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 recognizable 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

T HE 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.

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 (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

1More 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: \( H(\theta, \phi, \omega) = \frac{\Phi(\theta, \phi, \omega)}{\Phi_f(\omega)} \), where \( (\theta, \phi) \) indicates the angular position of the source relative to the listener, and \( \omega \) is the angular frequency.
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 to 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 (≈17°) for unfamiliar sound sources, 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). These results find accordance in a 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. 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 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.

2The 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.

3In 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.
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 sinusoidal (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.

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 $\theta = 0^\circ$) HRIRs, with the elevation angle $\phi$ varying from $\phi = -45^\circ$ to $\phi = 45^\circ$ at 5.625-degree steps (17 HRIRs per subject). We choose to consider the median plane because relative azimuthal variations up to at least $\Delta \theta = 30^\circ$ 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 ($\phi = 45^\circ$) was chosen because of the high degree of uncertainty in elevation judgement for sources at $\phi > 45^\circ$ [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 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, then the ear canal lies around $\theta = \pm 90^\circ$. 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 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 $\theta = 0^\circ$) HRIRs, with the elevation angle $\phi$ varying from $\phi = -45^\circ$ to $\phi = 45^\circ$ at 5.625-degree steps (17 HRIRs per subject). We choose to consider the median plane because relative azimuthal variations up to at least $\Delta \theta = 30^\circ$ 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 ($\phi = 45^\circ$) was chosen because of the high degree of uncertainty in elevation judgement for sources at $\phi > 45^\circ$ [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 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, then the ear canal lies around $\theta = \pm 90^\circ$. 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

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.

Since human ears typically lie slightly behind and below the x axis [42], the source-ear angular distance is certainly greater than $90^\circ$ for sources between $\phi = 0^\circ$ and $\phi = 45^\circ$ at least.
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 \(\phi = 11.25^\circ\), then grow until \(\phi = 28.125^\circ\), and decrease at higher elevations on average. Despite the significant variance in the central frequencies of the three tracks (\(T_3\) in particular), these 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 \(-45^\circ\) 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

\[
d_r(\phi) = \frac{ct_d(\phi)}{2}
\]

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 \( \pi \):

\[
f_n(\phi) = \frac{2n + 1}{2d_r(\phi)} = \frac{c(2n + 1)}{4d_r(\phi)}, \quad n = 0, 1, \ldots
\]

Hence the first notch falls at frequency

\[
f_0(\phi) = \frac{c}{4d_r(\phi)}.
\]

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

\[
f_n(\phi) = \frac{n + 1}{2d_r(\phi)} = \frac{c(n + 1)}{2d_r(\phi)}, \quad n = 0, 1, \ldots
\]

\[
f_0(\phi) = \frac{c}{2d_r(\phi)}.
\]

Note that since our extracted notch tracks are pairwise in non-harmonic 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 recognizable 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 \( d_r \) 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 \( d_r \) 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, \( C_5 \) was the hardest to recognize due to the lower 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;
- the rotation \( \rho \) 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 \( \mathbf{s} = [s_1, s_2, s_3] \) (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 \( s_j = \{0, 1, 2, 3\} \)) 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 \( s = [s_1, s_2, s_3] \) 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 \( \rho \in I = -5^\circ, 5^\circ \) at 1-degree steps is considered. More in detail, for each track \( T_j \) the corresponding notch frequencies \( f_j^\rho(\phi) \), \( j = \{1, 2, 3\} \), are first translated into Euclidean distances (in pixels) through a sign-dependent combination of (3) and (5),

\[
d_x^\rho(\phi) = \frac{c}{2(s_x + 1)f_x^\rho(\phi)},
\]

and subsequently projected onto the point

\[
p_x^\rho(\phi) = (\psi_x + d_x^\rho(\phi) \cos(\phi + \rho), \psi_y + d_y^\rho(\phi) \sin(\phi + \rho))
\]

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

\[
\min_{\rho \in I, \phi \in I} \sum_{j=1}^{3} \min_{i} d_{\rho,j}^i(T_j, C_i),
\]

where \( d_{\rho,j}^i(T_j, C_i) \) is the distance between track \( T_j \) and contour \( C_i \), which is defined as the average of distances \( d(p_x^\rho(\phi), C_i) \) 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

\[
F(s) = \frac{1}{3} \sum_{j=1}^{3} \min_{i} d_{\rho,j}^i(T_j, C_i)/2 - s_j,
\]

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 \( s_{opt} \) 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 \( s = [0, 0, 0] \), 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 $s^{\text{det}} = [0, 0, 0]$ 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 $s = [1, 1, 1]$ are more than doubled with respect to the complementary configuration $s = [0, 0, 0]$, a result which catalogues the hypothesis of an overall positive reflection sign as unlikely. Also, note that the second best configuration is generally $s = [1, 0, 0]$. Moreover, tracks $T_2$ and $T_3$ always best match with $C_4$ and $C_5$, 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 $s^{\text{det}} = [0, 0, 0]$ is further supported by the following observations. First, if $s_1 = 1$, $T_1$ would fall near to contour $C_3$ just like $T_2$ (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 $\phi = 45^\circ$ 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 $\phi = 0^\circ$ 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 $\chi_1/\chi_2 < 1$ 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. 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 $\phi = 0^\circ$ (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 $\phi = -45^\circ$ and $\phi = 0^\circ$ 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
TABLE I

CONTOUR MATCHING PROCEDURE RESULTS

<table>
<thead>
<tr>
<th>Subject</th>
<th>( F(0,0,0) )</th>
<th>( F(0,0,1) )</th>
<th>( F(0,1,0) )</th>
<th>( F(1,0,0) )</th>
<th>( F(1,0,1) )</th>
<th>( F(1,1,0) )</th>
<th>( F(1,1,1) )</th>
<th>Nearest contours</th>
</tr>
</thead>
<tbody>
<tr>
<td>003</td>
<td>4.03</td>
<td>5.10</td>
<td>9.27</td>
<td>8.27</td>
<td>12.93</td>
<td>13.63</td>
<td>14.73</td>
<td>1,4,3</td>
</tr>
<tr>
<td>008</td>
<td>2.95</td>
<td>4.86</td>
<td>5.33</td>
<td>7.30</td>
<td>3.69</td>
<td>7.89</td>
<td>5.58</td>
<td>10.04</td>
</tr>
<tr>
<td>009</td>
<td>2.55</td>
<td>5.18</td>
<td>4.79</td>
<td>7.02</td>
<td>2.95</td>
<td>5.08</td>
<td>2.94</td>
<td>5.01</td>
</tr>
<tr>
<td>010</td>
<td>1.88</td>
<td>5.18</td>
<td>2.36</td>
<td>6.02</td>
<td>3.57</td>
<td>5.69</td>
<td>4.46</td>
<td>6.70</td>
</tr>
<tr>
<td>011</td>
<td>2.02</td>
<td>5.10</td>
<td>5.60</td>
<td>9.53</td>
<td>3.16</td>
<td>5.79</td>
<td>4.07</td>
<td>9.22</td>
</tr>
<tr>
<td>012</td>
<td>2.08</td>
<td>4.21</td>
<td>4.76</td>
<td>7.30</td>
<td>2.70</td>
<td>5.32</td>
<td>3.20</td>
<td>6.78</td>
</tr>
<tr>
<td>017</td>
<td>2.81</td>
<td>6.35</td>
<td>4.53</td>
<td>8.12</td>
<td>2.99</td>
<td>5.02</td>
<td>5.03</td>
<td>6.79</td>
</tr>
<tr>
<td>019</td>
<td>1.64</td>
<td>6.64</td>
<td>4.85</td>
<td>8.00</td>
<td>1.64</td>
<td>6.64</td>
<td>4.85</td>
<td>8.00</td>
</tr>
<tr>
<td>020</td>
<td>1.15</td>
<td>1.15</td>
<td>5.27</td>
<td>5.27</td>
<td>1.85</td>
<td>1.85</td>
<td>5.45</td>
<td>5.45</td>
</tr>
<tr>
<td>021</td>
<td>2.90</td>
<td>6.40</td>
<td>4.06</td>
<td>8.44</td>
<td>3.30</td>
<td>8.97</td>
<td>6.25</td>
<td>11.54</td>
</tr>
<tr>
<td>022</td>
<td>2.07</td>
<td>6.53</td>
<td>5.04</td>
<td>8.56</td>
<td>2.32</td>
<td>5.27</td>
<td>2.80</td>
<td>4.25</td>
</tr>
<tr>
<td>028</td>
<td>1.71</td>
<td>3.54</td>
<td>4.21</td>
<td>5.57</td>
<td>3.79</td>
<td>4.02</td>
<td>5.62</td>
<td>6.10</td>
</tr>
<tr>
<td>033</td>
<td>2.51</td>
<td>4.73</td>
<td>6.66</td>
<td>6.61</td>
<td>3.42</td>
<td>7.08</td>
<td>9.08</td>
<td>9.98</td>
</tr>
<tr>
<td>040</td>
<td>1.74</td>
<td>5.48</td>
<td>2.59</td>
<td>5.35</td>
<td>2.57</td>
<td>5.80</td>
<td>3.30</td>
<td>7.08</td>
</tr>
<tr>
<td>044</td>
<td>1.88</td>
<td>2.84</td>
<td>5.33</td>
<td>4.81</td>
<td>2.86</td>
<td>2.49</td>
<td>4.13</td>
<td>3.74</td>
</tr>
<tr>
<td>048</td>
<td>2.02</td>
<td>5.33</td>
<td>5.45</td>
<td>7.86</td>
<td>3.70</td>
<td>5.06</td>
<td>5.27</td>
<td>6.97</td>
</tr>
<tr>
<td>050</td>
<td>3.25</td>
<td>6.29</td>
<td>7.68</td>
<td>10.52</td>
<td>4.37</td>
<td>7.39</td>
<td>7.57</td>
<td>11.23</td>
</tr>
<tr>
<td>134</td>
<td>1.64</td>
<td>6.11</td>
<td>5.18</td>
<td>8.56</td>
<td>3.38</td>
<td>6.31</td>
<td>4.56</td>
<td>7.87</td>
</tr>
<tr>
<td>165</td>
<td>1.09</td>
<td>5.35</td>
<td>3.08</td>
<td>5.03</td>
<td>3.43</td>
<td>3.89</td>
<td>3.00</td>
<td>2.99</td>
</tr>
</tbody>
</table>

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].

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 explored 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:

\[
H_{\text{head}}(z) = \frac{\beta + \alpha f_s}{\beta - \alpha f_s} \frac{z^{-1}}{1 + \frac{z^{-1}}{eta f_s^2}},
\]

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

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

\( \theta_{\text{inc}} \) is the incidence angle that, assuming the interaural axis to coincide with the \( z \) axis for sake of brevity, relates to azimuth \( \theta \) as \( \theta_{\text{inc}} = 90^\circ - \theta \) for the right ear and \( \theta_{\text{inc}} = 90^\circ + \theta \) for the left ear. A reasonably good approximation of real diffraction curves in our range of interest for the azimuth angle \(-30^\circ < \theta < 30^\circ\) is heuristically found for parameters \( \alpha_{\text{min}} = 0.1 \) and \( \theta_{\text{min}} = 180^\circ \). 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 \( \phi \), which drives the evolution of resonances’ center frequency \( F_2^1(\phi) \), dB bandwidth \( B_2^1(\phi) \), and gain \( G_2^1(\phi) \), and of the corresponding notch parameters \( F_2^i(\phi), B_2^i(\phi), G_2^i(\phi), i = 1, 2, 3 \). 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 \( \phi \) values.\(^7\) Second, a fifth order polynomial \( P_5^i \) or \( P_3^j \), 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 \( F_5^i(\phi) \) 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 \( F_5^i(\phi), j = 1, 2, 3 \). In the remainder of this Section we refer to HRTFs given by the fully resynthesized model (without contour extraction) as \( H^a \), while HRTFs resulting from the contour-parameterized model as \( H^c \).

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]

\[
H_{\text{res}}^{1}(z) = \frac{1 + (1 + k) \frac{B_4^i}{F_4^i} + (1 - k) z^{-1} + (-k - (1 + k) \frac{B_2^i}{F_2^i}) z^{-2}}{1 + l(1 - k) z^{-1} - k z^{-2}}, \tag{12}
\]

where

\[
k = \frac{\tan \left( \frac{\pi B_4^i(\phi)}{f_s} \right) - 1}{\tan \left( \frac{\pi B_2^i(\phi)}{f_s} \right) + 1}, \quad l = - \cos \left( 2\pi \frac{F_4^i(\phi)}{f_s} \right), \tag{13}
\]

\[
V_0 = 10 e^{-k \frac{f_s}{\pi}}, \quad H_0 = V_0 - 1, \tag{14}
\]

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

\[
H_{\text{res}}^{2}(z) = \frac{V_0(1 - h)(1 - z^{-2})}{1 + 2h z^{-1} + \{2h - 1\} z^{-2}}, \tag{15}
\]

\(^7\)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. 6. The structural HRTF model. Customization is performed through parameter extraction from anthropometric measurements and a pinna picture.

\[
h = \frac{1}{1 + \tan \left( \frac{\pi B_2^i(\phi)}{f_s} \right)}, \tag{16}
\]

while \( t \) 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 \( H_{\text{res}}^1 \) with the only differences in the parameters’ description. In order to keep notation correct, polynomials \( P_n^1 \) must be substituted by the corresponding notch counterparts \( P_n^2 \), \( j = 1, 2, 3 \), and parameter \( k \) defined in (13) replaced by its “cut” version

\[
k = \frac{\tan \left( \frac{\pi B_2^i(\phi)}{f_s} \right)}{\tan \left( \frac{\pi B_1^i(\phi)}{f_s} \right) + V_0}. \tag{17}
\]

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:

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

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, \( H_{1,\text{cont}}^c \);

\[
\text{Fig.} 7. \text{ Example plots of PRTF resynthesis with similar filter structures can be found in [44].}
\]

\[
\text{B. Results and Discussion}
\]

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

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

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, \( H_{1,\text{cont}}^c \).
Fig. 7. Spectral distortion between reconstructed and measured median-plane HRTFs (mean and standard deviation over 18 CIPIC subjects).

Table II

<table>
<thead>
<tr>
<th>Subject</th>
<th>( m(T_1, C_j) )</th>
<th>( m(T_2, C_4) )</th>
<th>( m(T_3, C_3) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>003</td>
<td>11.42%</td>
<td>-</td>
<td>12.02%</td>
</tr>
<tr>
<td>008</td>
<td>8.80%</td>
<td>8.69%</td>
<td>14.67%</td>
</tr>
<tr>
<td>010</td>
<td>4.80%</td>
<td>2.96%</td>
<td>18.76%</td>
</tr>
<tr>
<td>011</td>
<td>8.75%</td>
<td>7.77%</td>
<td>12.20%</td>
</tr>
<tr>
<td>012</td>
<td>-</td>
<td>8.98%</td>
<td>8.60%</td>
</tr>
<tr>
<td>017</td>
<td>7.80%</td>
<td>3.44%</td>
<td>17.97%</td>
</tr>
<tr>
<td>019</td>
<td>-</td>
<td>4.48%</td>
<td>5.92%</td>
</tr>
<tr>
<td>020</td>
<td>-</td>
<td>4.27%</td>
<td>11.73%</td>
</tr>
<tr>
<td>021</td>
<td>-</td>
<td>10.16%</td>
<td>11.73%</td>
</tr>
<tr>
<td>027</td>
<td>-</td>
<td>2.06%</td>
<td>7.63%</td>
</tr>
<tr>
<td>028</td>
<td>-</td>
<td>8.05%</td>
<td>14.79%</td>
</tr>
<tr>
<td>033</td>
<td>4.52%</td>
<td>3.55%</td>
<td>16.44%</td>
</tr>
<tr>
<td>040</td>
<td>2.98%</td>
<td>5.56%</td>
<td>12.92%</td>
</tr>
<tr>
<td>044</td>
<td>-</td>
<td>6.49%</td>
<td>8.10%</td>
</tr>
<tr>
<td>048</td>
<td>4.01%</td>
<td>-</td>
<td>16.19%</td>
</tr>
<tr>
<td>050</td>
<td>-</td>
<td>7.28%</td>
<td>18.95%</td>
</tr>
<tr>
<td>134</td>
<td>2.59%</td>
<td>5.16%</td>
<td>10.13%</td>
</tr>
<tr>
<td>165</td>
<td>3.91%</td>
<td>4.11%</td>
<td>6.44%</td>
</tr>
</tbody>
</table>

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

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

As expected, \( H_{\text{tot}}^{c} \) is the response with the highest average SD. As a matter of fact, errors in the resynthesized resonant \( (H_{\text{res}}^{n}) \) and contour-parameterized reflective \( (H_{\text{eff}}^{c}) \) components combine together yielding the SD for \( H_{\text{tot}}^{c} \), 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 \( H_{\text{tot}}^{c} \) tends to decrease apart from a new rise at the last elevation angles, which is conversely due to greater errors in the resonant component \( H_{\text{res}}^{n} \). An informal inspection of resonant components at higher elevations revealed indeed that the second modeled 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 \( H_{\text{eff}}^{c} \) 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 \( H_{\text{tot}}^{c} \).

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:

\[
m(T_j, C_i) = \frac{1}{n(T_j)} \sum_{\phi} \frac{f_n^{j}(\phi) - F_n^{j}(\phi)}{f_n^{j}(\phi)} \cdot 100\%,
\]

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 might 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 psychoacoustically 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 \( H_{\text{tot}}^{c} \) 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 $-45^\circ$. Alternative HRTF data sets or BEM simulations should be considered in order to extend the ray tracing procedure to the range $-90^\circ < \phi < -45^\circ$. 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 $\phi = 45^\circ$ up to $\phi = 90^\circ$ 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


