Structural Modeling of Pinna-Related Transfer Functions for 3-D Sound Rendering

Simone Spagnol spagnols@dei.unipd.it

Michele Geronazzo geronazz@dei.unipd.it

Federico Avanzini avanzini@dei.unipd.it

Università di Padova

ABSTRACT

This paper considers the general problem of modeling pinna-related transfer functions (PRTFs) for 3-D sound rendering. Following a structural approach, we present an algorithm for the decomposition of PRTFs into ear resonances and frequency notches due to reflections over pinna cavities and exploit it in order to deliver a method to extract the frequencies of the most important spectral notches. Ray-tracing analysis reveals a convincing correspondence between extracted frequencies and pinna cavities of a bunch of subjects. We then propose a model for PRTF synthesis which allows to control separately the evolution of resonances and spectral notches through the design of two distinct filter blocks. The resulting model is suitable for future integration into a structural head-related transfer function model, and for parametrization over anthropometrical measurements of a wide range of subjects.

1. 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 an iterative algorithm that separates 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, followed by a discussion on the possible relation between notch frequencies and anthropometry. Finally, a structural model of the pinna is proposed.

2. PREVIOUS WORKS

According to Batteau [6], 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 [7]. Unhappily, this model lacks the description of pinna resonant modes: as Shaw argued [8],

since pinna cavities act as resonators the frequency content of both the direct and the reflected sound waves is significantly altered. Batteau's model has accordingly been improved by Barreto *et al.*, with a new reflection structure [9] 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 [10]. 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.* [11]. 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

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

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.* [12]. 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

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

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.

3. PRTF ANALYSIS

Taking the last two works described in the previous section as an inspiration and a "resonance-plus-delay" PRTF model as starting point, the main and final goal of our work is the construction of an essential multi-notch filter suitable for anthropometric parametrization. This obviously requires a PRTF analysis step. In order to analyze PRTFs, we consider measured HRIRs from the CIPIC database [13], a public domain database of high spatial resolution HRIR measurements at 1250 directions for 45 different subjects along with their anthropometry. We choose to investigate the behaviour of pinna features in subjects 010, 027, 134, and 165 in order to facilitate comparison with previous works on notch frequencies extraction (the same subjects' PRTFs were analyzed in [11]).

3.1 The Separation Algorithm

For purpose of analysis we focus on HRIRs sampled on the median plane, with elevation varying from -45° to 90°. As a matter of fact, since sensitivity of PRTFs to azimuth is weak [12], we roughly expect PRTFs to be elevation dependent only. Such an assumption makes the PRTF model suitable for all azimuths.

Knowing that the magnitude response of an earless head with respect to a sound source in the median plane is ideally flat if the head is modeled as a rigid sphere, the only preprocessing step we apply to obtain a raw estimate of the PRTF is windowing the corresponding HRIR using a 1.0 ms Hann window [11]. In this way, spectral effects due to reflections caused by shoulders and torso are removed from the PRTF estimate.

In order to isolate the spectral notches in the so built PRTFs we exploit an ad-hoc designed algorithm that returns an estimate of the separated resonant and reflective components. Figure 1 reports the complete flow chart of this analysis algorithm. The idea beyond it is to iteratively compensate the PRTF magnitude spectrum with an approximate multi-notch filter until no significant notches are left. Once convergence is reached (say at iteration $\underline{i}$), the PRTF spectrum $H_{res}^{(\underline{i})}$ will contain the resonant component, while the combination $H_{refl}^{(\underline{i})}$ of the multi-notch filters will provide the reflective component. The algorithm's initial conditions heavily influence the final result; three parameters have to be chosen:

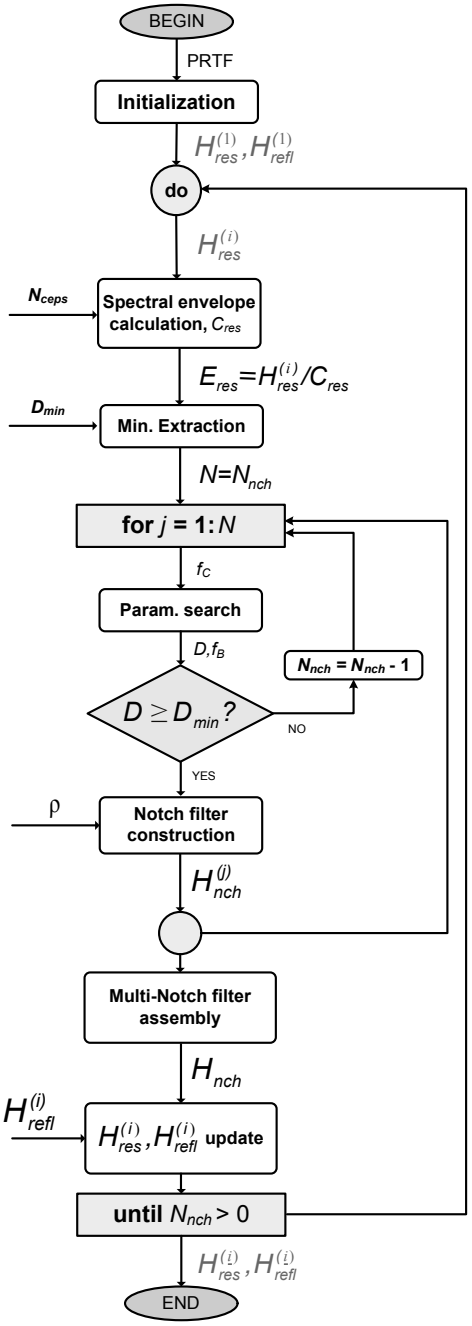


Figure 1. Flow chart of the analysis algorithm.

- N_{ceps}, the number of cepstral coefficients used for estimating the PRTF spectral envelope at each iteration;
- D_{min}, the minimum dB depth threshold for notches to be considered;
- ρ, the reduction factor for every notch filter bandwidth (its purpose will be discussed below).

Before entering the core of the algorithm, let $H_{res}^{(1)}$ match the PRTF and set $H_{refl}^{(1)}$ to 1. These two frequency responses will be updated at each iteration, resulting in $H_{res}^{(i)}$ and $H_{refl}^{(i)}$ at the beginning of the i-th iteration. If $N_{nch}^{(i)}$ is the number of "valid" notches algebraically identified at the end of it, the algorithm will terminate at iteration $\underline{i}$ if $N_{nch}^{(\underline{i})} = 0$, while $H_{res}^{(\underline{i})}$ and $H_{refl}^{(\underline{i})}$ will respectively contain the resonant and reflective components of the PRTF. As one may expect, both the number of iterations and the quality of our decomposition strongly rely on a good choice of the above parameters. For instance, choosing D_{min} too close to zero may lead to an unacceptable number of iterations; conversely, a high value of D_{min} could result in a number of uncompensated notches in the resonant part of the PRTF. In the following, we present the step-by-step analysis procedure on $H_{res}^{(i)}$, assuming that $N_{nch}^{(i-1)} > 0$. For the sake of simplicity, in the following the apex (i) indicating iteration number is dropped from all notation.

3.1.1 Residue computation

First, in order to extract properly the local minima due to pinna notches in the PRTF, the resonant component of the spectrum must be compensated for. To this end, the real cepstrum of H_{res} is calculated; then, by liftering the cepstrum with the first N_{ceps} cepstral coefficients and performing the FFT, an estimate of the spectral envelope of H_{res} is obtained, which we call C_{res}.

The parameter N_{ceps} must be chosen adequately, since it is crucial in determining the degree of detail of the spectral envelope. As N_{ceps} increases, the notches' contribution is reduced both in magnitude and in passband while the resonance plot becomes more and more detailed. We experimentally found that the optimal number of coefficients that capture the resonant structure of the PRTF while leaving all the notches out of the spectral envelope is $N_{ceps} = 4$. This number also matches the maximum number of modes identified by Shaw which appear at one specific spatial location: for elevations close to zero, modes 1, 4, 5, and 6 are excited. Once C_{res} is computed, we subtract it from the dB magnitude of H_{res} and obtain the residue E_{res}.

3.1.2 Multi-notch filter parameter search

At this point E_{res} should present an almost flat spectrum with a certain number of notches. Parameter N_{nch} is first set to the number of local minima in E_{res} deeper than D_{min}, extracted by a simple notch picking algorithm. Our aim is to compensate each notch with a second-order notch filter, defined by three parameters: central frequency f_C, bandwidth f_B, and notch depth D.

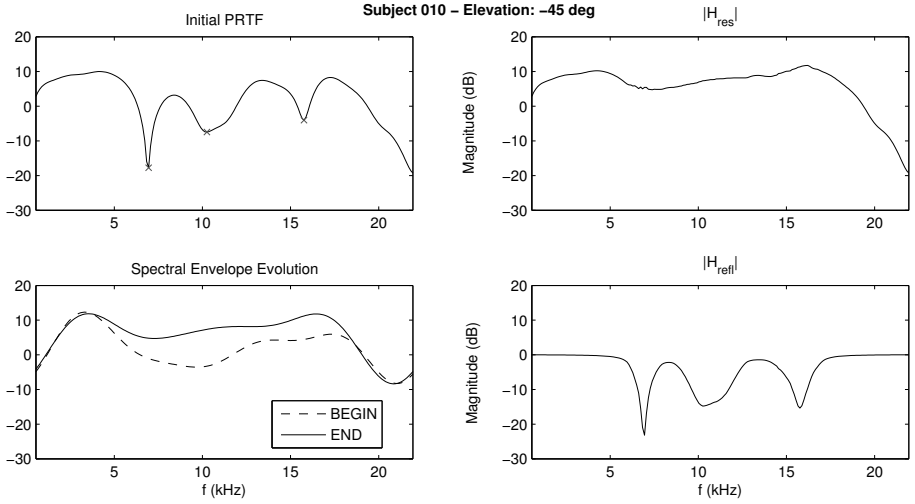


Figure 2. An example of the algorithm's evolution. The PRTF magnitude in the top left panel is decomposed into resonances (top right panel) and frequency notches (bottom right panel). The bottom left panel shows the evolution of the PRTF spectral envelope from the first iteration to convergence.

Consider the j-th local minimum. The central frequency of the corresponding notch filter f_C is immediately determined, while notch depth is found as $D = |E_{res}(f_C)|$. Computation of f_B is less straightforward. Indeed, f_B is calculated as the standard 3-dB bandwidth, i.e. $f_B = f_r - f_l$, where f_l and f_r are respectively the left and right +3 dB level points relative to f_C in E_{res}, except for the following situations:

1. if $D < 3$ dB, the 3-dB bandwidth is not defined. Then f_r and f_l are placed at an intermediate dB level, halfway between 0 and $-D$ in a linear scale;

2. if the local maximum of E_{res} immediately preceding (following) f_C does not lie above the 0-dB line while the local maximum immediately following (preceding) does, f_B is calculated as twice the half-bandwidth between f_C and f_r (f_l);

3. if both local maxima do not lie above the 0-dB line, we vertically shift E_{res} until the 0-dB level meets the closest of the two. Then, f_B is calculated as before except if the new notch depth is smaller than D_{min} in the shifted residue plot, in which case the parameter search procedure for the current notch is aborted and N_{nch} is decreased by one.

Note that case 1 may occur simultaneously with respect to case 2 or 3: in this situation, both corresponding effects are considered when calculating f_B.

3.1.3 Multi-notch filter construction

The so found parameters f_C, D, and f_B need to uniquely define a filter structure. To this end, we use a second-order notch filter implementation of the form [14]

$$H_{nch}^{(j)}(z) = \frac{1+(1+k)\frac{H_0}{2}+l(1-k)z^{-1}+(-k-(1+k)\frac{H_0}{2})z^{-2}}{1+l(1-k)z^{-1}-kz^{-2}}, \tag{6}$$

where

$$k = \frac{\tan(\pi\frac{f_B}{f_s}) - V_0}{\tan(\pi\frac{f_B}{f_s}) + V_0}, \tag{7}$$

$$l = -\cos(2\pi\frac{f_C}{f_s}), \tag{8}$$

$$V_0 = 10^{\frac{D}{20}}, \tag{9}$$

$$H_0 = V_0 - 1, \tag{10}$$

and f_s is the sampling frequency. Using such an implementation allows us to fit our parameters directly to the filter model. Clearly, not every combination of the three parameters is accurately approximated by the second-order filter: if the notch to be compensated is particularly deep and sharp, the filter will produce a shallower and broader notch, having a center frequency which is slightly less than f_C.

Although moderate frequency shift and attenuation is not detrimental to the estimation algorithm (an underestimated notch will be fully compensated through the following iterations), an excessive notch bandwidth could lead to undesired artifacts in the final resonance spectrum. Here is where parameter ρ comes into play: if we divide f_B by $\rho > 1$, the new bandwidth specification will produce a filter whose notch amplitude will be further reduced, allowing us to reach a smaller bandwidth. Typically, in order to achieve a satisfactory trade-off between the size of ρ and the number of iterations, we set it to 2.

Consequently, the parameters to be fed to the filter are $(f_C, D, f_B/\rho)$, yielding coefficients vectors $\boldsymbol{b}^{(j)}$ and $\boldsymbol{a}^{(j)}$ for $H_{nch}^{(j)}$. We iterate the parameter search and notch filter construction procedures for all N_{nch} notches. In order to build the complete multi-notch filter H_{nch},

$$H_{nch}(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{a_0 + a_1 z^{-1} + a_2 z^{-2}} = \prod_{j=1}^{N_{nch}} H_{nch}^{(j)}(z), \tag{11}$$

it is now sufficient to convolve all the coefficient vectors computed during iteration i:

$$\boldsymbol{b} = [b_0, b_1, b_2] = \boldsymbol{b}^{(1)} * \boldsymbol{b}^{(2)} * \cdots * \boldsymbol{b}^{(N_{nch})} \tag{12}$$

$$\boldsymbol{a} = [a_0, a_1, a_2] = \boldsymbol{a}^{(1)} * \boldsymbol{a}^{(2)} * \cdots * \boldsymbol{a}^{(N_{nch})}. \tag{13}$$

Finally, before considering the next iteration, we must update the global multi-notch filter $H_{refl}^{(i+1)} = H_{refl}^{(i)} \cdot H_{nch}$ and compensate the PRTF by applying $H_{res}^{(i+1)} = H_{res}^{(i)} / H_{nch}$.

3.1.4 Algorithm evolution example

Figure 2 illustrates the algorithm's evolution for a particular PRTF. The specific choice of the initial parameters was $N_{ceps} = 4$, $D_{min} = 0.1$ dB, and $\rho = 2$. The top left panel illustrates Subject 010 PRTF for an elevation of -45 degrees. The bottom left panel reports the spectral envelope evolution, where we can see how interfering spectral notches negatively influence the initial estimate. The panels on the right represent the resonant (H_{res}) and reflective (H_{refl}) parts of the PRTF at the end of the algorithm.

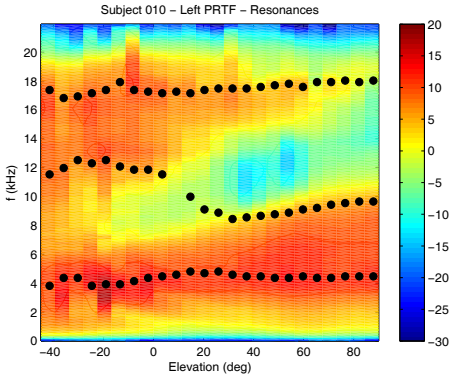
Consider the range where acoustic effects of the pinna are relevant, i.e. the range from 3 to 18 kHz approximately. Figure 2 shows that inside such range the algorithm has produced a realistic decomposition: the gain of the reflective component is unitary outside the notch regions, while the peaks appearing in the resonant component have a good correspondence to Shaw's modes (this point is further discussed in the next section). Outside the relevant range for the pinna, there is a sharp gain decrease in the resonant part and further imperfections that appear for different subjects and elevations. Nevertheless, this is not a problem as long as we consider the pinna contribution to the HRTF alone.

The behavior exemplified in figure 2 is observed for different elevations and subjects too.

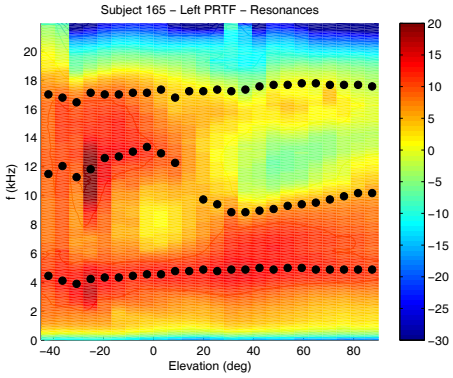
3.2 Resonances

We now discuss the PRTF features identified by the decomposition carried out through the separation algorithm. From the 3-D plots in Figure 3 we can study how the resonances' contribution for Subjects 010 and 165 varies throughout all available elevations. The center frequency of each resonance was extracted with the help of an identification system based on a sixth-order ARMA model [15] and spatially tracked along elevation, resulting in the dotted tracks superposed on the plots.

We can easily identify two major hot-colored areas in these plots. The first one, centered around 4 kHz, appears to be very similar amongst subjects since it spans all elevations. One may immediately notice that this area includes



(a) Subject 010.



(b) Subject 165.

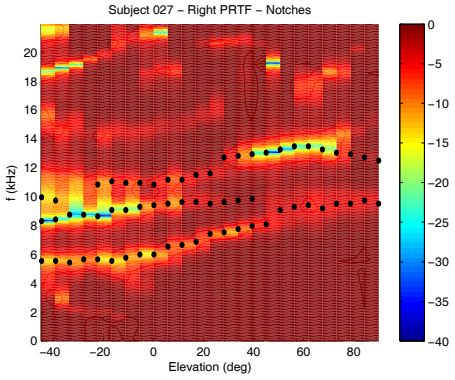
Figure 3. Resonance plots for different elevations.

Shaw's omnidirectional mode 1. The resonance's bandwidth appears to increase with elevation; however, knowledge of pinna modes implies that a second resonance is likely to interfere within this frequency range, specifically Shaw's mode 2 (centered around 7 kHz with a magnitude of 10 dB). On the other hand, the second hot-colored area differs both in shape and shade amongst subjects. Still it is most prominent at low elevations between 12 and 18 kHz, a frequency range which is in general agreement with Shaw's horizontal modes 4, 5, and 6.

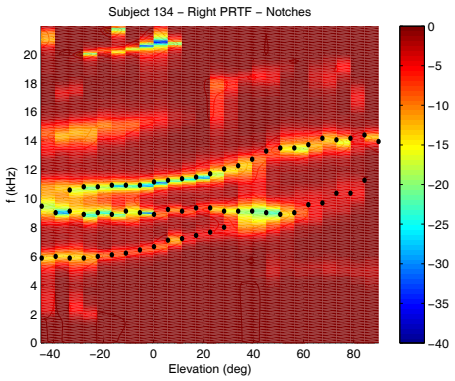
Note that the higher resonance may be perceptually irrelevant since it lies near the upper limit of the audible range. In addition, since the resonances at 12 and 7 kHz are excited in mutually exclusive elevation ranges, we may look forward to a double-resonance filter design.

3.3 Notches

Similarly to the resonance plots, those in Figure 4 represent the frequency notches' contribution for Subjects 027 and 134. As expected, reflection patterns strongly depend on elevation and pinna shape. While PRTFs generally exhibit poor notch structures when the source is above the head, as soon as the elevation angle decreases the number and



(a) Subject 027.



(b) Subject 134.

Figure 4. Spectral notch plots for different elevations.

depth of frequency notches grows to an extent that varies among subjects.

However, several analogies can be noticed here too. In order to investigate such common trends, we inherit an analysis tool that is widely used in the field of sinusoidal modeling, specifically the McAulay-Quatieri partial tracking algorithm (see [16] for details), to track the most prominent notches' patterns along all elevations. Originally, this algorithm was used to group sinusoidal partials (extracted through a peak detection algorithm) along consecutive temporal windows according to their spectral location. We implemented the original version [16] of the algorithm; obviously, since in our case elevation dependency replaces temporal evolution and spectral notches take the role of partials, we call it "notch tracking" algorithm. The notch detection step simply locates all of the local minima in the reflective component's spectrum, while the matching interval for the notch tracking procedure is set to $\Delta = 3$ kHz.

Since it is preferable to restrict our attention to the frequency range where reflections due to the pinna alone are most likely seen, and ignore notches which are overall feeble, two post-processing steps are performed on the obtained tracks:

- delete the tracks which are born and die outside the range $4 - 14$ kHz;
- delete the tracks that do not present a notch deeper than 5 dB.

The outputs of the notch tracking algorithm are the dotted tracks superposed on the plots in Figure 4. Results are definitely akin to those found in [11] with the use of an elaborated DSP-based algorithm. Three major tracks are seen for both subjects, whereas the shorter track in Subject 027's plot very probably represents the continuation of the missing track at those elevations. Reasonably, the gap between tracks is caused by the algorithm's impossibility of locating proper minima in that region (due e.g. to superposition of two different notches or the presence in the magnitude plot of valleys which are not notch-like). However, the three longer tracks suggest that similar reflection patterns occur in different PRTFs.

4. REFLECTIONS AND ANTHROPOMETRY

We now move to a possible 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.

4.1 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 [11] when tracing reflection points over pinna images based on the extracted notch frequencies.

Nevertheless, Satarzadeh [17] 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)).

4.2 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 [11].

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) [13]. 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

$$d(\phi) = \frac{c}{2 f_0(\phi)}, \qquad (14)$$

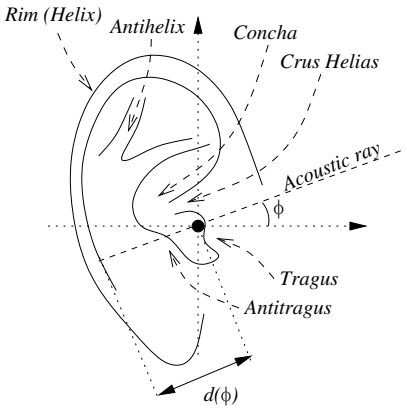


Figure 5. Anatomy of the pinna.

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 [11]. Then, if we consider the 2-D polar coordinate system illustrated in Figure 5 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 6. 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:

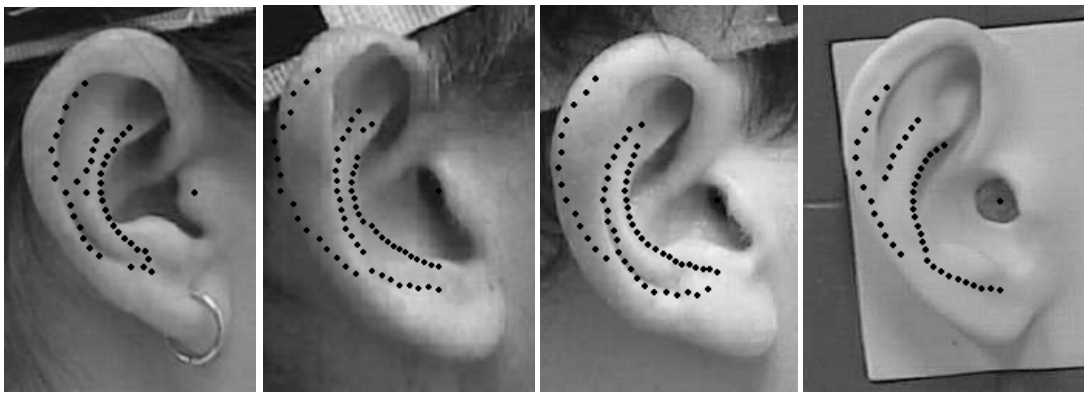
- in the case of subjects 010 and 165 the track is faint and probably due to the antihelix;
- conversely, subjects 027 and 134 present a longer and deeper track, that we visually associate to a reflection on the rim's edge.

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.

4.3 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



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

Figure 6. Reflection points on four CIPIC subjects' right pinnae.

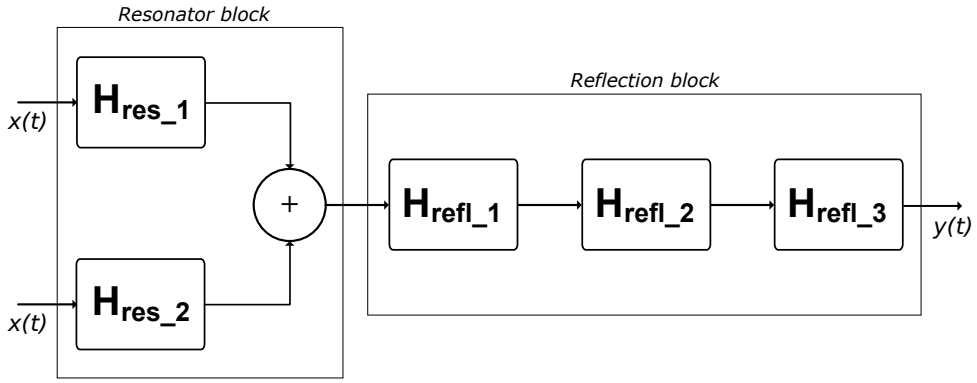


Figure 7. General model for the reconstruction of PRTFs.

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.

5. A STRUCTURAL MODEL OF THE PINNA

The information gathered from the outputs of the decomposition and notch tracking algorithms allows to model the PRTF with two resonances and three spectral notches. As Figure 7 depicts, our final aim is to design two distinct filter blocks, one accounting for resonances and one for reflections. Clearly, in order to reach complete control of the filter parameters, full parametrization of the model on anthropometrical measurements is needed. Hence for the moment we shall present the PRTF re-synthesis procedure driven by the outputs of the two above algorithms.

5.1 Filter design

In Section 3.2 we have shown that a PRTF at one specific elevation includes two main resonances in the frequency range of interest for the pinna. It is then possible to approximate the effective resonances by deducing center frequency f_C and magnitude G of each resonance from the dotted tracks and directly using the so found parameters to design two second-order peak filters with fixed bandwidth $f_B = 5$ kHz of the form [18]

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

where

$$h = \frac{1}{1+\tan(\pi\frac{f_B}{f_s})}, \qquad (16)$$

$$d = -\cos(2\pi\frac{f_C}{f_s}), \qquad (17)$$

$$V_0 = 10^{\frac{G}{20}}, \qquad (18)$$

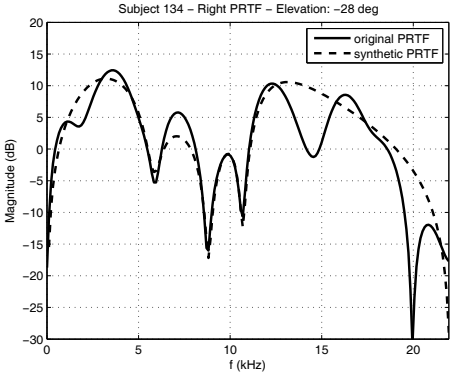
and f_s is the sampling frequency. A posteriori analysis of the synthesized resonances has revealed that PRTFs for high elevations only need the first resonance to be synthesized, being the second very close to it. We thus choose to bypass the second resonant filter when $\phi \geq 20°$.

Similarly, for what concerns the reflection block, we feed the center frequency f_C, notch depth D, and bandwidth f_B parameters coming from the notch tracking algorithm to three second-order notch filters of the form in Eq. 6, each accounting for a different spectral notch. 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.

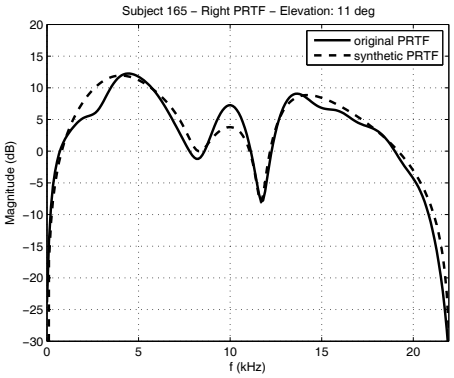
5.2 Results

Figure 8 reports the comparison between original and re-synthesized PRTF magnitudes for three distinct subjects, each at a different elevation. Adherence rate to the original PRTFs is overall satisfactory in the frequency range up to 14 kHz. Still, several types of imperfections need to be adjusted: as a first example, deep frequency notches that appear at low elevations complicate the notch filter design procedure. In point of fact, if the notch to be approximated is particularly deep and sharp, the second-order filter will produce a shallower and broader notch whose bandwidth may interfere with adjacent notches, resulting in underestimating the PRTF magnitude response in the frequency interval between them. Figures 8(a) and 8(b) show this behaviour around 7.5 and 10 kHz, respectively. Using a filter design procedure which forces to respect the notch bandwidth specification during re-synthesis would grant a better rendering of resonances, at the expense of worsening notch depth accuracy.

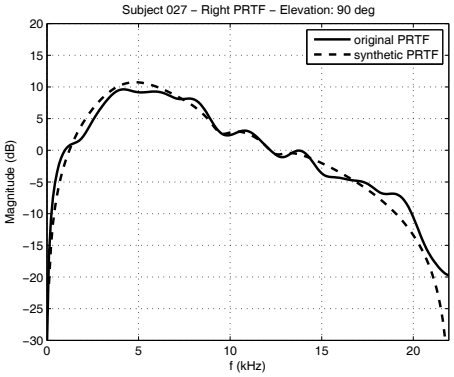
The absence of modeled notches over the upper frequency threshold is another cause of imprecision. For instance, Figure 8(a) presents an evident mismatch between original and modeled PRTF just after the 12.5-kHz peak, due to the cut of the frequency notch at 14.5 kHz. This problem may be corrected by increasing the 14-kHz threshold in order to take into account a higher number of notches. However, being the psychoacoustic relevance of this fre-



(a) Subject 134, elevation $-28°$.



(b) Subject 165, elevation $11°$.



(c) Subject 027, elevation $90°$.

Figure 8. Original vs Synthetic PRTF plots.

quency range relatively low, the effective weight of the mismatch is reduced.

Last but not least, resonance modeling may bring approximation errors too. In particular, the possible presence of non-modeled interfering resonances and the fixed-bandwidth specification both represent a limitation to the re-synthesis procedure. Furthermore, center frequencies extracted by the ARMA identification method mentioned in Section 3.2 do not always coincide with peaks in the PRTF. Thus a stronger criterion for extracting the main parameters of each resonance is needed. Nevertheless, the approximation error seems to be negligible in all those cases where resonances are distinctly identifiable in the PRTF.

In conclusion, the above presented re-synthesis model appears to be overall effective, especially for PRTFs which clearly show one or two main resonant modes and moderately deep notches. Figure 8(c) supports this assertion.

6. CONCLUSIONS AND FUTURE WORK

In this paper we presented an approach for structural PRTF modeling, which exploits an algorithm that separates the resonant and reflective parts of the PRTF spectrum. We used such decomposition to re-synthesize the original PRTF through a low-order filter model, whose results show an overall suitable approximation. In a parallel manner, our attempt towards the explanation of the scattering process resulting in the most important spectral notches in the PRTF provided visually convincing results. Besides improving the synthesis step, ongoing and future work includes understanding of the reflection coefficient and relating the resonant component of the PRTF to anthropometry.

7. 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] D. W. Batteau, "The role of the pinna in human localization," *Proc. R. Soc. London. Series B, Biological Sciences*, vol. 168, pp. 158–180, August 1967.

[7] A. J. Watkins, "Psychoacoustical aspects of synthesized vertical locale cues," *J. Acoust. Soc. Am.*, vol. 63, pp. 1152–1165, April 1978.

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

[9] 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.

[10] 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.

[11] 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, pp. 364–374, July 2005.

[12] 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.

[13] 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), pp. 1–4, 2001.

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

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

[18] S. J. Orfanidis, ed., *Introduction To Signal Processing*. Prentice Hall, 1996.