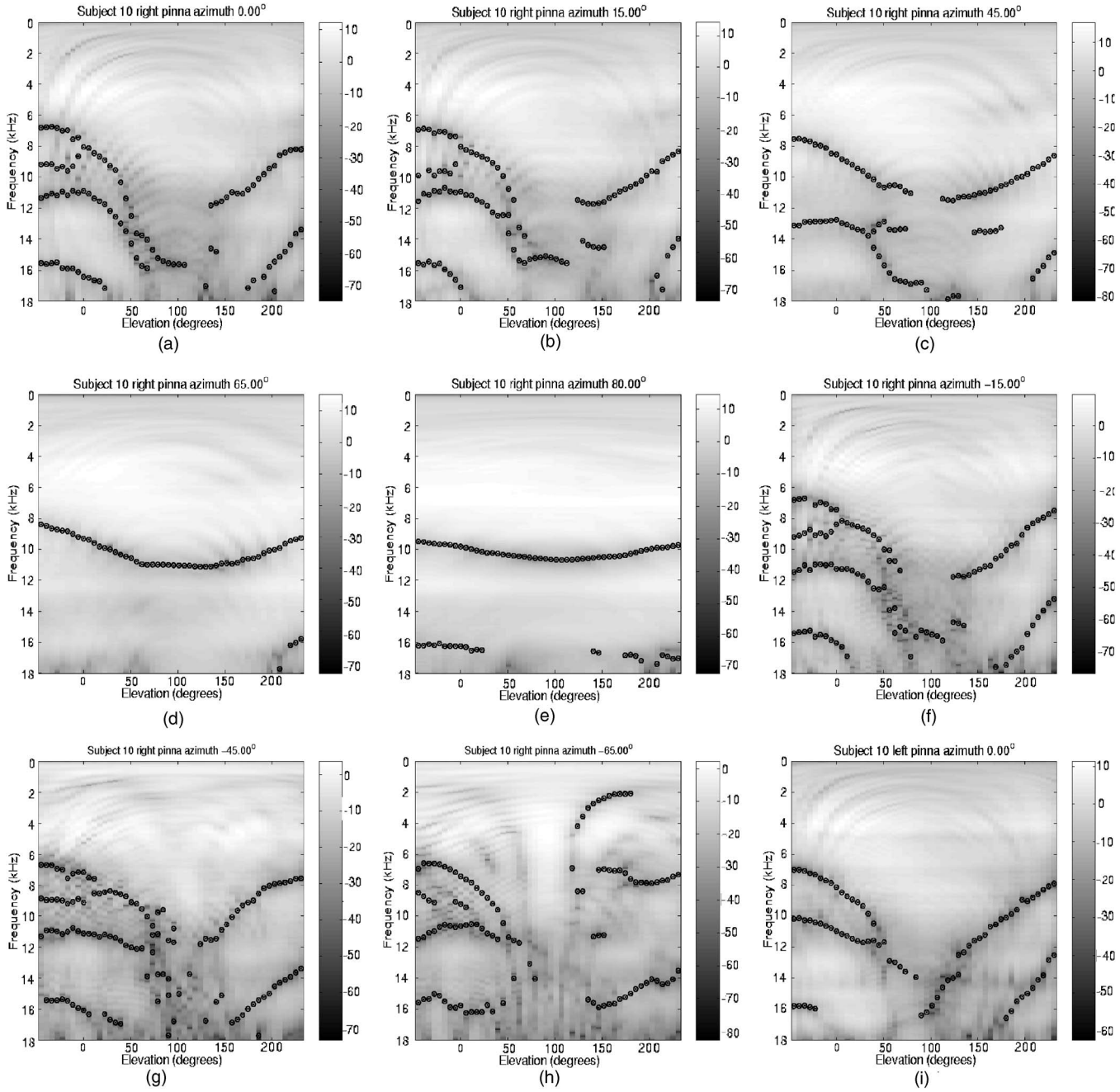

FIG. 7. The spectral notch frequencies extracted for subject 10 *right* pinna in the CIPIC database for azimuth angles (a)0°, (b)15°, (c)45°, (d)65°, (e)80°, (f) −15°, (g)−45°, and (h)−65°. (i) The spectral notch frequencies corresponding to the *left* pinna of subject 10 for azimuth 0°.

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 Fig. 7(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 (Lopez-Poveda and Meddis, 1996). However, since elevation perception is essentially thought to be monaural (Middlebrooks and Green, 1991) 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 cues for vertical localization. Figure 7(i) shows the notch frequencies for the left pinna for subject 10 and azimuth 0°. It was observed for most subjects that left and the right pinna do not have the same shape and dimensions (Algazi *et al.*, 2001c). The frequencies of the notches and their variation with elevation are different for the left and the right pinna. Figures 8 show the same results for nine different subjects and for azimuth 0°. Similar results are obtained when the analysis is applied to all the subjects in the database.

We note that LP analysis can also be used to extract spectral peaks in the spectrum. The poles extracted by LP analysis appear to correspond to the resonances of the pinna reported by Shaw (1997), who identified the normal modes

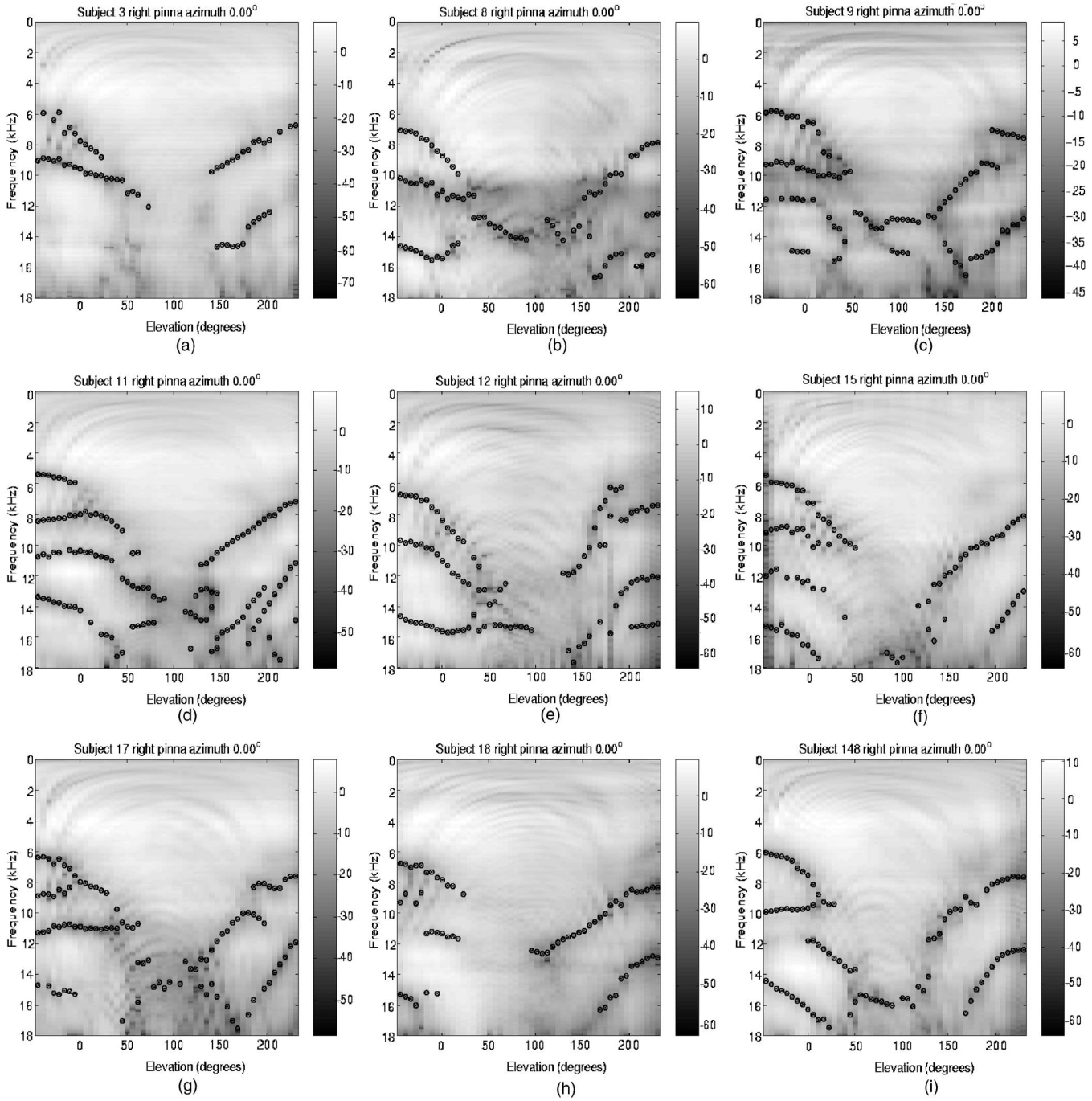

FIG. 8. The spectral notch frequencies for different elevations extracted for the right pinna for azimuth 0° for different subjects in the CIPIC database.

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 9 shows the frequency response of the 12th order all-pole model for the subject 10 for azimuth 0° as a function of different elevations as a mesh plot. These six modes are marked in the plot.

V. SPECTRAL NOTCHES AND PINNA SHAPE

The proposed procedure was successful in extracting the pinna spectral notches, visible to the human eye. We hope this would be a useful tool for researchers to study the significance of the pinna spectral notches to location perception and also to build structural models for the pinna. While perceptual tests are beyond the scope of this paper, we demonstrate a potential use of our procedure by showing that the pinna spectral notches are indeed related to the shape and anthropometry of the pinna.

The structure of the pinna is fairly complicated and difficult to characterize by simple models. To a first approxima-

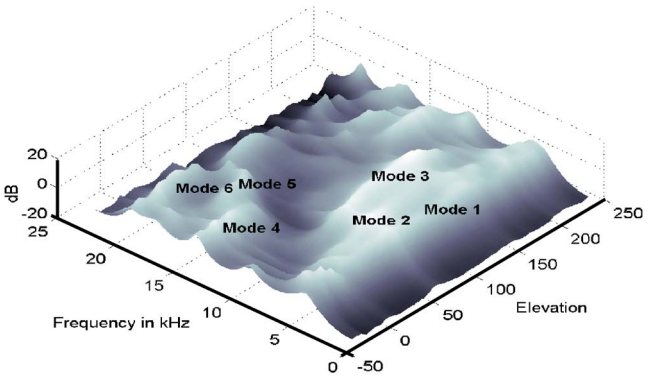

FIG. 9. Frequency response of the 12th order all-pole model for azimuth 0° as a function of different elevations. The six modes are approximately marked.

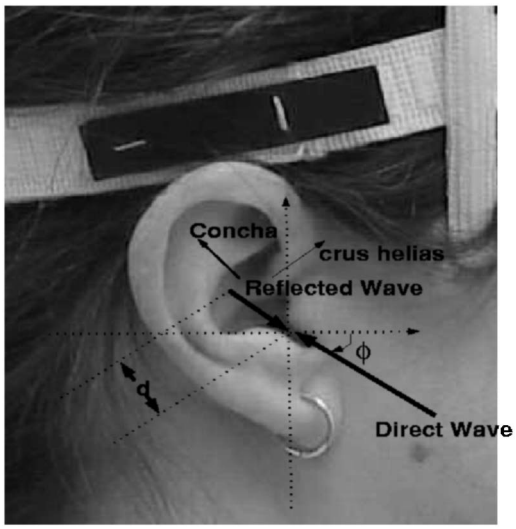

FIG. 10. A simple reflection model for the pinna spectral notches. The direct wave incident at an angle ϕ gets reflected from the concha. The time delay corresponds to a length of $2d$. The pinna image is taken from the CIPIC database.

tion the response can be characterized by peaks and notches observed in the spectrum. Figure 10 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 measured signal $y(t)$ is the sum of the direct and the reflected wave,

$$y(t) = x(t) + ax(t - t_d(\varphi)), \tag{10}$$

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

$$t_d(\varphi) = \frac{2d(\varphi)}{c}, \tag{11}$$

where $2d(\varphi)$ is the distance corresponding to the delay and c is the speed of the sound (approximately 343 m/s). The dis-

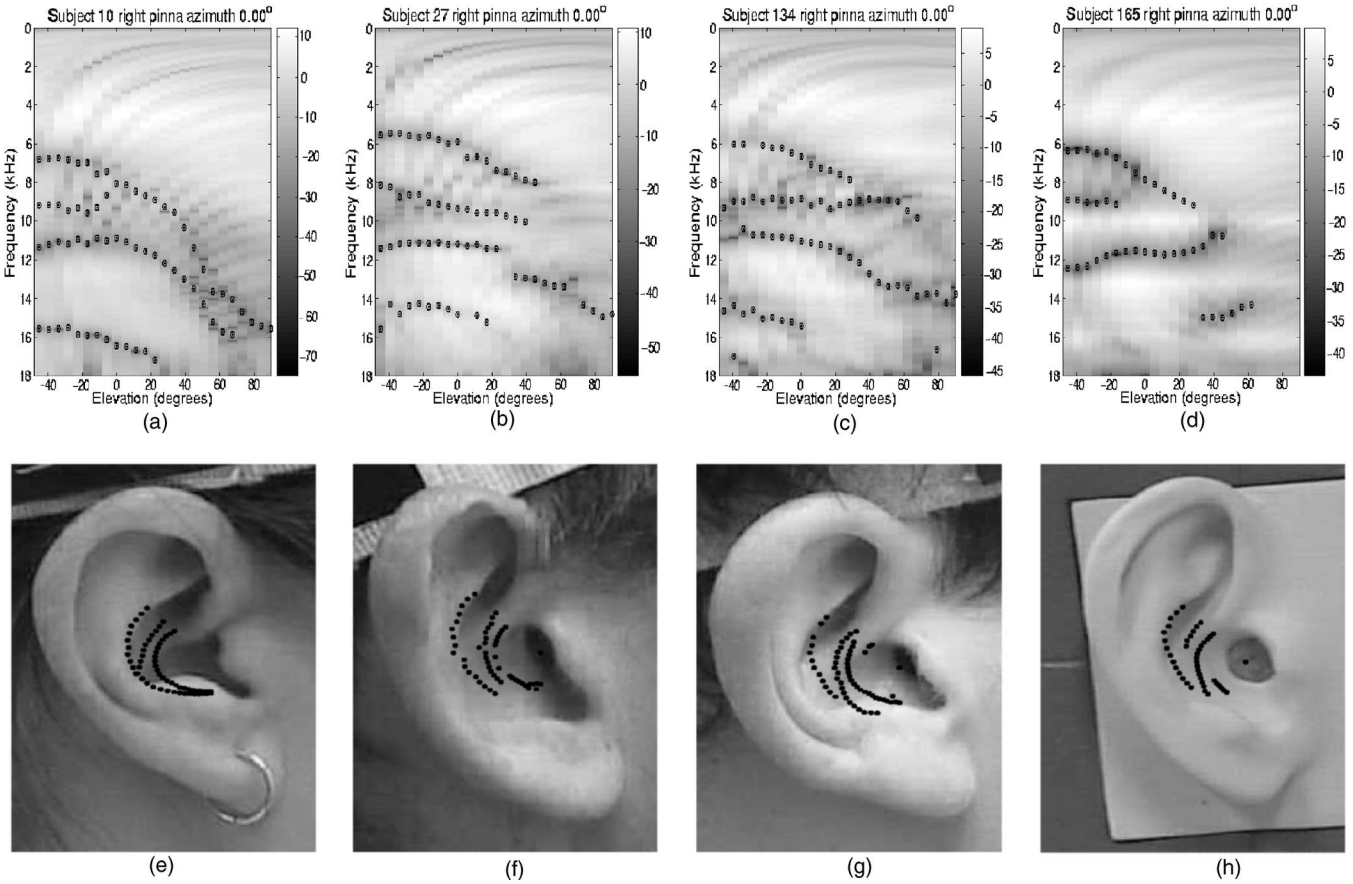

FIG. 11. The spectral notch frequencies for different elevations (from −45° to 90°) extracted for the right pinna for azimuth 0° (a) subject 10, (b) subject 27, (c) subject 134, and (d) subject 165 in the CIPIC database. The dimensions corresponding to the spectral notches marked on the pinna image for (e) subject 10, (f) subject 27, (g) subject 134, and, (h) subject 165, respectively. The pinna images are taken from the CIPIC database.

tance $d(\varphi)$ 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

$$f_n(\varphi) = \frac{(2n+1)}{2t_d(\varphi)} = \frac{c(2n+1)}{4d(\varphi)}, \quad n = 0, 1, \ldots . \tag{12}$$

The frequency of the first spectral notch is given by

$$f_0(\varphi) = \frac{c}{4d(\varphi)}. \tag{13}$$

In practice there are multiple reflections occurring in the pinna. Each reflection gives rise to a series of periodic spectral notches. In the previous section we extracted the frequencies of the spectral notches. From Eq. (13) we can calculate the distance $d(\phi)$ corresponding to the notch frequency $f_0(\phi)$. As 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 database. The distance can be marked on the pinna image approximately. Figure 11(a) shows the notch frequencies extracted for aximuth 0° and elevation varying from −45° to 90° (subject 10 right pinna). We consider only elevations in front of the head, since for elevations behind the ear the mechanism of the spectral notches is not clear. For each of the extracted notches the corresponding distance is plotted on the image of the pinna [Fig. 11(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 11 shows the same results for three other subjects in the CIPIC database, and exhibit similar trends.

These results suggest that the shape of the different cavities in the pinna are 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 particularly appealing method for synthesizing the HRIR would be to take the image of a pinna and obtain the notch frequencies by analyzing the pinna anthropometry.

VI. SUMMARY

We proposed signal processing algorithms for extraction of the pinna spectral notch frequencies from experimentally measured HRIRs. The difficulties in the analysis of HRIR caused by the combined effects of several components are discussed, and windowing in the time domain was proposed to reduce the effects of the reflected components. 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 spectral notch frequencies are related to the shape of the pinna. The code is made available to the research community on the first author's website.

ACKNOWLEDGMENTS

The support of NSF Award No. ITR-0086075 is gratefully acknowledged. We would also like to thank associate editor and the two reviewers for their comments and suggestions which helped to improve the clarity and quality of the paper. The first author would like to thank Prof. Richard Duda for the discussions on the CIPIC database.

Algazi, V. R., Avendano, C., and Duda, R. O. (**2001a**). "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**, 1110–1122.

Algazi, V. R., Duda, R. O., Duraiswami, R., Gumerov, N. A., and Tang, Z. (**2002**). "Approximating the head-related transfer function using simple geometric models of the head and torso," J. Acoust. Soc. Am. **112**, 2053–2064.

Algazi, V. R., Duda, R. O., Morrison, R. P., and Thompson, D. M. (**2001b**). "Structural Composition and Decomposition of HRTF's," in Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (Mohonk Mountain House, New Paltz, NY), pp. 103–106.

Algazi, V. R., Duda, R. O., Thompson, D. M., and Avendano, C. (**2001c**). "The CIPIC HRTF Database," in Proceedings of the 2001 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (Mohonk Mountain House, New Paltz, NY), pp. 99–102.

Asano, F., Suzuki, Y., and Sone, T. (**1990**). "Role of spectrum cues in median plane localization," J. Acoust. Soc. Am. **88**, 159–168.

Batteau, D. W. (**1967**). "The role of the pinna in human localization," Proc. R. Soc. London, Ser. B **168**, 158–180.

Blauert, J. (**1996**). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT, Cambridge, MA).

Blommer, M. A. and Wakefield, G. H. (**1997**). "Pole-zero approximations for head-related transfer functions using a logarithmic error criterion," IEEE Trans. Speech Audio Process. **5**, 278–287.

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

Duda, R. O. and Martens, W. L. (**1998**). "Range dependence of the response of a spherical head model," J. Acoust. Soc. Am. **104**, 3048–3058.

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

Gardner, M. B. and Gardner, R. S. (**1974**). "Problem of localization in the median plane: Effect of pinna cavity occlusion," J. Acoust. Soc. Am. **53**, 400–408.

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

Hebrank, J. and Wright, D. (**1974a**). "Are two ears necessary for localization of sound sources on the median plane?," J. Acoust. Soc. Am. **56**, 935–938.

Hebrank, J. and Wright, D. (**1974b**). "Spectral cues used in the location of sound sources on the median plane," J. Acoust. Soc. Am. **56**, 1829–1834.

Hofman, P., Van Riswick, J., and Van Opstal, A. (**1998**). "Relearning sound localization with new ears," Nat. Neurosci. **1**, 417–421.

Kahana, Y., Nelson, P. A., Petyt, M., and Choi, S. (**1999**). "Numerical modelling of the transfer functions of a dummy-head and of the external ear," in Proceedings of the AES 16th International Conference on Spatial Sound Reproduction, Rovaneimi, pp. 330–345.

Kuhn, G. F. (**1977**). "Model for interaural time differences in the azimuthal plane," J. Acoust. Soc. Am. **62**, 157–167.

Kulkarni, A. and Colburn, H. S. (**2004**). "Infinite-impulse-response models of the head-related transfer function," J. Acoust. Soc. Am. **115**, 1714–1728.

Langendijk, E. H. A. and Bronkhorst, A. W. (**2002**). "Contribution of spectral cues to human sound localization," J. Acoust. Soc. Am. **112**, 1583–1596.

Lopez-Poveda, E. A. and Meddis, R. (**1996**). "A physical model of sound diffraction and reflections in the human concha," J. Acoust. Soc. Am. **100**, 3248–3259.

Makhoul, J. (**1975**). "Linear prediction: A tutorial review," Proc. IEEE **63**, 561–580.

Middlebrooks, J. C. (**1999**). "Virtual localization improved by scaling non-individualized external-ear functions in frequency," J. Acoust. Soc. Am. **106**, 1493–1509.

Middlebrooks, J. C. and Green, D. M. (**1991**). "Sound localization by human listeners," Annu. Rev. Psychol. **42**, 135–159.
Moore, B. C. J., Oldfield, S. R., and Dooley, G. (**1989**). "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," J. Acoust. Soc. Am. **85**, 820–836.
Musicant, A. and Butler, R. (**1984**). "The influence of pinnae-based spectral cues on sound localization," J. Acoust. Soc. Am. **75**, 1195–1200.
Oppenheim, A. and Schafer, R. W. (**1989**). *Discrete-Time Signal Processing* (Prentice-Hall, Englewood Cliffs, NJ).
Poon, P. and Brugge, J. F. (**1993a**). "Sensitivity of auditory nerve fibers to spectral notches," J. Neurophysiol. **70**, 655–666.
Poon, P. and Brugge, J. F. (**1993b**). "Virtual-space receptive fields of single auditory nerve fibers," J. Neurophysiol. **70**, 667–676.
Raykar, V. C., Yegnanarayana, B., Duraiswami, R., and Davis, L. (**2003**). "Extracting significant features from the HRTF," in Proceedings of the 2003 International Conference on Auditory Display, pp. 115–118.
Shaw, E. A. G. (**1997**). "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. (Erlbaum, Mahwah, NJ), pp. 25–47.
Shinn-Cunningham, B. G., Santarelli, S. G., and Kopco, N. (**2000**). "Tori of confusion: Binaural cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636.
Steiglitz, K. and McBride, L. E. (**1965**). "A technique for the identification of linear systems," IEEE Trans. Autom. Control **10**, 461–464.
Strutt, J. W. (**1907**). "On our perception of sound direction," Philos. Mag. **13**, 214–232.
Tollin, D. J. and Yin, T. C. T. (**2003**). "Spectral cues explain illusory elevation effects with stereo sounds in cats," J. Neurophysiol. **90**, 525–530.
Wenzel, E. M., Arruda, M., Kistler, D. J., and Wightman, F. L. (**1993**). "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**, 111–123.
Wightman, F. L. and Kistler, D. J. (**1997**). "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 T. B. Anderson (Erlbaum, Mahwah, NJ).
Wright, D., Hebrank, J. H., and Wilson, B. (**1974**). "Pinna reflections as cues for localization," J. Acoust. Soc. Am. **56**, 957–962.
Yegnanarayana, B. (**1978**). "Formant extraction from linear prediction phase spectra," J. Acoust. Soc. Am. **63**, 1638–1640.
Yegnanarayana, B., Saikia, D. K., and Krishnan, T. R. (**1984**). "Significance of group delay functions in signal reconstruction from spectral magnitude or phase," IEEE Trans. Acoust., Speech, Signal Process. **32**, 610–623.
Zotkin, D. N., Duraiswami, R., Davis, L., Mohan, A., and Raykar, V. C. (**2002**). "Virtual audio system customization using visual matching of ear parameters," in Proceedings of the 2002 International Conference on Pattern Recognition, pp. 1003–1006.