On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features

Simone Spagnol, Michele Geronazzo, and Federico Avanzini

***Abstract*—This paper studies the relationship between head-related transfer functions (HRTFs) and pinna reflection patterns in the frontal hemispace. A pre-processed database of HRTFs allows extraction of up to three spectral notches from each response taken in the median sagittal plane. Ray-tracing analysis performed on the obtained notches' central frequencies is compared with a set of possible reflection surfaces directly recognizeable from the corresponding pinna picture. Results of such analysis are discussed in terms of the reflection coefficient sign, which is found to be most likely negative. Based on this finding, a model for real-time HRTF synthesis that allows to control separately the evolution of different acoustic phenomena such as head diffraction, ear resonances, and reflections is proposed through the design of distinct filter blocks. Parameters to be fed to the model are derived either from analysis or from specific anthropometric features of the subject. Finally, objective evaluations of reconstructed HRTFs in the chosen spatial range are performed through spectral distortion measurements.**

***Index Terms*—Acoustic signal processing, anthropometry, auditory displays, head-related transfer functions (HRTFs), spatial hearing.**

I. Introduction

THE ability of the human auditory system to estimate the spatial location of sound sources in an acoustic scene has high survival value as well as a relevant role in several everyday tasks: detecting potential dangers in the environment, selectively focusing attention on one stream of information, and so on. Audition performs remarkably at this task, complementing the information provided by the visual channel: as an example, it can provide localization information on targets that are out of sight.

Accordingly, in recent years spatial sound has become increasingly important in several application domains. Spatial rendering of sound is recognized to greatly enhance the effectiveness of auditory human-computer interfaces [1], particularly in cases where the visual interface is limited in extension and/or resolution, as in mobile devices [2]; it improves the sense of presence in augmented/virtual reality systems [3], and adds engagement to computer games.

Manuscript received October 25, 2011; revised May 30, 2012; accepted October 16, 2012. Date of publication November 15, 2012; date of current version December 31, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Patrick A. Naylor.

S. Spagnol is with Iuav University of Venice, 30123 Venice, Italy (e-mail: sspagnol@iuav.it).

M. Geronazzo and F. Avanzini are with the Department of Information Engineering, University of Padova, 35131 Padua, Italy (e-mail: geronazzo@dei.unipd.it; avanzini@dei.unipd.it).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASL.2012.2227730

Auditory cues related to directional information include binaural cues, such as interaural level and time differences, and monaural cues, such as the spectral coloration resulting from filtering effects of the human body, especially from the external ear. All these features are summarized into the so-called *Head Related Transfer Functions (HRTFs)* [4], i.e., the frequency- and space-dependent acoustic transfer functions between the sound source and the eardrum.[1] Binaural spatial sound can be synthesized by convolving an anechoic sound signal with the corresponding left and right HRTFs.

Non-individualized HRTF sets are typically recorded using "dummy heads", i.e., mannequins constructed from averaged anthropometric measures, and represent a cheap and straightforward mean of providing 3-D rendering in headphone reproduction. However, they are known to produce evident sound localization errors [5], including incorrect perception of elevation, front-back reversals, and lack of externalization [6], especially when head tracking is not utilized in the reproduction [7]. Therefore, individual anthropometric features have a key role in characterizing HRTFs. On the other hand, HRTF measurements on a significant number of subjects are both expensive and inconvenient.

Structural HRTF modeling [8] represents an attractive solution to these shortcomings. By isolating the effects of different components (head, pinnae, ear canals, shoulders/torso), and modeling each one of them with a corresponding filtering element, the global HRTF is approximated through a proper combination of all the considered effects. Moreover, by relating the temporal/spectral features (or equivalently, the filter parameters) of each component to corresponding anthropometric quantities, one can in principle obtain a HRTF representation that is both computationally economical and customizable.

Following the structural modeling approach, this work investigates the contribution of the external ear to the HRTF, the Pinna-Related Transfer Function (PRTF). While the pinna is known to play a primary role in the perception of source elevation, the relation between PRTF features—resonances associated to cavities and spectral notches resulting from reflections [9]—and anthropometry is not fully understood. Recent related works [10]–[12] adopt a physical modeling approach in which PRTFs are simulated through computationally intensive techniques, such as finite-difference time-domain (FDTD) methods, or boundary elements methods (BEM). Other works [13]–[15] utilize series expansions, such as principal component analysis

[1]More formally, the HRTF at one ear is the frequency-dependent ratio between the sound pressure level (SPL) $\Phi(\theta, \phi, \omega)$ at the eardrum and the free-field SPL at the center of the head $\Phi_f(\omega)$ as if the listener were absent: $H(\theta, \phi, \omega) = \Phi(\theta, \phi, \omega)/\Phi_f(\omega)$, where (θ, ϕ) indicates the angular position of the source relative to the listener, and ω is the angular frequency.

(PCA) or surface spherical harmonics (SSH) representations of HRTFs and PRTFs.

Alternatively, the relationship between PRTF features and pinna geometry can be studied by directly analyzing real measured HRTFs, and by relating relevant extracted spectral features to known anthropometric data [16], [17]. In this paper we follow this latter approach: we estimate and analyze PRTFs of 20 subjects from a public domain database, and focus on the relationship between PRTF notches and pinna contours. The results of this work are the first step in the development of a parametric PRTF model that can be customized according to individual anthropometric data, which in turn can be automatically estimated through straightforward image analysis.

The remainder of the paper is organized as follows. Section II discusses the relevant literature on source elevation perception, pinna mechanisms, and structural modeling of PRTFs, while Section III focuses on data collection and feature extraction. In Section IV we study the relation between pinna reflection patterns and anthropometry. Finally, a structural model of the pinna is proposed and objectively evaluated in Section V.

II. Background and Previous Works

Directional hearing in the median vertical plane has long been known to have a coarser resolution compared with the horizontal plane [18]. The threshold for detecting changes in the direction of a sound source (known as "localization blur") along the median plane was found to be never less than 4°, reaching a much larger threshold ($\approx 17°$) for unfamiliar speech sounds, as opposed to a localization blur of approximately 1°–2° in the horizontal plane for a vast class of sounds [19]. Such a poor resolution is motivated by two basic observations:

- the theoretically nonexistent interaural differences between the signals arriving at the left and right ear, which conversely play a primary role in horizontal perception;
- the need of high-frequency content (above 4–5 kHz) for accurate vertical localization [20]–[22].

It is undisputed that vertical localization ability is brought by the presence of the pinnae [23]. Even though localization in any plane involves pinna cavities of both ears [24], determination of the perceived vertical angle of a sound source in the median plane is essentially a monaural process [25]. The external ear plays an important role by introducing peaks and notches in the high-frequency spectrum of the HRTF, whose center frequency, amplitude, and bandwidth greatly depend on the elevation angle of the sound source [26], to a remarkably minor extent on azimuth [27], and are almost independent on distance between source and listener beyond a few centimeters from the ear [28].

Following two historical theories of localization, the pinna can be seen both as a filter in the frequency domain [19] and a delay-and-add reflection system in the time domain [9] as long as typical pinna reflection delays for elevation angles, clearly detectable by the human hearing apparatus [29], are seen to produce spectral notches in the high-frequency range.

The evolution of notches in the median plane was studied by Raykar *et al.* [16]. Robust digital signal processing techniques based on the residual of a linear prediction model were applied to measured head-related impulse responses (HRIRs) in order to extract the frequencies of those spectral notches caused by the presence of the pinna. The authors exploited a simple ray-tracing law (borrowed from [20]) to show that the estimated spectral notches, each assumed to be caused by its own reflection path, are related to the shape of the concha and crus helias, at least on the frontal side of the median plane. However, there is no clear one-to-one correspondence between pinna contours and notch frequencies in the available plots.

Additionally to reflections, pinna resonances and diffraction inside the concha were also seen to contribute to the HRTF spectral shape. Shaw [30] identified six resonant modes of the pinna excited at different directions which clearly produce the most prominent HRTF spectral peaks: an omnidirectional resonance at 4.2 kHz (mode 1), two vertical resonances at 7.1 and 9.6 kHz (modes 2 and 3), and three horizontal resonances at 12.2, 14.4, and 16.7 kHz (modes 4, 5, and 6).[2] These results find accordance in a more recent study by Kahana *et al.* [11] on numerical simulation of PRTFs using BEM over baffled pinna meshes.

Concerning diffraction effects, Lopez-Poveda and Meddis [27] motivated the slight dependence of spectral notches on azimuth through a diffraction process that scatters the sound within the concha cavity, allowing reflections on the posterior wall of the concha to occur for any direction of the sound. Presence of diffraction around the tragus area has also been recently hypothesized by Mokhtari *et al.* [12], [31].

Nevertheless, the relative importance of major peaks and notches in elevation perception has been disputed over the past years.[3] A recent study [32] showed how a parametric HRTF recomposed using only the first, omnidirectional peak in the HRTF spectrum (corresponding to Shaw's mode 1) coupled with the first two notches yields almost the same localization accuracy as the corresponding measured HRTF. Additional evidence in support of the lowest-frequency notches' relevance is given in [21], which states that the threshold for perceiving a shift in the central frequency of a spectral notch is consistent with the localization blur on the median plane. Also, in [20] the authors judge increasing frontal elevation apparently cued by the increasing central frequency of a notch, and determine two different peak/notch patterns for representing the above and behind direction. In general, hence, both peaks and notches seem to play an important function in vertical localization of a sound source.

With the purpose of creating the best possible approximation to the above patterns, different physical and structural models of the pinna have been proposed in the past. The former class aims at recreating the physics lying behind the production of the aforementioned spectral patterns either by approximating the pinna as a cavity configuration or as a reflecting surface. Examples of the first approach are the simple geometric (cylindrical or rectangular) concha/pinna models by Teranishi and

[2]The reported center frequencies were averaged among 10 different pinnae. Vertical modes are excited by sources above the head; horizontal modes by sources in the vicinity of the horizontal plane.

[3]In this context, it is important to point out that both peaks and notches in the high-frequency range are perceptually detectable as long as their amplitude and bandwidth are sufficiently marked [21], which is the case for most measured HRTFs.

Shaw [33], which progressively led to Shaw's notable flange-and-cavity model [34], and the recent "three-step" model by Takemoto *et al.* [35], simulated through the Finite-Difference Time Domain (FDTD) method, which qualitatively recreates typical peak/notch patterns along the median plane. The second approach is best exemplified by the rigorous diffraction/reflection model by Lopez-Poveda and Meddis [27] based on diffraction theory applied to both a half-cylinder shape and a realistic concha shape. Despite the objectively good approximations that physical models can provide, their main drawback is the difficulty in introducing effective customizations to the physical structure.

The history of structural models, one of which will be described in this paper, begins with Batteau's reflection theory [9]. Following Batteau's observations, Watkins [36] designed a very simple double-delay-and-add time-domain model of the pinna where the first reflection path is characterized by a fixed time delay of 15 μs while the second path includes an elevation-dependent delay calculated from empirical data. Beside considering a very limited amount of reflections, no method for extracting parametric time delays and gain factors was proposed. Furthermore, simple delay-and-add approximations were proven to be inadequate to predict both the absolute position of the spectral minima and the relative position between them [27]. Nonetheless, the pioneering novelty of such model is undisputed.

A similar time-domain structural model, proposed by Faller *et al.* [37], is composed of multiple parallel reflection paths each including a different time delay, a gain factor, and a low-order resonance block. The model is fitted by decomposing a measured HRIR into a heuristic number of damped and delayed sinusoidals (DDS) using an adaptation of the Hankel Total Least Squares (HTLS) decomposition method, and associating the parameters of each DDS to the corresponding parameters of its relative model path. Still, no relation between model parameters and human anthropometry was explicitly found.

Moving from time domain to frequency domain, the approach followed by Satarzadeh *et al.* [17] approximates PRTFs at elevations close to zero degrees through a structural model composed of two low-order bandpass filters and one comb filter which account for two resonance modes (Shaw's modes 1 and 4) and one main reflection, respectively. What's more relevant, a cylindrical approximation to the concha is exploited for fitting the model parameters to anthropometric quantities. Specifically, depth and width of the cylinder uniquely define the first resonance, while the second resonance is thought to be correlated to the main reflection's time delay, depending on whether the concha or the rim is the significant reflector. The authors show that their model has sufficient adaptability to fit both PRTFs with rich and poor notch structures. One limitation is that no direction of the sound wave other than the frontal one is considered; moreover, the presence of an unique reflection (and thus a single delay-and-add approximation) limits the generality of the representation. Nonetheless it represents, in the authors' opinion, the only valuable anthropometry-based pinna model available to date.

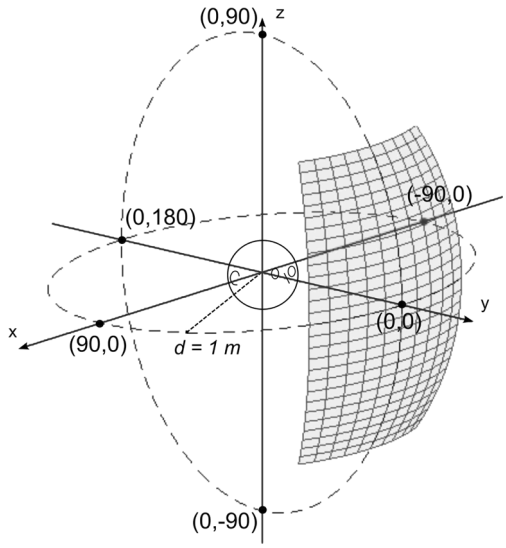


Fig. 1. Interaural polar coordinate system (reported through six (θ, ϕ) coordinates) and spatial range of validity of the model.

III. PRTF Analysis

Satarzadeh's filter model [17] can be generalized through consideration of multiple reflection paths, and extended to a wider frontal space. From this section onwards we describe the steps that allow construction of a multi-notch filter suitable for anthropometric parametrization as a replacement to the simpler comb filter.

A. Data Collection and Pre-Processing

Extraction of notches' parameters first requires a PRTF analysis step. Our initial data set consists of measured HRIRs taken from the CIPIC database [38], a public-domain database of high spatial resolution HRIR measured at 1250 directions for 45 different subjects. Since our work involves the anthropometry of these subjects in the form of a picture of their left or right pinna, we restrict our attention to the 20 of them for which the corresponding photograph is available [39]: subjects 003, 008, 009, 010, 011, 012, 015, 017, 019, 020, 021 (KEMAR with large pinna), 027, 028, 033, 040, 044, 048, 050, 134, and 165 (KEMAR with small pinna). Taking as reference system the interaural polar coordinate system defined in [38] and sketched in Fig. 1, we focus on median-plane (azimuth angle $\theta = 0°$) HRIRs, with the elevation angle ϕ varying from $\phi = -45°$ to $\phi = 45°$ at 5.625-degree steps (17 HRIRs per subject). We choose to consider the median plane because relative azimuthal variations up to at least $\Delta\theta = 30°$ at fixed elevation cause very slight spectral changes in the PRTF [16], [27], [31], hence we expect PRTFs in this region to be elevation-dependent-only. The upper elevation limit ($\phi = 45°$) was chosen because of the high degree of uncertainty in elevation judgement for sources at $\phi > 45°$ [19], [24] and the general lack of deep spectral notches in PRTFs in this region [11], [16], [40], which may besides be two faces of the same coin. Thus the angular range of validity of our model will be at least as broad as the shaded area depicted in Fig. 1.

The first problem that needs to be addressed is how to extract the PRTF from the corresponding (left or right, depending on

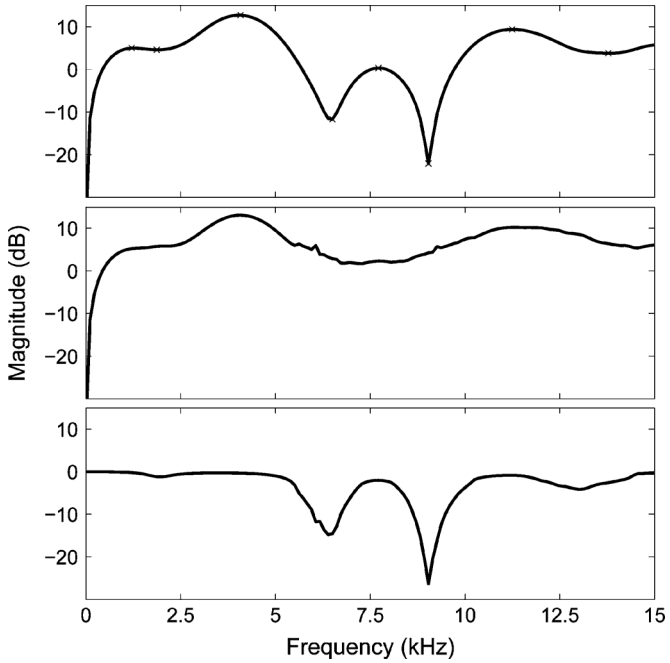


Fig. 2. Top panel: right PRTF of CIPIC subject 003 ($\theta = 0°, \phi = -28.125°$), magnitude spectrum. Middle panel: the PRTF resonant component extracted by the separation algorithm [44]. Bottom panel: the PRTF reflective component extracted by the separation algorithm.

the available pinna image) HRIR: basically, the head, torso and shoulders contributions need to be discarded from the response. Knowing that pinna reflection delays usually range between 100 and 300 μs in the median plane [9], we shorten the HRIR by applying a 1-ms Hann window starting from the HRIR onset [16]. In this way spectral effects due to reflections caused by shoulders and torso are removed from the response, while those due to the pinna are preserved. Concerning head diffraction compensation, if we virtually treat the pinnaless head as a sphere,[4] then the ear canal lies around $\theta = \pm 90°$–$100°$.[5] It can be directly seen [43] that the corresponding responses of spherical diffraction for a source in the frontal side of the median plane at 1 meter (where CIPIC HRTF measurements were taken) are approximately flat. Further evidence of such "flatness" is found in [31], where the authors provide graphical evidence that the spectral distance between FDTD-simulated HRTFs of a complete KEMAR head and PRTFs of its pinna alone is negligible in the median plane.

As a consequence, no further preprocessing step is applied to the windowed and zero-padded HRIR, whose FFT, calculated on a 512-sample window size, yields the estimated PRTF (see top panel of Fig. 2).

B. Feature Extraction

The next issue concerns feature extraction from the obtained PRTF. We choose to treat reflections and resonances as two separated phenomena and thus split the PRTF into a "resonant" and a "reflective" component by means of a separation algorithm, whose details are reported in [44]. The idea that drives the algorithm is the iterative compensation of the PRTF magnitude spectrum through a sequence of synthetic multi-notch filters until no local notches above a given amplitude threshold are left. Each multi-notch filter is fitted to the shape of the PRTF spectrum at the current iteration with its spectral envelope removed and subtracted to it, giving the spectrum for the next iteration. Eventually, when convergence is reached the final spectrum contains the resonant component, while the reflective component is given by direct combination of all the calculated multi-notch filters. An example of the algorithm output is reported in Fig. 2.

Analysis of the resonant component in different CIPIC subjects reveals common trends with respect to elevation. In particular, two prominent peaks at quasi-steady central frequencies can be distinctly identified in the considered frequency range, the first around 4 kHz corresponding to Shaw's omnidirectional mode, and the second around 12 kHz corresponding to the first horizontal mode. By contrast, since common trends cannot be identified in the evolution of spectral notches, and following the common idea that notches are of major relevance for elevation detection in the frontal region [20], [21], [29], [32], we focus our attention onto the reflective component.

Similarly to [16], we choose to treat each notch as the result of a distinct reflection path. Also, similarly to previous works on reflection modeling [16], [17] we consider as the most relevant notch feature its own central frequency. Inspection of different PRTF plots reveals that the notch moves continuously along the frequency axis depending on the elevation angle [20], [26] to an extent that can definitely be detected by the human auditory system [21]. Conversely, changes in notch bandwidth and amplitude along elevation are seen to be far less systematic [45], and their perceptual relevance is little understood in previous literature.

Notch frequencies are obtained through a simple notch picking algorithm [46]. In order to have a consistent labeling along subsequent PRTFs, extracted notches need to be grouped into tracks evolving through elevation. To this end, we exploit the McAulay-Quatieri partial tracking algorithm [47] and fit it to our needs. The original formulation of the algorithm can be used to track the most prominent notch patterns along elevation, with elevation dependency conceptually replacing temporal evolution, and spectral notches taking the role of sinusoidal partials. The obtained notch track collection is reduced by keeping only those tracks which remain inside the range 4–16 kHz, where pinna cues are most likely to be detected [20]. Further details are given in [46].

As a result, the majority of the 20 considered CIPIC subjects exhibits three notch tracks at a given elevation. Only subjects 019 and 020 lack of one track, the lowest and the highest in frequency respectively. Average notch frequencies in the three tracks at each available elevation are reported in Fig. 3, along with their standard deviation: frequencies in the first two tracks (T_1 and T_2) monotonically grow with elevation, while frequencies in the third track (T_3) remain almost constant up to $\phi = -11.25°$, then grow until $\phi = 28.125°$, and decrease at higher elevations on average. Despite the significant variance in the central frequencies of the three tracks (T_3 in particular), these

[4]In [41] it is shown that there is roughly no difference between FDTD-simulated responses on an unmodified KEMAR head and on a head shape morphed towards a sphere in the median plane.

[5]Since human ears typically lie slightly behind and below the x axis [42], the source-ear angular distance is certainly greater than 90° for sources between $\phi = 0°$ and $\phi = 45°$ at least.

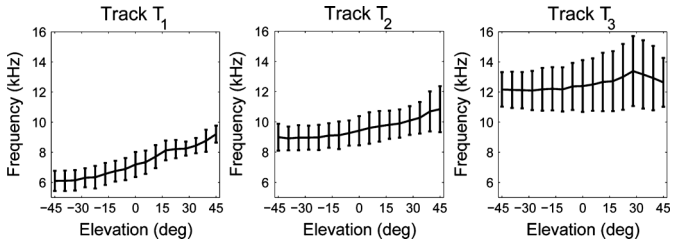


Fig. 3. Mean and standard deviation of notch frequencies per elevation and track across 20 subjects.

trends were seen to be consistent across subjects. Not reported in the figure is the number of subjects that exhibit a notch for each track/elevation coordinate: for the sake of brevity, suffice it to mention that all tracks begin at $-45°$ except for three cases only, that T_1 terminates earlier than T_2 on average, and the same applies to T_2 with respect to T_3.

IV. Reflections and Anthropometry

Ray-tracing reflection models [20] assume ray-like rather than wave-like behavior of sound, providing a crude approximation of the wave equation. Despite this, the approach conveyed by such models is valid as long as the wavelength of the sound is small when compared to the dimensions of the involved reflection surfaces. This is definitely the case of the audible spectrum's higher frequencies, where spectral notches due to pinna reflections appear. In this context, one can intuitively observe that the elevation-dependent temporal delay $t_d(\phi)$ between the direct and the reflected wave projects the point of reflection at distance

$$d_c(\phi) = \frac{ct_d(\phi)}{2} \tag{1}$$

from the ear canal (where c is the speed of sound). Assuming the reflection coefficient to be positive, then we will have destructive interference (i.e., a notch) at all those frequencies where the reflection's phase shift equals π:

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

Hence the first notch falls at frequency

$$f_0(\phi) = \frac{c}{4d_c(\phi)}. \tag{3}$$

The positive reflection assumption was also adopted by Raykar [16] when tracing reflection points over pinna images based on the extracted notch frequencies.

Nevertheless, Satarzadeh [48] drew attention to the fact that almost 80% of CIPIC subjects exhibit a clear negative reflection in their HRIRs and proposed a physical explanation to this phenomenon. In case of negative reflection, destructive interference would not appear at half-wavelength delays anymore, but at full-wavelength delays. Equations (2) and (3) would then become

$$f_n(\phi) = \frac{n+1}{t_d(\phi)} = \frac{c(n+1)}{2d_c(\phi)}, \quad n = 0, 1, \ldots \tag{4}$$

and

$$f_0(\phi) = \frac{c}{2d_c(\phi)}. \tag{5}$$

Note that since our extracted notch tracks are pairwise in nonharmonic relationship, both on average (see again Fig. 3) and for every single subject, we cannot assign a single reflection path to any pair of tracks. Hence our previous assumption that each notch in the considered frequency range is the result of a distinct reflection path is well-grounded.

In the following, we treat each extracted notch frequency as the f_0 of its respective reflection, and investigate the correspondence between pinna anatomy and theoretical reflection points under different reflection sign conditions on a wide morphological variety of CIPIC subjects' pinnae. We now present the formal analysis procedure, which was informally sketched in an earlier work [46] on four subjects only. Results are presented and discussed at the end of the Section.

A. Contour Matching Procedure

The basic assumption that drives our analysis procedure is that each notch track is associated with a distinct reflection surface on the subject's pinna. Since the available data for each subject is a side-view of his/her head showing the left or right pinna, extraction of the "candidate" reflection surfaces must be reduced to a two-dimensional basis. We choose to investigate as possible reflection surfaces a set of three contours directly recognizeable from the pinna photograph, together with two hidden surfaces approximating the real inner back walls of the concha and helix. Specifically, as Fig. 4 depicts, we consider the following contours:

1) helix border (C_1), visible on picture;
2) helix inner wall (C_2), following the jutting light surface at the helix approximately halfway between the rim border and the rim outer wall;
3) concha outer border (C_3), visible on picture;
4) antihelix and concha inner wall (C_4), following the jutting light surface just behind the concha outer border up to the shaded area below the antitragus;
5) crus helias inferior surface (C_5), visible on picture.

Since automatic contour extraction is beyond the scope of this paper, the extraction procedure was performed by manual tracing through a pen tablet. Photographs were accurately resized to match a 1:1 scale based on the quantitative pinna height parameter (d_5 in [38]) available from the HRTF database's anthropometric data, or based on the measuring tape pictured in the photograph close to the pinna in those cases where d_5 was not defined. Right pinna photographs were horizontally mirrored so that all pinnae headed left, and contours were drawn and stored as sequences of pixels in the post-processed image. Of all the contours, C_4 was the hardest to recognize due to the low resolution of the pictures; it is therefore necessary to point out that in some cases the lower part of this contour was almost blindly traced.

Before describing the contour matching procedure, let us formally state some useful definitions.

- the *focus* $\psi = (\psi_x, \psi_y)$ is the reference point where the direct and reflected waves meet, usually set at the entrance of the ear canal where the microphone is assumed to have been placed during HRTF measurements;

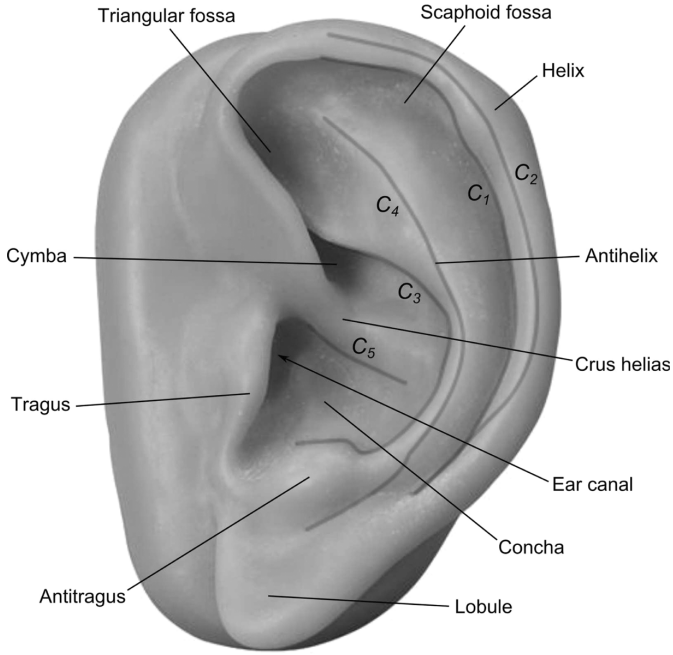


Fig. 4. Pinna anatomy and the five chosen contours for the matching procedure. C_1: helix border; C_2: helix wall; C_3: concha border; C_4: antihelix and concha wall; C_5: crus helias.

- the rotation ρ is a tolerance on elevation that counterbalances possible angular mismatches between the actual orientation of the subject's ear and the picture's x-axis;
- a *reflection sign configuration* $\boldsymbol{s} = [s_1, s_2, s_3]$ (with $s_j = \{0, 1\}$), abbreviated as *configuration*, is the combination of reflection coefficient signs attributed to the three notch tracks $\{T_1, T_2, T_3\}$. Here s_j takes 0 value if a negative sign is attributed to T_j and 1 otherwise;
- the *distance* $d(p, C_i)$ *between a point* p *and a contour* C_i is defined as the Euclidean distance between p and the nearest point of C_i.

Our goal is to discover which of the 8 configurations ($2 \times 2 \times 2$ possible combinations of the three reflection signs $s_j = \{0, 1\}, j = 1, 2, 3$) is the most likely to hold according to an error measure between extracted contours and ray-traced notch tracks.

First, in order to perform ray tracing for each configuration $\boldsymbol{s} = [s_1, s_2, s_3]$ the focus needs to be known. Unfortunately, no documentation on the exact microphone position is provided with the CIPIC database; hence, in order to avoid blind focus fixing, an optimization procedure is run pixelwise over a rectangular search area A of the pinna photograph covering the whole ear canal entrance. Also, a rotation tolerance $\rho \in I = [-5°, 5°]$ at 1-degree steps is considered. More in detail, for each track T_j the corresponding notch frequencies $f_0^j(\phi), j = \{1, 2, 3\}$, are first translated into Euclidean distances (in pixels) through a sign-dependent combination of (3) and (5),

$$d_c^j(\phi) = \frac{c}{2(s_j + 1) f_0^j(\phi)}, \tag{6}$$

and subsequently projected onto the point

$$p_{\psi,\rho}^j(\phi) = (\psi_x + d_c^j(\phi)\cos(\phi + \rho), \psi_y + d_c^j(\phi)\sin(\phi + \rho)) \tag{7}$$

on the pinna image. The optimal focus and rotation of the configuration, $(\psi_{\mathbf{s}}^{\mathrm{opt}}, \rho_{\mathbf{s}}^{\mathrm{opt}})$, are then defined as those satisfying the following minimization problem:

$$\min_{\psi \in A, \rho \in I} \sum_{j=1}^{3} \min_i d_{\psi,\rho}(T_j, C_i)^2, \tag{8}$$

where $d_{\psi,\rho}(T_j, C_i)$ is the distance between track T_j and contour C_i, which is defined as the average of distances $d(p_{\psi,\rho}^j(\phi), C_i)$ across all the track points.

Having fixed the eight optimal foci and rotations, one per configuration, we now use a simple scoring function to indicate the *fitness* of each configuration. This is defined as

$$F(\boldsymbol{s}) = \frac{1}{3} \sum_{j=1}^{3} \min_i \frac{d_{\psi_{\mathbf{s}}^{\mathrm{opt}}, \rho_{\mathbf{s}}^{\mathrm{opt}}}(T_j, C_i)}{2 - s_j}, \tag{9}$$

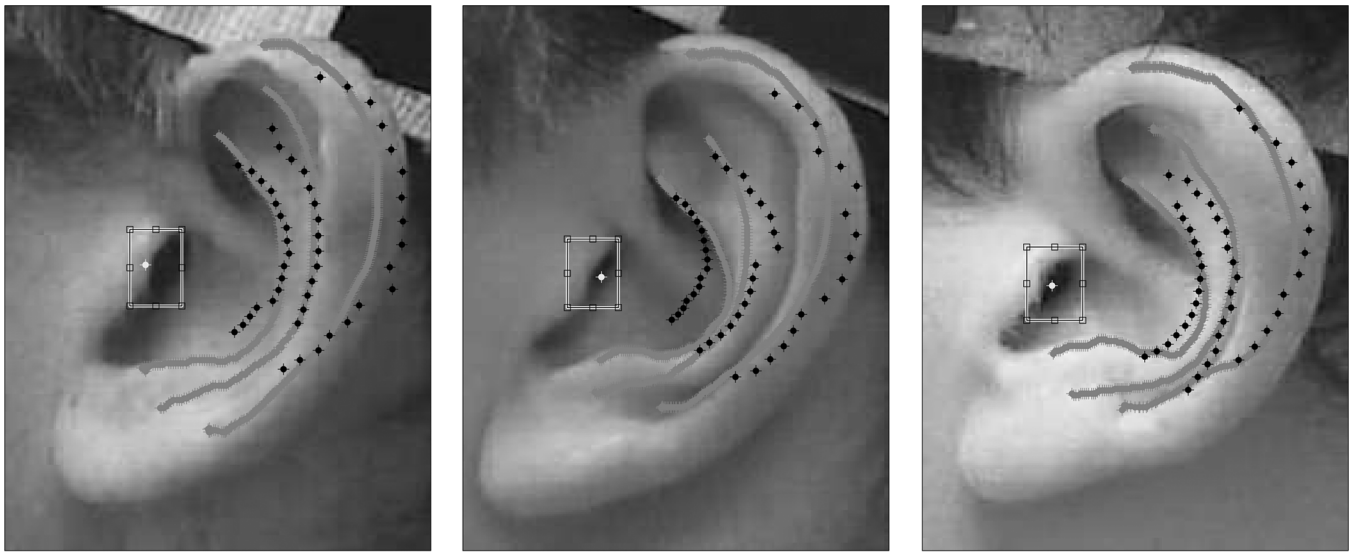
that is, the mean of all the (linear) distances between each ray-traced track $T_j, j = 1, 2, 3$, and its nearest contour $C_i, i = 1, \ldots, 5$. Note that the innermost quantity in (9) is scaled by a factor of 1/2 if the reflection sign is negative; this factor takes into account the halved resolution of the ray-traced negative reflection with respect to a positive reflection. Clearly, the smaller the fitness value, the better the fit.

B. Results

The above contour matching procedure was run for all our 20 CIPIC subjects. Table I summarizes the final scores (fitness values) for all possible configurations, along with the resulting "best" configuration $\boldsymbol{s}^{\mathrm{opt}}$ and the corresponding best matching contours. For subjects with two tracks only we conventionally label the missing track's reflection sign with "$*$". As an example, Fig. 5 shows the optimal ray-traced tracks for three subjects: 027 (having a final score close to the median), 050 (second worst subject), and 134 (third best subject).

We can immediately notice that configuration $\boldsymbol{s} = [0, 0, 0]$, i.e., negative coefficient sign for all reflections, obtains the best score in all cases except for Subject 015. However, we noticed that for both this subject and Subject 009 the optimal focus of the winning configuration is located well outside the ear canal area, even when the search area A is widened. Closer inspection of the corresponding pinna pictures revealed that they were taken from an angle which is far from being approximately aligned to the interaural axis, resulting in focus points much displaced towards the back of the head. As an effect, the pinna image is stretched with respect to all other cases. Consequently, as no consistent matching can be defined on these two pinna pictures, in the following we regard Subject 009 and Subject 015 as outliers.

All the remaining subjects exhibit $\boldsymbol{s}^{\mathrm{opt}} = [0, 0, 0]$ as the winning configuration. Quantitative correspondence between tracks and contours varies from subject to subject, e.g., assigning a much lower score to Subject 165 with respect to Subject 003; still, scores were defined as above with the aim to give an indication of the probability of a configuration for a series of subjects rather than an intersubjective fitness measure. Interestingly, in all cases except one, scores for $\boldsymbol{s} = [1, 1, 1]$ are more than doubled with respect to the complementary configuration



(a) (b) (c)

Fig. 5. Optimal ray-tracing for three subjects. The light grey point surrounded by the search area A is the optimal focus of the winning configuration $\boldsymbol{s}^{\mathrm{opt}} = [0, 0, 0]$. Black points indicate the three projected tracks, and dark grey points the hand-traced nearest contours to the tracks. (a) Subject 027, (b) Subject 050, (c) Subject 134.

$\boldsymbol{s} = [0, 0, 0]$, a result which catalogues the hypothesis of an overall positive reflection sign as unlikely. Also, note that the second best configuration is generally $\boldsymbol{s} = [1, 0, 0]$. Moreover, tracks T_2 and T_3 always best match with C_4 and C_3, respectively, while T_1 matches best with C_1 in 47% of subjects and with C_2 in 53% of subjects. These results enforce the hypothesis of negative reflection sign for T_2 and T_3 while leaving a halo of uncertainty on T_1's actual reflection sign.

Nevertheless, the optimality of $\boldsymbol{s}^{\mathrm{opt}} = [0, 0, 0]$ is further supported by the following observations. First, if $s_1 = 1, T_1$ would fall near to contour C_3 just like T_3 (see e.g., Fig. 5 for graphical evidence), hence the hypothesis of two different signs for reflections onto the same surface seems unlikely. Second, as mentioned in Section III.B T_1 terminates on average earlier than T_2 and T_3. This indicates that for elevations approaching $\phi = 45°$ the incoming wave hardly finds a perpendicular reflection surface, and this is compatible with a reflection on the helix, which normally ends just below the eye level. Last but not least, if $s_1 = 0, T_1$ falls near C_2 for all those subjects having a protruding ear; this would mean that reflections are most likely to happen on the wide helix wall rather than the border C_1, which conversely is the significant reflector for subjects with a narrow helix.

Another quantitative result that deserves to be commented is the score per track, averaged on the 18 "good" subjects: 2.37 for T_1, 1.84 for T_2, and 2.57 for T_3. Surprisingly, the best score is obtained for C_4, which was harder to trace in the preprocessing phase. By contrast, one of the clearest contours, C_3, is also the one that exhibits the greatest mismatch with respect to its relative track. This is mainly due to a number of track points around elevation $\phi = 0°$ being projected nearer to the ear canal than C_3 on the pinna image, a common trend that is observed in 11 subjects over 18 and is clearly detectable in the three cases depicted in Fig. 5, Subject 050 showing the greatest mismatch. This point is further discussed next.

C. Discussion

The above results numerically give credit to Satarzadeh's negative reflection hypothesis. Three main notches apparently due to three different reflections on the concha border, antihelix/concha wall, and helix are seen in most HRTFs. One may think of the pinna seen from the median plane as a sequence of three protruding borders: concha border, antihelix, and helix border. These are regarded by Satarzadeh as boundaries between skin and air, that in a mechanical wave transmission analogy would introduce an impedance discontinuity $Z_1/Z_2 < 1$ at the reflection point [48]. Thus, a part of the wave would follow a straight path while another with diminished amplitude and inverted phase would be reflected back to the ear canal. Despite the clever intuition, there is no evidence of the fact that waves are only reflected at borders and not onto inner pinna walls.

A recent study by Takemoto *et al.* on pressure distribution patterns in median-plane PRTFs [49] reveals through FDTD simulations on four different subjects' pinnae the existence of vast negative pressure anti-nodes inside pinna cavities at the first notch frequency. Specifically, when the source is below the horizontal plane the cymba, triangular fossa, and scaphoid fossa resonate in the same phase which is reverse to that of the incoming wave, while when the source is placed in the anterosuperior direction the same phenomenon appears at the back of the concha. The authors then observe that these negative pressure zones cancel the wave and, as a consequence, a pressure node appears at the ear canal entrance. Thus, we can speculate about the following generation mechanism for notches in track T_1, all of which we refer to as N_1: a given frequency component of the incoming sound wave forms a negative pressure area in the vicinity of the helix wall or border, reflects back with inverted phase, and encounters the direct wave at the ear canal entrance after a full period delay canceling that frequency component.

TABLE I
CONTOUR MATCHING PROCEDURE RESULTS

Subject	$F(0,0,0)$	$F(0,0,1)$	$F(0,1,0)$	$F(0,1,1)$	$F(1,0,0)$	$F(1,0,1)$	$F(1,1,0)$	$F(1,1,1)$	$\boldsymbol{s}^{\text{opt}}$	Nearest contours
003	4.03	9.19	9.27	13.78	7.83	12.45	13.03	17.54	$[0,0,0]$	$1,4,3$
008	2.95	4.86	5.33	7.30	3.69	7.89	5.58	10.64	$[0,0,0]$	$1,4,3$
009	2.55	5.18	4.79	7.02	2.95	5.08	2.94	5.01	$[0,0,0]$	$2,4,4$
010	1.88	5.18	2.26	6.02	3.57	5.69	4.46	6.70	$[0,0,0]$	$1,4,3$
011	2.62	5.10	5.60	9.53	3.16	5.79	4.97	9.25	$[0,0,0]$	$1,4,3$
012	2.08	4.21	4.76	7.30	2.70	5.32	3.20	6.78	$[0,0,0]$	$2,4,3$
015	4.99	9.92	6.14	10.59	3.02	6.70	3.39	3.19	$[1,0,0]$	$3,1,4$
017	2.81	6.35	4.53	8.12	2.99	5.02	5.63	6.79	$[0,0,0]$	$1,4,3$
019	1.64	6.64	4.85	8.00	1.64	6.64	4.85	8.00	$[*,0,0]$	$-,4,3$
020	1.15	1.15	5.27	5.27	1.85	1.85	5.45	5.45	$[0,0,*]$	$2,4,-$
021	2.90	6.40	4.06	8.44	3.30	8.97	6.25	11.54	$[0,0,0]$	$2,4,3$
027	2.07	6.53	5.04	8.56	2.32	5.27	2.80	4.25	$[0,0,0]$	$2,4,3$
028	1.71	3.54	4.21	5.57	3.79	4.02	5.62	6.10	$[0,0,0]$	$2,4,3$
033	2.51	4.73	6.66	6.61	3.42	7.68	9.08	9.98	$[0,0,0]$	$1,4,3$
040	1.74	5.48	2.59	5.35	2.57	5.86	3.30	5.96	$[0,0,0]$	$1,4,3$
044	1.88	2.84	5.33	4.81	2.86	2.49	4.13	3.74	$[0,0,0]$	$2,4,3$
048	2.02	5.33	5.45	7.86	3.70	5.06	5.27	6.97	$[0,0,0]$	$1,4,3$
050	3.25	6.29	7.68	10.52	4.37	7.59	7.57	11.23	$[0,0,0]$	$2,4,3$
134	1.64	6.11	5.18	8.56	3.38	6.31	4.56	7.37	$[0,0,0]$	$2,4,3$
165	1.09	5.35	3.08	5.93	3.43	3.89	3.00	2.99	$[0,0,0]$	$2,4,3$

Unfortunately, similar pressure distribution patterns for notches in T_2 and T_3 (respectively N_2 and N_3) have not been studied in [49]; still we can think of analogous generation mechanisms for these tracks too.

Shifting our focus to actual pinna contours that are responsible for spectral notches, one further clue confirms contour C_3 as most likely associated to track T_3. The observed "anticipation" of contour C_3 exhibited by T_3 at elevations close to $\phi = 0°$ (see Fig. 5) may be regarded as a delay that affects the direct wave alone due to diffraction across the tragus. Evidence of this phenomenon is also conjectured in [12]. Concerning track T_1, our findings seem to conflict with the common idea that N_1 is due to a reflection on the concha wall [16], [20], [27]. In two works by Mokhtari *et al.* [12], [31], micro-perturbations to pinna surface geometry in the form of 2-mm voxels are introduced at each possible point on a simulated KEMAR pinna. The authors observe that perturbations across the whole area of the pinna, helix included, introduce positive or negative shifts in the center frequency of N_1, especially at elevations between $\phi = -45°$ and $\phi = 0°$ in the median plane. Such shifts do not appear if voxels are introduced over the helix area in higher order notches, whose center frequency sensitively varies for perturbations introduced within the concha, cymba and triangular fossa only. This result clearly indicates that the reflection path responsible for N_1 crosses the whole pinna area, calling into question the above common belief and giving credit to our result instead.

Admittedly, as [12] points out, the last result also suggests that ray-tracing models are based on a wrong assumption, i.e., that a single path is responsible for a notch. The dependence of N_1 on the whole pinna surface clearly indicates that multiple reflection paths concur in determining the distinctive parameters of the notch. However, even if multiple paths are responsible for the exact frequency location of the notch, thanks to the concave shape of the considered contours one may think of a specific time delay for which the greatest portion of reflections counteract the direct wave as an approximation to a single, direct ray.

Another objectionable point of our approach is the adequateness of using a 2-D representation for contour extraction. As a matter of fact, since in most cases the pinna structure does not lie on a parallel plane with respect to the head's median plane, especially in subjects with protruding ears, a 3-D model of the pinna would allow to investigate its horizontal section. Beside the unavailability of such kind of reconstruction for the considered subjects, our original aim was to keep the contour extraction procedure as low-cost and accessible as possible; furthermore, additional results in the following Section will confirm that the 2-D approximation is, on a theoretical basis at least, already satisfactory.

To conclude this short discussion, it should be emphasized that the results of the ray-tracing analysis do not conclusively prove that negative reflections effectively occur in reality. In particular, it remains to be explained from an acoustical point of view why negative reflection coefficients are likely to be produced. Clearly, a negative reflection coefficient c_r will not have unitary magnitude in real conditions because of the soft reflective surfaces involved, hence it will always satisfy $-1 < c_r < 0$. This results in a partial cancellation of the frequency where the notch falls: the closer the reflection coefficient to -1 is, the deeper the corresponding frequency notch will be. In order to characterize the magnitude of the coefficient, it could be therefore worthy to study how notch depths change with elevation.[6]

V. THE STRUCTURAL MODEL

In this Section we propose an extension of Satarzadeh's structural filter model, which includes contributions by the head and pinna into two separate structures. In light of the previously discussed invariance of PRTFs to azimuth up to 30° from the median plane we introduce a fundamental assumption, i.e., elevation and azimuth cues are handled orthogonally throughout the considered frontal workspace (see again Fig. 1). Vertical control

[6]Unfortunately, common HRTF recordings do not have a frequency resolution that allows detection of the exact local minimum characterizing a notch, i.e., notch depth is always underestimated. A previous work of the authors did not reveal clear trends for notch depth (when considering the frequency resolution of the CIPIC database) except for a known general decrease with increasing elevation [45].

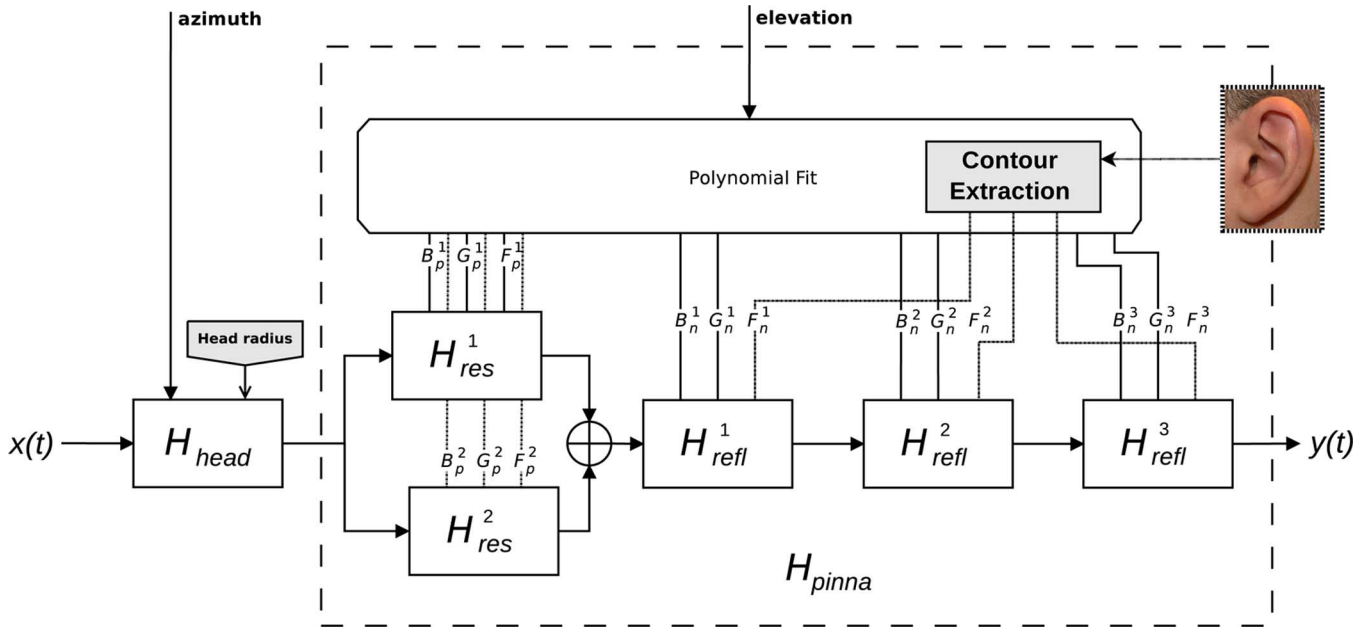


Fig. 6. The structural HRTF model. Customization is performed through parameter extraction from anthropometric measurements and a pinna picture.

is associated with the acoustic effects of the pinna while the horizontal one is delegated to head diffraction. No modeling for the shoulders and torso is considered, even though their presence would generally add low-frequency secondary HRTF cues for elevation perception [42]. Furthermore, dependence on source distance is negligible in the pinna model but critical in the head counterpart in the near field [43]: since the current head model does not integrate such dependence, the overall structure is assumed to be valid only for sources at 1 m from the center of the head or farther. Two instances (one per ear) of such model, appropriately synchronized through interaural time delay (ITD) estimation methods, allow for real-time binaural rendering.

A. Filter Model

Fig. 6 reports a global view of the model. From left to right, the first block is the head model. Different possible existing models can be exploited here; in order to keep the overall structure as computationally efficient as possible, we choose to use the digital counterpart of the single-pole, single-zero minimum-phase analog filter that approximates head shadowing described in [8], obtained through the bilinear transform:

$$H_{\text{head}}(z) = \frac{\frac{\beta+\alpha f_s}{\beta+f_s} + \frac{\beta-\alpha f_s}{\beta+f_s} z^{-1}}{1 + \frac{\beta-f_s}{\beta+f_s} z^{-1}}, \tag{10}$$

where f_s is the sampling frequency, β depends on the head radius parameter a as $\beta = c/a$, and α is defined as in [8],

$$\alpha(\theta_{\text{inc}}) = 1 + \frac{\alpha_{\min}}{2} + \left(1 - \frac{\alpha_{\min}}{2}\right) \cos\left(\frac{\theta_{\text{inc}}}{\theta_{\min}}\pi\right). \tag{11}$$

θ_{inc} is the incidence angle that, assuming the interaural axis to coincide with the x axis for sake of brevity, relates to azimuth θ as $\theta_{\text{inc}} = 90° - \theta$ for the right ear and $\theta_{\text{inc}} = 90° + \theta$ for the left ear. A reasonably good approximation of real diffraction curves in our range of interest for the azimuth angle $-30° < \theta < 30°$ is heuristically found for parameters $\alpha_{\min} = 0.1$ and $\theta_{\min} = 180°$. Furthermore, the head radius parameter a, whose value influences the cutoff frequency for the head shadowing, is defined by a weighted sum of the subject's head dimensions using the optimal weights obtained in [50] through a regression on the CIPIC subjects' anthropometric data.

Coming to the pinna block, the only independent parameter used here is source elevation ϕ, which drives the evolution of resonances' center frequency $F^i_p(\phi)$, 3 dB bandwidth $B^i_p(\phi)$, and gain $G^i_p(\phi), i = 1, 2$, and of the corresponding notch parameters $(F^j_n(\phi), B^j_n(\phi), G^j_n(\phi), j = 1, 2, 3)$. For each subject, these parameters are derived as follows. First, they are straightforwardly estimated from the separated resonant or reflective (i.e., notch tracks) component of median-plane PRTFs for all the available ϕ values.[7] Second, a fifth order polynomial $\mathcal{P}^i_p$ or $\mathcal{P}^j_n$, where $\mathcal{P} \in \{F, B, G\}$, is best fitted to the corresponding sequence of parameter values, yielding a complete parametrization of the filters. Obviously, all the polynomials must be computed offline previous to the rendering process.

However, following our findings in the previous Section, functions $F^j_n(\phi)$ can alternatively be extracted from the subject's anthropometry (in the form of a pinna picture): contours C_2 or C_1 (depending on whether the subject's ear is respectively protruding or not), C_4, and C_3 are converted into distances with respect to the ear canal entrance, and then translated into sequences of frequencies through (5), thus assuming overall negative reflection coefficients. Again, a fifth order polynomial is best fitted to these sequences, resulting in functions $F^j_n(\phi), j = 1, 2, 3$. In the remainder of this Section we refer to HRTFs given by the fully resynthesized model (without contour extraction) as H^s, while HRTFs resulting from the contour-parameterized model as H^c.

[7]In order to avoid bad outcomes in the design of notch filters, gaps in notch tracks are assigned a gain equal to 0 dB while bandwidth and center frequency are given the value of the previous notch feature in the track.

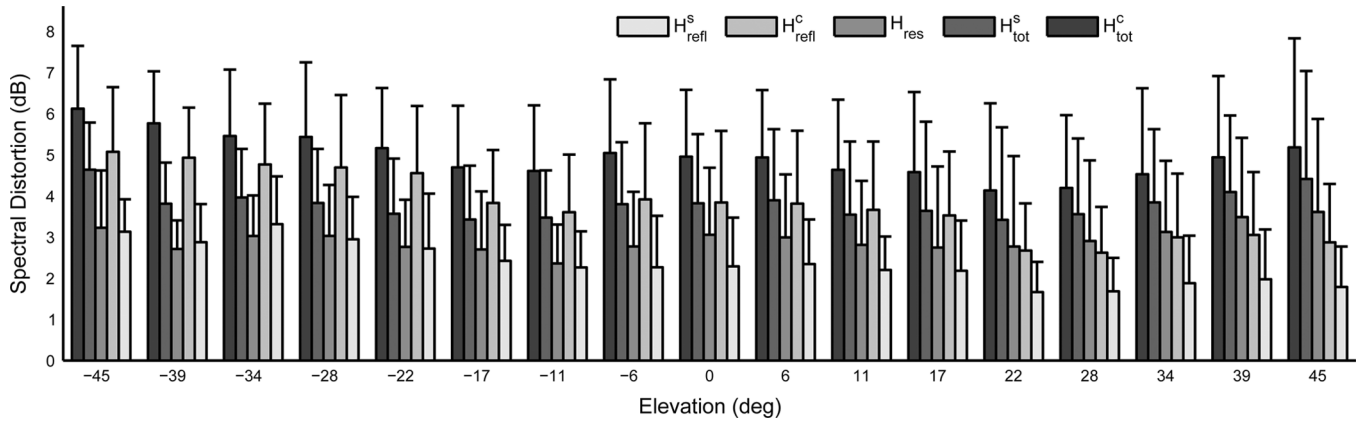


Fig. 7. Spectral distortion between reconstructed and measured median-plane HRTFs (mean and standard deviation over 18 CIPIC subjects).

The resonant part of the pinna model is represented as a parallel of two different second-order peak filters. The first peak $(i = 1)$ has the form [51]

$$H_{\rm res}^{(1)}(z) = \frac{1+(1+k)\frac{H_0}{2}+l(1-k)z^{-1}+(-k-(1+k)\frac{H_0}{2})z^{-2}}{1+l(1-k)z^{-1}-kz^{-2}}, \tag{12}$$

where

$$k = \frac{\tan\left(\pi\frac{B_p^1(\phi)}{f_s}\right)-1}{\tan\left(\pi\frac{B_p^1(\phi)}{f_s}\right)+1}, \quad l = -\cos\left(2\pi\frac{F_p^1(\phi)}{f_s}\right), \tag{13}$$

$$V_0 = 10^{\frac{G_p^1(\phi)}{20}}, \quad H_0 = V_0 - 1, \tag{14}$$

and f_s is the sampling frequency. The second peak $(i = 2)$ is implemented as in [52],

$$H_{\rm res}^{(2)}(z) = \frac{V_0(1-h)(1-z^{-2})}{1+2lhz^{-1}+(2h-1)z^{-2}}, \tag{15}$$

$$h = \frac{1}{1+\tan\left(\pi\frac{B_p^2(\phi)}{f_s}\right)}, \tag{16}$$

while l and V_0 are defined as in (13) and (14) with polynomial index $i = 2$. The reason for this distinction lies in the low-frequency behavior we need to model: the former implementation has unitary gain at low frequencies so as to preserve such characteristic in the parallel filter structure, while the latter has a negative dB magnitude in the same frequency range. In this way, the all-round pinna filter does not alter low-frequency components in the signal forwarded by the head shadow filter.

The notch filter implementation is of the same form as peak filter $H_{\rm res}^{(1)}$ with the only differences in the parameters' description. In order to keep notation correct, polynomials $\mathcal{P}_p^1$ must be substituted by the corresponding notch counterparts $\mathcal{P}_n^j, j = 1, 2, 3$, and parameter k defined in (13) replaced by its "cut" version

$$k = \frac{\tan\left(\pi\frac{B_n^j(\phi)}{f_s}\right)-V_0}{\tan\left(\pi\frac{B_n^j(\phi)}{f_s}\right)+V_0}. \tag{17}$$

Example plots of PRTF resynthesis with similar filter structures can be found in [44].

B. Results and Discussion

In order to objectively evaluate the model against the original measured HRTFs in the CIPIC database we consider an error measure widely used in recent literature [37], [53], [54], i.e., spectral distortion:

$$\mathrm{SD} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(20\log_{10}\frac{|H(f_i)|}{|\tilde{H}(f_i)|}\right)^2}\ [\mathrm{dB}], \tag{18}$$

where H is the original response, $\tilde{H}$ is the reconstructed response, and N is the number of available frequencies in the considered range, that we limit between 500 Hz and 16 kHz.

Fig. 7 reports SD values, averaged across the 18 non-outlier CIPIC subjects, of five different median-plane reconstructed responses:

1) the all-round response of the contour-parameterized model, $H_{\rm tot}^c$;
2) the reflective component of the contour-parameterized model given by notch filters, $H_{\rm refl}^c$;
3) the resonant component of the model (either contour-parameterized or resynthesized) given by peak filters, $H_{\rm res}$;
4) the all-round response of the fully resynthesized model, $H_{\rm tot}^s$;
5) the reflective component of the fully resynthesized model given by notch filters, $H_{\rm refl}^s$.

Resonant and reflective components are obviously compared to their counterparts extracted by the separation algorithm.

As expected, $H_{\rm tot}^c$ is the response with the highest average SD. As a matter of fact, errors in the resynthesized resonant $(H_{\rm res})$ and contour-parameterized reflective $(H_{\rm refl}^c)$ components combine together yielding the SD for $H_{\rm tot}^c$, which ranges from 4 to 6 dB on average and is worse at low elevations. This fact can be explained by the occurrence of very deep notches at low elevations, that causes large errors in the SD when a notch extracted from a contour is not perfectly reconstructed at its proper frequency.

In proof of this note that, as notches become fainter and fainter with increasing elevation, the mean SD of $H_{\rm tot}^c$ tends to decrease apart from a new rise at the last elevation angles, which is conversely due to greater errors in the resonant component $H_{\rm res}$. An informal inspection of resonant components at higher elevations revealed indeed that the second modeled

TABLE II
NOTCH FREQUENCY MISMATCH BETWEEN TRACKS AND CONTOURS

Subject	$m(T_1, C_1)$	$m(T_1, C_2)$	$m(T_2, C_4)$	$m(T_3, C_3)$
003	11.42%	–	12.02%	18.25%
008	8.98%	–	8.69%	14.07%
010	4.80%	–	2.90%	18.74%
011	8.75%	–	7.77%	12.20%
012	–	5.57%	8.98%	8.69%
017	7.80%	–	3.44%	17.97%
019	–	–	4.48%	5.92%
020	–	5.50%	4.27%	–
021	–	9.18%	10.16%	11.73%
027	–	8.14%	2.09%	7.63%
028	–	7.39%	8.05%	14.79%
033	4.52%	–	3.55%	16.44%
040	2.98%	–	5.50%	12.92%
044	–	9.63%	6.49%	8.10%
048	4.01%	–	3.18%	16.19%
050	–	8.62%	7.28%	18.95%
134	–	2.59%	5.10%	10.13%
165	–	3.91%	4.11%	6.44%

high-frequency peak (horizontal mode) disappears, gradually letting non-modeled lower-frequency vertical modes in. The appearance of such modes also brings a significant rise of the SD variance in the all-round responses at the highest elevation angles.

As a further confirmation of the criticality of the exact notch frequency location in SD computation, note that when frequencies are extracted from real HRTFs the SD of the reflective component H^s_{refl} distinctly decreases both in mean (3 dB or less) and variance, resulting in a noticeably lower average SD (about 4 dB) in the total response H^s_{tot}.

We now introduce another error measure to show that, even if contour-extracted notch frequencies do not exactly correspond to their measured counterparts, the effective frequency shift is almost everywhere not likely to result in a perceptual difference. Specifically, we define the *mismatch* between a computed notch track T_j and its associated contour C_i as the percentual ratio between the aforementioned frequency shift and the measured notch frequency, averaged on all the elevations where the notch is present:

$$m(T_j, C_i) = \frac{1}{n(T_j)} \sum_{\phi} \frac{|f_0^j(\phi) - F_n^j(\phi)|}{f_0^j(\phi)} \cdot 100\%, \qquad (19)$$

where $n(T_j)$ is the number of available notch frequencies in track T_j and $F_n^j(\phi)$ is extracted from the associated contour C_i as described in Section V.A.

Table II shows frequency mismatches computed for the usual 18 CIPIC subjects. We can directly compare these results to the findings by Moore *et al.* included in Experiment V in [21]: two steady notches in the high-frequency range (around 8 kHz) differing just in central frequency are not distinguishable on average if the mismatch is less than approximately 9%, regardless of notch bandwidth. Although these results were found for just one high-frequency location, we may informally compare mismatches of T_1 and T_2 with the 9%-threshold and conclude that only 5 tracks over 35 exhibit a mismatch greater than the threshold, suggesting that the frequency shift caused by contour extraction is not perceptually relevant on average.

Conversely, track T_3 shows much greater mismatches, mostly due to the "contour anticipation" effect discussed in Subsection IV.C. Beside possible improvements that may take into account such an effect while extracting contour C_3 and lower the mismatch, no results are available in the literature about notch perception in the region between 10 and 15 kHz. However, as already mentioned in Section II, the third notch is of lesser importance than the first two in elevation perception [32], hence psychoacoustical criticality of its center frequency is somehow questionable.

As a conclusion to the presented results, if we assume that the aforementioned mismatches are in most cases not perceptually relevant, we can then consider the mean SD of 4 dB in H^s_{tot} as a satisfactory result, being comparable to SD values found in similar works that deal with HRTF resynthesis by means of HRIR decomposition [37] or anthropometric parametrization through multiple regression analysis on HRTF decomposition [53]. What's more, our model is composed of first- and second-order filters only: given that many responses exhibit sharp notches whose slope cannot be reached by a second-order filter, increasing the order of notch filters in particular would further improve the SD score. However, low-order filters allow cheap and fast real-time simulation, which is a valuable merit of the model.

The model as it was proposed in this paper represents a notable extension of the one in [17] as it includes a large portion of the frontal hemispace, and could thus be suitable for real-time control of virtual sources in a number of applications involving frontal auditory displays, such as a sonified screen [55]. Further extensions of the model, such as to include source positions behind, above, and below the listener, may be obtained in different ways.

The HRTF database used in this study does not include elevation data below $-45°$. Alternative HRTF data sets or BEM simulations should be considered in order to extend the ray tracing procedure to the range $-90° < \phi < -45°$. It ought to be noted that in this range the inclusion of the shoulders and torso's contribution becomes crucial, adding relevant shadowing effects to the incoming waves [56]. Thus, it should be verified whether a model of the torso can effectively compensate for the lack of a model for reflections due to the pinna at very low elevations, not forgetting that low-elevation HRTFs are usually heavily influenced by posture [56].

Concerning source positions above the listener, the attenuation of frequency notches with increasing elevation observed in the literature [16], [44] and directly in HRTF sets suggests that notches could simply be gradually extinguished starting from $\phi = 45°$ up to $\phi = 90°$ while keeping their central frequency fixed. However, particular care should be reserved to the modeling of resonances in this elevation range, where the second peak generally disappears in favour of a broader first peak [44]. Finally, the role of notches for posterior sources is not completely understood in current literature, although a regular presence of spectral notches has been observed in posterior HRTFs too [11]. An assessment of the applicability of the ray tracing procedure to this elevation range is therefore left for future work.

VI. Conclusion

In this paper we performed an analysis of real HRTF data in order to study the relation between HRTF features and anthropometry in the frontal median plane. Our findings support the hypothesis that reflections occurring on pinna surfaces can be reduced for the sake of design to three main contributions, each carrying a negative reflection coefficient. Based on this observation an approach to HRTF customization, mainly based on structural modeling of the pinna contribution, was proposed. Spectral distortion and notch frequency mismatch measures indicate that our approximation is objectively satisfactory.

Beside subjective evaluations of the model, which were outside the scope of this paper and will need new HRTF measurements as well as model reconstruction onto a number of physical subjects, ongoing and future work includes automatic pinna contour extraction and extension of the model to a wider spatial range, including the upper and back side of the sagittal plane. Understanding the influence of notch depth and bandwidth in elevation perception along with the relation between the resonant component of the PRTF and the shape of pinna cavities is also required to have a complete anthropometric parametrization of the pinna model. Last but not least, an extension of the head model that includes near-field dependence on source distance is currently being studied in order to allow a complete representation of the auditory scene surrounding the listener.

Acknowledgment

The authors would like to thank Professor Ralph V. Algazi for the kind provision of the 20 pinna pictures of CIPIC HRTF database subjects.

References

[1] D. R. Begault, *3-D Sound for Virtual Reality and Multimedia*. San Diego, CA: Academic, 1994.

[2] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *J. Audio Eng. Soc.*, vol. 52, no. 6, pp. 618–639, 2004.

[3] F. Avanzini and P. Crosato, "Integrating physically-based sound models in a multimodal rendering architecture," *Comp. Anim. Virtual Worlds*, vol. 17, no. 3–4, pp. 411–419, Jul. 2006.

[4] C. I. Cheng and G. H. Wakefield, "Introduction to head-related transfer functions (HRTFs): Representations of HRTFs in time, frequency, and space," *J. Audio Eng. Soc.*, vol. 49, no. 4, pp. 231–249, Apr. 2001.

[5] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 94, no. 1, pp. 111–123, 1993.

[6] H. Møller, M. F. Sørensen, C. B. Jensen, and D. Hammershøi, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc.*, vol. 44, no. 6, pp. 451–469, 1996.

[7] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds," *J. Acoust. Soc. Amer.*, vol. 42, no. 2, pp. 480–488, Aug. 1967.

[8] C. P. Brown and R. O. Duda, "A structural model for binaural sound synthesis," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 5, pp. 476–488, Sep. 1998.

[9] D. W. Batteau, "The role of the pinna in human localization," in *Proc. R. Soc. London. Ser. B, Biol. Sci.*, Aug. 1967, vol. 168, no. 1011, pp. 158–180.

[10] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Amer.*, vol. 110, no. 5, pp. 2440–2448, Nov. 2001.

[11] Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," *J. Sound Vibr.*, vol. 300, no. 3–5, pp. 552–579, 2007.

[12] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Pinna sensitivity patterns reveal reflecting and diffracting surfaces that generate the first spectral notch in the front median plane," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP '11)*, Prague, Czech Republic, May 2011.

[13] D. J. Kistler and F. L. Wightman, "A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction," *J. Acoust. Soc. Amer.*, vol. 91, no. 3, pp. 1637–1647, 1992.

[14] S. Hwang, Y. Park, and Y. Park, "Modeling and customization of head-related impulse responses based on general basis functions in time domain," *Acta Acustica United With Acustica*, vol. 94, no. 6, pp. 965–980, Nov. 2008.

[15] M. J. Evans, J. A. S. Angus, and A. I. Tew, "Analyzing head-related transfer function measurements using surface spherical harmonics," *J. Acoust. Soc. Amer.*, vol. 104, no. 4, pp. 2400–2411, Oct. 1998.

[16] V. C. Raykar, R. Duraiswami, and B. Yegnanarayana, "Extracting the frequencies of the pinna spectral notches in measured head related impulse responses," *J. Acoust. Soc. Amer.*, vol. 118, no. 1, pp. 364–374, July 2005.

[17] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Conv. Audio Eng. Soc.*, Vienna, Austria, May 5–8, 2007.

[18] A. Wilska, "Studies on directional hearing," Ph.D. dissertation, Aalto Univ. School of Sci. and Technol., Dept. of Signal Process. and Acoust., Univ. of Helsinki, Helsinki, Finland, 2010, 1938, English translation, originally published in German as Untersuchungen über das Richtungshören.

[19] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1983.

[20] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Amer.*, vol. 56, no. 6, pp. 1829–1834, Dec. 1974.

[21] B. C. J. Moore, S. R. Oldfield, and G. J. Dooley, "Detection and discrimination of spectral peaks and notches at 1 and 8 kHz," *J. Acoust. Soc. Amer.*, vol. 85, no. 2, pp. 820–836, Feb. 1989.

[22] F. Asano, Y. Suzuki, and T. Sone, "Role of spectral cues in median plane localization," *J. Acoust. Soc. Amer.*, vol. 88, no. 1, pp. 159–168, July 1990.

[23] M. B. Gardner and R. S. Gardner, "Problem of localization in the median plane: Effect of pinnae cavity occlusion," *J. Acoust. Soc. Amer.*, vol. 53, no. 2, pp. 400–408, 1973.

[24] M. Morimoto, "The contribution of two ears to the perception of vertical angle in sagittal planes," *J. Acoust. Soc. Amer.*, vol. 109, no. 4, pp. 1596–1603, Apr. 2001.

[25] J. Hebrank and D. Wright, "Are two ears necessary for localization of sound sources on the median plane?," *J. Acoust. Soc. Amer.*, vol. 56, no. 3, pp. 935–938, Sep. 1974.

[26] E. A. G. Shaw and R. Teranishi, "Sound pressure generated in an external-ear replica and real human ears by a nearby point source," *J. Acoust. Soc. Amer.*, vol. 44, no. 1, pp. 240–249, 1968.

[27] E. A. Lopez-Poveda and R. Meddis, "A physical model of sound diffraction and reflections in the human concha," *J. Acoust. Soc. Amer.*, vol. 100, no. 5, pp. 3248–3259, Nov. 1996.

[28] D. S. Brungart and W. M. Rabinowitz, "Auditory localization of nearby sources. Head-related transfer functions," *J. Acoust. Soc. Amer.*, vol. 106, no. 3, pp. 1465–1479, Sep. 1999.

[29] D. Wright, J. H. Hebrank, and B. Wilson, "Pinna reflections as cues for localization," *J. Acoust. Soc. Amer.*, vol. 56, no. 3, pp. 957–962, Sep. 1974.

[30] E. A. G. Shaw, "Acoustical features of human ear," in *Binaural and Spatial Hearing in Real and Virtual Environments*, R. H. Gilkey and T. R. Anderson, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 1997, pp. 25–47.

[31] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic sensitivity to micro-perturbations of KEMAR's pinna surface geometry," in *Proc. 20th Int. Congr. Acoust. (ICA '10)*, Sydney, Australia, Aug. 2010.

[32] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, "Median plane localization using a parametric model of the head-related transfer function based on spectral cues," *Appl. Acoust.*, vol. 68, pp. 835–850, 2007.

[33] R. Teranishi and E. A. G. Shaw, "External-ear acoustic models with simple geometry," *J. Acoust. Soc. Amer.*, vol. 44, no. 1, pp. 257–263, 1968.

[34] E. A. G. Shaw, "The acoustics of the external ear," in *Acoustical Factors Affecting Hearing Aid Performance*, G. A. Studebaker and I. Hochberg, Eds. Baltimore, MD: Univ. Park Press, 1980.

[35] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "A simple pinna model for generating head-related transfer functions in the median plane," in *Proc. 20th Int. Congr. Acoust. (ICA '10)*, Sydney, Australia, Aug. 2010.
[36] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Amer.*, vol. 63, no. 4, pp. 1152–1165, Apr. 1978.
[37] K. J. Faller, II, A. Barreto, and M. Adjouadi, "Augmented Hankel total least-squares decomposition of head-related transfer functions," *J. Audio Eng. Soc.*, vol. 58, no. 1/2, pp. 3–21, Jan./Feb. 2010.
[38] R. V. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *Proc. IEEE Workshop Appl. Signal Process. Audio, Acoust.*, New Paltz, NY, 2001, pp. 1–4.
[39] V. R. Algazi, 2010, private communications.
[40] S. Spagnol, M. Hiipakka, and V. Pulkki, "A single-azimuth pinna-related transfer function database," in *Proc. 14th Int. Conf. Digital Audio Effects (DAFx-11)*, Paris, France, Sep. 2011.
[41] P. Mokhtari, H. Takemoto, R. Nishimura, and H. Kato, "Acoustic simulation of KEMAR's HRTFs: Verification with measurements and the effects of modifying head shape and pinna concavity," in *Proc. Int. Work. Princ. Appl. Spatial Hearing (IWPASH)*, Zao, Miyagi, Japan, Nov. 2009.
[42] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *J. Acoust. Soc. Amer.*, vol. 109, no. 3, pp. 1110–1122, Mar. 2001.
[43] R. O. Duda and W. L. Martens, "Range dependence of the response of a spherical head model," *J. Acoust. Soc. Amer.*, vol. 104, no. 5, pp. 3048–3058, Nov. 1998.
[44] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and modeling of pinna-related transfer functions," in *Proc. 13th Int. Conf. Digital Audio Effects (DAFx-10)*, Graz, Austria, Sep. 2010.
[45] M. Geronazzo, S. Spagnol, and F. Avanzini, "A head-related transfer function model for real-time customized 3-D sound rendering," in *Proc. 7th Int. Conf. Signal Image Technol. and Internet-Based Syst. (SITIS '11)*, Dijon, France, Nov.–Dec. 2011, pp. 174–179.
[46] S. Spagnol, M. Geronazzo, and F. Avanzini, "Fitting pinna-related transfer functions to anthropometry for binaural sound rendering," in *Proc. IEEE Int. Workshop Multi. Signal Process.*, Saint-Malo, France, Oct. 2010, pp. 194–199.
[47] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 4, pp. 744–754, Aug. 1986.
[48] P. Satarzadeh, "A study of physical and circuit models of the human pinnae," M.S. thesis, Univ. of California Davis, Davis, 2006.
[49] H. Takemoto, P. Mokhtari, H. Kato, R. Nishimura, and K. Iida, "Pressure distribution patterns on the pinna at spectral peak and notch frequencies of head-related transfer functions in the median plane," in *Proc. Int. Work. Princ. Appl. Spatial Hearing (IWPASH)*, Zao, Miyagi, Japan, Nov. 2009.
[50] V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc.*, vol. 49, no. 6, pp. 472–479, 2001.
[51] U. Zölzer, *Digital Audio Effects*. New York, NY: Wiley, 2002.
[52] S. J. Orfanidis, *Introduction to Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[53] T. Nishino, N. Inoue, K. Takeda, and F. Itakura, "Estimation of HRTFs on the horizontal plane using physical features," *Acoust. Science Technol.*, vol. 68, pp. 897–908, 2007.
[54] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu, "Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 6, pp. 1124–1132, Aug. 2009.
[55] A. Walker and S. Brewster, "Spatial audio in small screen device displays," *Personal and Ubiquitous Computing*, vol. 4, pp. 144–154, 2000.
[56] V. R. Algazi, R. O. Duda, and D. M. Thompson, "The use of head-and-torso models for improved spatial sound synthesis," in *Proc. 113th Conv. Audio Eng. Soc.*, Los Angeles, CA, 2002.

Simone Spagnol received the B.S. degree in computer engineering in 2006 and the M.S. degree in computer engineering in 2008 from the University of Padova, Italy. He was Visiting Scholar at the Laboratory of Acoustics and Audio Signal Processing, Aalto University, Finland in 2010. He received the Ph.D. degree in information engineering (Curriculum in information and communication technology) at the University of Padova in April 2012. He is currently a Postdoctoral Fellow at Iuav University of Venice. His research interests include binaural sound localization and synthesis and sonic interaction design.

Michele Geronazzo received the B.S. degree in computer engineering in 2006 and the M.S. degree in computer engineering in 2009 from the University of Padova, Italy. He is currently working towards his Ph.D. degree in information engineering (Curriculum in information and communication technology) at the University of Padova, Italy, where he is with the Sound and Music Computing group. His research interests include binaural technologies and multimodal interaction in virtual environments.

Federico Avanzini received the Laurea degree (cum laude) in physics from the University of Milano, Italy, in 1997 and the Ph.D. degree in information engineering from the University of Padova, Italy, in 2001 with a research project on sound and voice synthesis by physical modeling. During his doctoral studies he also worked as a visiting researcher at the Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology. Since 2002 he has been with the Sound and Music Computing group at the Department of Information Engineering of the University of Padova, where he is currently Assistant Professor, teaching courses in computer science and sound and music computing. His main research interests are in the area of sound synthesis and processing, with particular focus on musical sound synthesis, nonspeech sound in multimodal interfaces, voice synthesis and analysis. He has authored more than 90 publications on peer-reviewed international journals and conferences. He has been key researcher in numerous European projects (FP5, FP6) and national projects, and PI of the EU project DREAM (Culture2007) and of industry-funded projects. He serves on a regular basis in conference program committees and editorial committees, and was the General Chair of the 2011 International Sound and Music Computing Conference. He is in the Board of Directors of the Center of Computational Sonology (CSC) of the University of Padova, and the Italian Music Informatics Association (AIMI). He is CTO of Bloop, spin-off company of the University of Padova.