Fitting Pinna-Related Transfer Functions to Anthropometry for Binaural Sound Rendering

Simone Spagnol [1], Michele Geronazzo [2], Federico Avanzini [3]

Department of Information Engineering, Università di Padova
Padova, Italy
[1] spagnols@dei.unipd.it
[2] geronazz@dei.unipd.it
[3] avanzini@dei.unipd.it

***Abstract*—This paper faces the general problem of modeling pinna-related transfer functions (PRTFs) for 3-D sound rendering. Following a structural approach, we aim at constructing a model for PRTF synthesis which allows to control separately the evolution of ear resonances and spectral notches through the design of two distinct filter blocks. Taking such model as endpoint, we propose a method based on the McAulay-Quatieri partial tracking algorithm to extract the frequencies of the most important spectral notches. Ray-tracing analysis performed on the so obtained tracks reveals a convincing correspondence between extracted frequencies and pinna geometry of a bunch of subjects.**

I. Introduction

At the beginning of the last century, Lord Rayleigh's studies on the scattering of sound waves by obstacles gave birth to the extensive and still partially misunderstood field of 3-D sound. Within the context of his notable Duplex Theory of Localization [1], a commonly known formula that approximates the behaviour of sound waves diffracting around the listener's head provided indeed a first glance of the today-called head-related transfer function (HRTF). Alas, despite the importance and applicative potential of such a centenary theory, most of the efforts towards efficient modeling of HRTFs were spent in the last few decades only.

Throughout these years, low-order rational functions [2] and series expansions of HRTFs [3] were proposed as tools for HRTF modeling. Albeit the straightforward nature and intrinsic simplicity of both techniques, real-time HRTF modeling requires fast computations which cannot undergo the complexity of filter coefficients and weights, respectively. Oppositely, structural modeling [4] represents nowadays the ultimate alternative approach for real-time HRTF rendering: if we isolate the contributions of the user's head, pinnae and torso to the HRTF in different subcomponents, each accounting for some well-defined physical phenomenon, then thanks to linearity we can reconstruct the global HRTF on-the-fly from a proper combination of all the considered effects. What we have is then a model which is both economical (if we assume that each physical phenomenon depends from few parameters) and well-suited to real-time implementations; as a further advantage, the intuitive nature of physical parameters enforces the chance to relate the model to simple anthropometrical measurements.

The present work exclusively deals with the contribution of the pinna to the HRTF. Even though head motion is perceptually a better discriminant, pinna cues are still of great importance in sound localization. A number of experiments have shown that, conversely to azimuth effects that can be reduced to simple binaural quantities, elevation effects - which are the result of a superposition of scattering waves influenced by a number of resonant modes - are basically monaural and heavily depend on the listener's anthropometry. Finding a suitable model for representing the pinna contribution to the HRTF (whose transfer function we commonly refer to as Pinna-Related Transfer Function - PRTF) is thus a crucial task, with the ultimate challenge in this direction being relating the model's parameters to easily obtainable anthropometric measurements on the user's pinnae. The resulting model, cascaded to a simple Head-and-Torso (HAT) model [5], will allow us to achieve a complete structural HRTF representation.

This paper lies its foundations on the iterative approach previously presented by the authors in [6] to separate resonance effects from pinna reflections in experimentally measured PRTFs. Moving from this start point, a method for extracting the frequencies of the most important notches is here developed (Section III), followed by a discussion on the possible relation between notch frequencies and anthropometry (Section IV).

II. Previous Works

According to Batteau [7], high-frequency tones are typically reflected by the outer ear, as long as their wavelength is small enough compared to the pinna dimensions. Consequently, interference between the direct and reflected waves causes sharp notches to appear in the high-frequency side of the received signal's spectrum with a periodicity that is inversely proportional to the time delay of each reflection. Such observation led to a first rough double-path model of the pinna [8]. Unhappily, this model lacks the description of pinna resonant modes: as Shaw argued [9], since pinna cavities act as resonators the frequency content of both the direct and the reflected sound

MMSP'10, October 4-6, 2010, Saint-Malo, France.
978-1-4244-8112-5/10/$26.00

waves is significantly altered. Batteau's model has accordingly been improved by Barreto *et al.*, with a new reflection structure [10] represented by four parallel paths cascaded to a low-order resonator block. Furthermore, the model parameters were associated to eight measured anthropometric features by means of multiple regression analysis [11]. The trouble is as well as providing no cloudless evidence of the physics behind the scattering phenomenon, the considered measures can only be acquired through the use of a 3-D laser scanner. In any case, these works surely endorse our final PRTF model's "resonance-plus-delay" architecture.

A different approach for reflection modeling, acting both in the time and frequency domains, was pursued by Raykar *et al.* [12]. Robust digital signal processing techniques are used here to extract the frequencies of the spectral notches due to the pinna alone: first the autocorrelation function of the HRIR's windowed LP residual is computed; then, frequencies of the spectral notches are found as the local minima of the group-delay function of the windowed autocorrelation. What's more, the authors advance a ray-tracing argument to attest that the so found spectral notches are related to the shape and anthropometry of the pinna. Specifically, knowing that the elevation-dependent temporal delay $t_d(\phi)$ between the direct and the reflected wave at the ear canal puts the point of reflection at a distance

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

where c is the speed of sound (approximately 343 m/s), and assuming the reflection coefficient to be positive, then each extracted frequency f_0 is considered as the first of a periodic series

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

in particular

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

The corresponding distance $d(\phi)$ was then projected onto the 2-D image of the pinna, resulting in a mapping consistent with reflections on the crus helias and concha wall.

Another important contribution on PRTF modeling was provided by Satarzadeh *et al.* [13]. In this work, PRTFs for elevation $\phi = 0^\circ$ are synthesized through a model composed of two second-order bandpass filters and one comb filter, which respectively approximate the two major resonances (Shaw's resonant modes 1 and 4) and one main reflection. The frequency of the comb filter's first tooth, f_0, is estimated from the spacing of consecutive notches in the PRTF spectrum: consequently, if the filter takes the form $[1 + \rho \exp(-st_d)]$ (where ρ is the reflection coefficient), then the time delay between direct and reflected wave is calculated as

$$t_d = \frac{1}{2f_0} \tag{4}$$

if $\rho > 0$ (according to Raykar *et al.*), or as

$$t_d = \frac{1}{f_0} \tag{5}$$

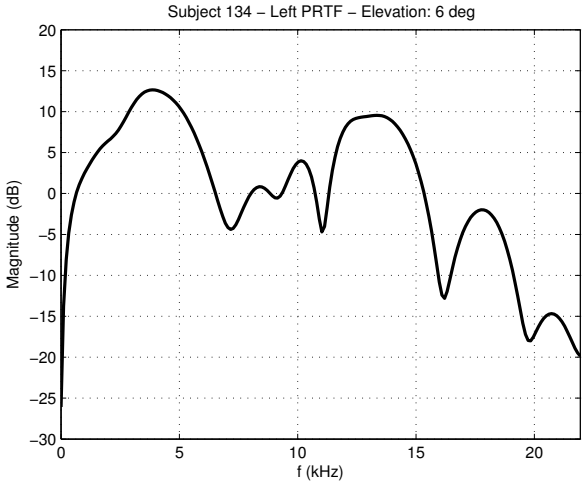


Fig. 1. *An example of PRTF.*

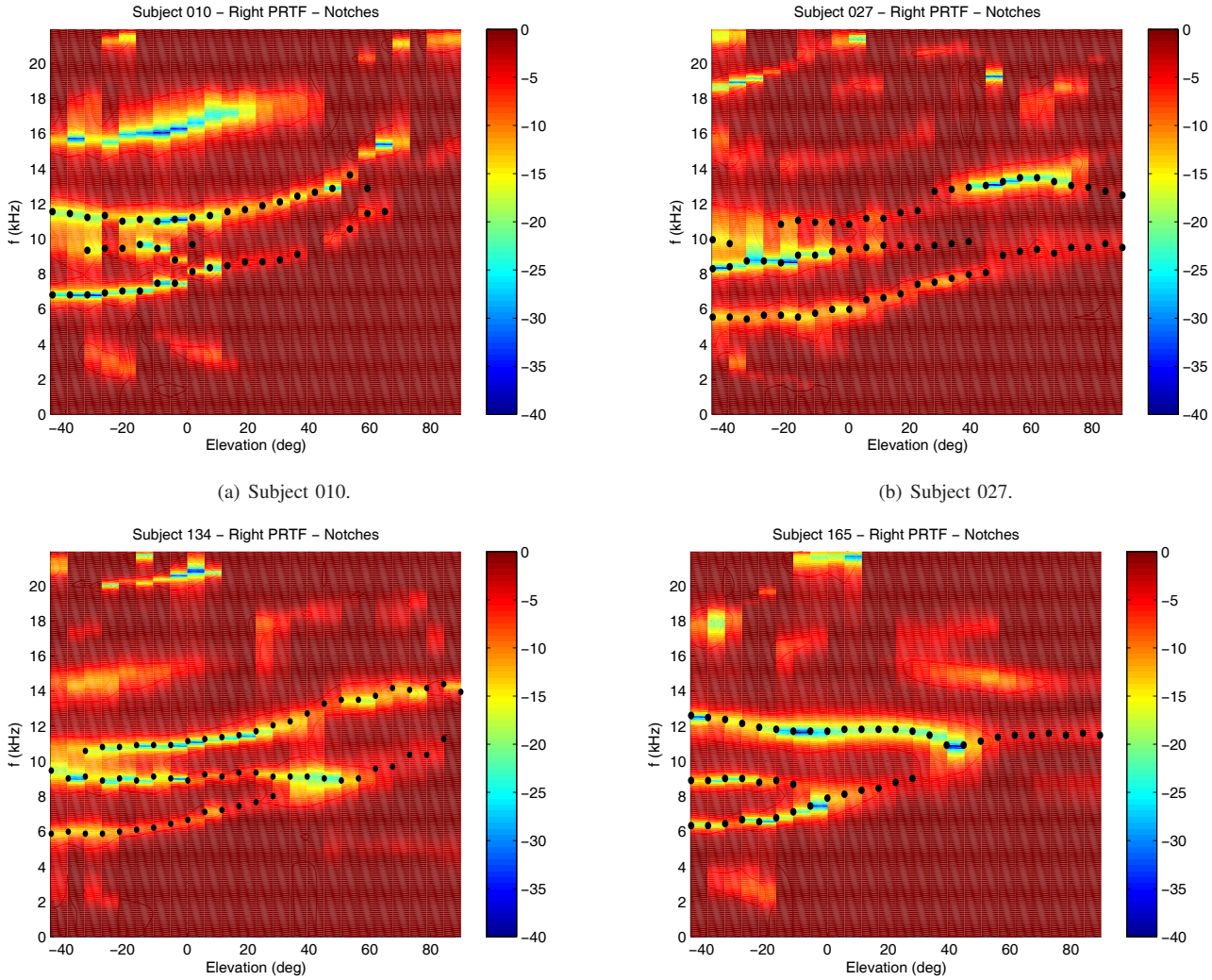
if $\rho < 0$. Once the sign of the reflection coefficient is determined from the impulse response, the distances inferred from Eq. (1) put the point of reflection either at the back of the concha or at the edge of the rim. In addition, a cylindrical approximation of the concha is used with the purpose of directly parameterizing the resonances' coefficients. In conclusion, such a low-order anthropometry-based filter provides a good fit to the experimental PRTF in all cases where the pinna has an approximately cylindrical shaped concha and a structure with a dominant reflection area (concha or rim). However, besides considering solely the frontal direction of the sound wave, taking into account a single reflection appears as a limiting factor.

III. PRTF Modeling

Taking [12], [13] and a "resonance-plus-delay" PRTF model as starting points, the final goal of our work is the construction of an essential multi-notch filter suitable for anthropometric parametrization. In order to analyze PRTFs we consider HRIRs from the CIPIC database [14], a public domain database of high spatial resolution HRIR evaluations at 1250 directions for 45 different subjects which also includes anthropometric measurements.

A. Pre-processing

Since sensitivity of PRTFs to azimuth is weak, we focus on HRIRs sampled on the median plane, with elevation varying from -45° to 90°. Given that the magnitude response of the earless head with respect to a sound source located in the median plane is ideally flat if the head is modeled as a rigid sphere, the only preprocessing step that is needed to obtain a raw estimate of the PRTF is to window the corresponding HRIR using a 1.0 ms Hann window. In this way, effects due to reflections caused by shoulders and torso are removed from the PRTF estimate. Figure 1 reports an example of PRTF, where spectral notches and resonances can be easily detected.



(a) Subject 010.

(b) Subject 027.

(c) Subject 134.

(d) Subject 165.

Fig. 2. *Spectral notch plots for different elevations.*

We implemented an algorithm for separating the reflective component of the spectrum from the resonant component in the so built PRTFs. The idea behind the algorithm is to iteratively compensate the PRTF magnitude spectrum with an approximate multi-notch filter until no significant notches are left. Once convergence is reached at iteration n, the PRTF spectrum contains the resonant component, while a combination of the n multi-notch filters provides the reflective component. Details of the separation algorithm can be found in [6].

An accurate inspection of the magnitude plots reveals that CIPIC subjects generally show a common trend with respect to elevation in the resonant component of the PRTF. Specifically, three major resonances can be distinctly identified up to 14 kHz:

- the first one centered around 4 kHz, spanning all elevations with almost equal intensity for all subjects. This resonance agrees with Shaw's omnidirectional mode 1;
- a second resonance whose bandwidth interferes with the first one's at high elevations around 7 kHz; this could be Shaw's mode 2;
- the third, occurring at low elevations around 12 kHz, is likely to incorporate Shaw's mode 4.

A more detailed analysis of resonances, which exploits an identification system based on a sixth-order ARMA model [15] to track center frequencies, can be found in [6]. In this context suffice it to observe that, since the last two resonances are excited in mutually exclusive elevation ranges, we may look forward to a double-resonance filter design. We now turn to a more extensive analysis of the reflective component of the PRTF spectrum.

B. Spectral notch analysis

The 3-D plots in Figure 2 represent the frequency notches' contribution for four different CIPIC subjects at all available elevations. The same subjects also appear in [12]; indeed, we

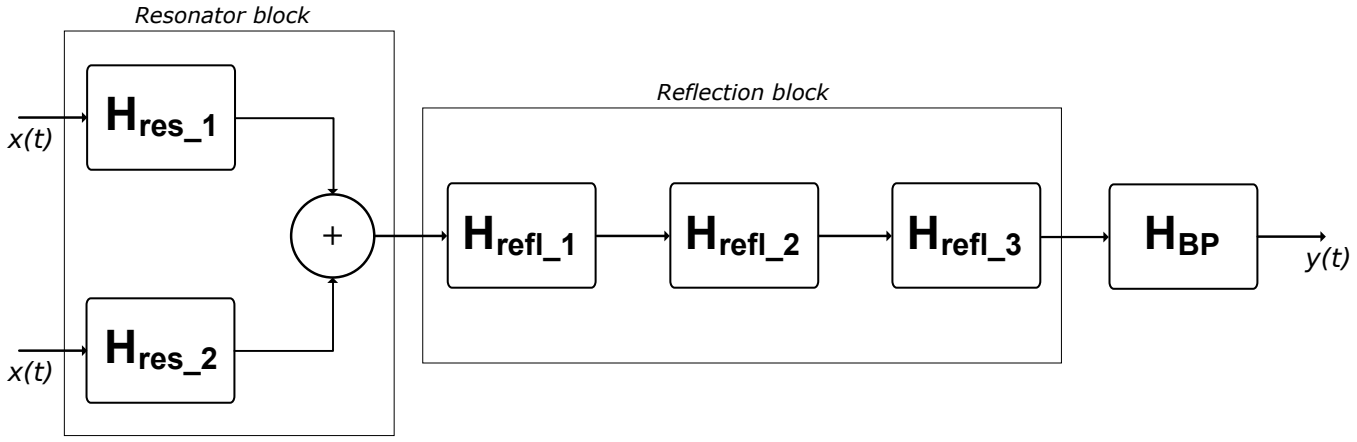


Fig. 3. *General model for the reconstruction of PRTFs.*

chose to report these results in order to facilitate comparison with Raykar's work on notch frequencies extraction.

As expected, reflection patterns strongly depend on pinna shape and elevation. Nevertheless, a number of common trends can be acknowledged here too. To the end of investigating them, we inherit a widely used analysis tool in the field of sinusoidal modeling, the McAulay-Quatieri partial tracking algorithm [16], originally used to group sinusoidal partials along consecutive temporal windows according to their spectral location. In our context, the very algorithm can be exploited to track the most marked notch patterns along elevation, as long as temporal evolution is conceptually replaced by elevation dependency and spectral notches take the role of partials: for this reason, we refer to it as "notch tracking" algorithm. With respect to its first formulation, suffice it to add that the notch detection (originally "peak detection") step trivially locates all of the local minima in the reflective component's spectrum, and that the matching interval for the notch tracking procedure is set to $\Delta = 3$ kHz.

Two post-processing steps are performed on the obtained tracks. First, since it is convenient to restrict our attention to the frequency range where reflections due to the pinna alone are most likely seen, we delete the tracks which are born and die outside the range $4-14$ kHz. Second, we delete the tracks that do not present a notch deeper than 5 dB, since overall shallow notches are not likely to be associated with a major reflection.

The dotted tracks superposed on the plots in Figure 2 represent the outputs of the notch tracking algorithm. We can immediately notice a great similarity between these tracks and those computed in [12] through a labyrinthine DSP-based algorithm. In particular, three main tracks are seen for all four subjects, while the shorter tracks in the plots of Subject 010 and Subject 027 very probably represent the continuation of the missing track at those specific elevations. Realistically, gaps between tracks may be caused by the algorithm's unlikelihood of locating proper minima due to uncontrollable events such as the superposition of two different notches or the presence of shallow valleys in the considered region of the magnitude plot. Nonetheless, the three aforementioned main tracks indicate us that congruous reflection patterns appear in different PRTFs. We will return to this point in Section IV.

C. A structural model of the pinna

The information collected from the outputs of the decomposition and notch tracking algorithms allows us to look forward to a structural model of the pinna which models two resonances and three spectral notches in the PRTF. As Figure 3 depicts, our aim is to design two filter blocks, one for resonances and one for reflections.

Assuming that the three major notches in the PRTF spectrum are caused by three different reflections, standard second-order peak and notch filters [17] can be used to model each resonance and reflection, respectively. The three notch filters must be placed in series and cascaded to the parallel of the two peak filters, resulting in an eighth-order global filter. Finally, since we have studied PRTF effects in the frequency range $4-14$ kHz, we are free to cascade a bandpass (BP) filter which cuts undesired frequencies to the whole structure. Application of this extra stage would become necessary when integrating the PRTF model with a HAT model.

Clearly, full parametrization of the model on anthropometrical measurements is needed in order to have complete control of the filter parameters. As for what concerns the reflective component of the PRTF, we move towards the characterization of a realistic mapping between notch frequencies and reflection points over the pinna, each having a reflection coefficient which controls the correlated notch filter's magnitude parameter.

IV. Notches and anthropometry

We dedicate this last section to a plausible explanation of the physical mechanism lying behind the production of frequency notches in the PRTF spectrum. As already pointed out, we relate each major notch to a distinct reflection, assuming it to be the first and most marked notch of a periodic series.

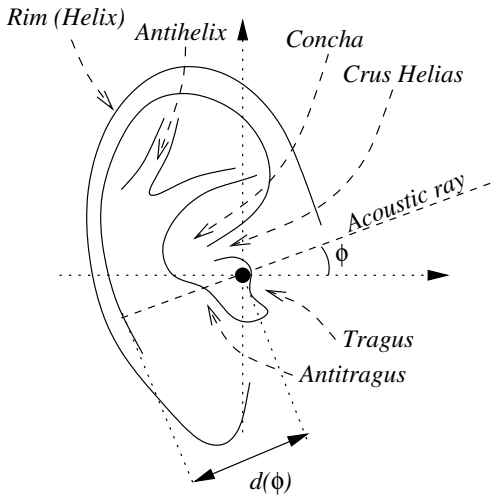


Fig. 4. *Anatomy of the pinna.*

A. Reflection coefficient sign

Reflection models usually assume all reflection coefficients to be positive. If this were the case, the extra distance travelled by the reflected wave with respect to the direct wave must be equal to half a wavelength in order for destructive interference to occur, which translates into spectral notches in the frequency domain (see Eq. (4)). This was the assumption taken by [12] when tracing reflection points over pinna images based on the extracted notch frequencies.

Nevertheless, Satarzadeh [18] drew attention to the fact that the majority of CIPIC subjects exhibit a clear negative reflection in the HRIR. He motivated this result by hypothesizing a boundary created by an impedance discontinuity between the pinna and air which could produce its own reflection, reversing the phase of the wave. In this latter case, destructive interference would not appear for half-wavelength delays anymore, yet only for full-wavelength delays (see Eq. (5)).

B. Ray tracing

Following Satarzadeh's hypothesis, we choose to use the negative reflection assumption in establishing a relation between notches and pinna geometry through a simple ray-tracing procedure, very similar to the one described in [12].

Right pinna images are taken from the CIPIC database and uniformly rescaled in order to match parameters d_5 (pinna height) and d_6 (pinna width) [14]. The distance of each reflection point with respect to the entrance of the ear canal is calculated through Eqs. (1) and (5), leading to the relation

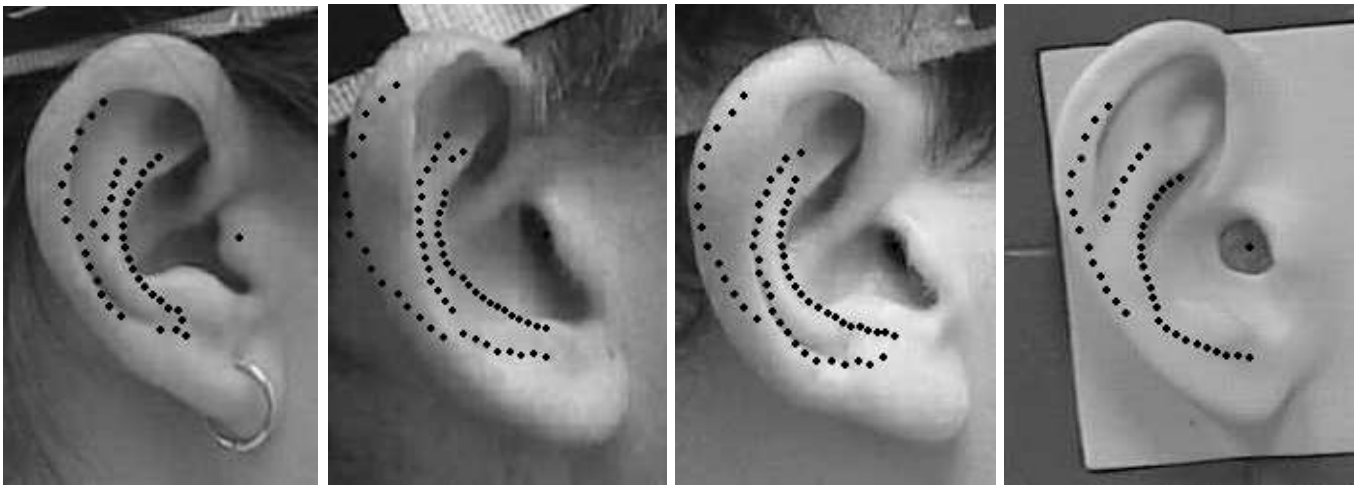
$$d(\phi) = \frac{c}{2f_0(\phi)}, \tag{6}$$

where $f_0(\phi)$ represents the frequency of the current notch at elevation ϕ. The negative reflection coefficient assumption causes distances to be approximately doubled with respect to those calculated in [12]. Then, if we consider the 2-D polar coordinate system illustrated in Figure 4 having the right ear canal entrance as origin, each notch is mapped to the point $(d(\phi), \pi + \phi)$.

Results for subjects 010, 027, 134, and 165 are reported in Figure 5. For all these subjects, the so-obtained mapping shows a high degree of correspondence between computed reflection points and pinna geometry. One can immediately notice that the track nearest to the ear canal very closely follows the concha wall of each subject for all elevations, except for a couple of cases:

- at low elevations, displacement of points may be caused by the little extra distance needed by the wave to pass over the crus helias;
- Subject 010's track disappears at around $\phi = 60°$ probably because of the insufficient space between tragus and antitragus that causes the incoming wave to reflect outside the concha.

The intermediate track falls upon the area between concha and rim, with variable length among subjects:



(a) Subject 010. (b) Subject 027. (c) Subject 134. (d) Subject 165.

Fig. 5. *Reflection points on four CIPIC subjects' right pinnae.*

- in the case of subjects 010 and 165 the track is faint and probably due to the antihelix;
- conversely, subjects 027 and 134 present a longer and deeper track, that we visually associate to a reflection on the rim's edge.

Finally, the furthest track follows the shape of the rim and is likely to be associated to a reflection in the inner wall of it, except for Subject 010 whose reflection occurs at the rim's edge. A strong evidence that validates the track's connection to the rim structure lies in the fact that the rim terminates in the vicinity of the point where the track disappears.

C. Model fitting to anthropometry

Further refinements should be applied to the above preliminary analysis for a more detailed account of the reflection structures of a vast test bed of subjects to be performed, in particular the use of a 3-D model of the pinna that allows to investigate its horizontal section. As a matter of fact, 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. Hence plotting distances on the side-view images should take into account the displacement caused by the flare angle of the pinna.

Nevertheless, our preliminary analysis has revealed a satisfactory correspondence between computed reflection points and reflective structures over the pinna. This opens the door for a very attractive approach to the parametrization of the structural PRTF model based on individual anthropometry. Indeed, given a 2-D image or a 3-D reconstruction of the user's pinna, one can easily trace the contours of the concha wall, antihelix and rim, compute each contour's distance with respect to the ear canal for all elevations, and extrapolate the notch frequencies by reversing Eq. 5. Obviously, since notch depth strongly varies within subjects and elevations, the reflection coefficient must also be estimated for each point. This problem theoretically requires strong physical arguments; alternatively, psychoacoustical criteria could be used in order to evaluate the perceptual relevance of notch depth, and potentially simplify the fitting procedure.

V. Conclusions and future work

In this paper we presented an approach for extracting the reflective component of the PRTF, taking a structural model of the pinna as endpoint. Our attempt towards the explanation of the scattering process resulting in the most important spectral notches in the PRTF provided visually convincing results. Clearly, in order to fully justify these findings, robust theoretical motivations and a rigorous analysis using a vast test bed of subjects are required. Ongoing and future work also includes understanding of the reflection coefficient and relating the resonant component of the PRTF to anthropometry.

Acknowledgment

The authors wish to acknowledge Professor Ralph Algazi for his support with the CIPIC database and the accompanying pinna photos.

References

[1] J. W. Strutt, "On our perception of sound direction," *Philosophical Magazine*, vol. 13, pp. 214–232, 1907.

[2] E. C. Durant and G. H. Wakefield, "Efficient model fitting using a genetic algorithm: pole-zero approximations of HRTFs," *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 1, pp. 18–27, 2002.

[3] 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. Am.*, vol. 91, no. 3, pp. 1637–1647, 1992.

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

[5] 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 Convention of the Audio Engineering Society*, Los Angeles, CA, USA, 2002.

[6] M. Geronazzo, S. Spagnol, and F. Avanzini, "Estimation and modeling of pinna-related transfer functions," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, Graz, Austria, September 6-10 2010, accepted for publication.

[7] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Series B, Biological Sciences*, vol. 168, no. 1011, pp. 158–180, August 1967.

[8] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, no. 4, pp. 1152–1165, April 1978.

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

[10] K. J. Faller II, A. Barreto, N. Gupta, and N. Rishe, "Time and frequency decomposition of head-related impulse responses for the development of customizable spatial audio models," *WSEAS Transactions on Signal Processing*, vol. 2, no. 11, pp. 1465–1472, 2006.

[11] N. Gupta, A. Barreto, and M. Choudhury, "Modeling head-related transfer functions based on pinna anthropometry," in *Proc. of the Second International Latin American and Caribbean Conference for Engineering and Technology (LACCEI)*, Miami, FL, USA, 2004.

[12] 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. Am.*, vol. 118, no. 1, pp. 364–374, July 2005.

[13] P. Satarzadeh, R. V. Algazi, and R. O. Duda, "Physical and filter pinna models based on anthropometry," in *Proc. 122nd Convention of the Audio Engineering Society*, Vienna, Austria, May 5-8 2007.

[14] R. V. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano, "The CIPIC HRTF database," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, 2001, pp. 1–4.

[15] P. A. A. Esquef, M. Karjalainen, and V. Välimäki, "Frequency-zooming ARMA modeling for analysis of noisy string instrument tones," *EURASIP Journal on Applied Signal Processing: Special Issue on Digital Audio for Multimedia Communications*, no. 10, pp. 953–967, 2003.

[16] R. J. McAulay and T. F. Quatieri, "Speech analysis/synthesis based on a sinusoidal representation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, 1986.

[17] U. Zölzer, Ed., *Digital Audio Effects*. New York, NY, USA: J. Wiley & Sons, 2002.

[18] P. Satarzadeh, "A study of physical and circuit models of the human pinnae," Master's thesis, University of California Davis, 2006.