LINEAR RELATIONSHIPS BETWEEN SPECTRAL CHARACTERISTICS AND ANTHROPOMETRY OF THE EXTERNAL EAR

Sergio G. Rodríguez

University of São Paulo
Signal Processing Laboratory – PSI - EPUSP
Av. Luciano Gualberto 158, tr.3, CEP 05508900
Butantã SP – São Paulo Brazil
sergio@lps.usp.br

Miguel A.Ramírez

University of São Paulo
Signal Processing Laboratory – PSI - EPUSP
Av. Luciano Gualberto 158, tr.3, CEP 05508900
Butantã SP – São Paulo Brazil
miguel@lps.usp.br

ABSTRACT

HRTF individualization is a critical issue in high fidelity virtual auditory spaces using binaural reproduction. Given the relative success of recently published models for the head-torso contribution to the HRTF, we are presenting an initial attempt for estimating the pinna contribution, called here as pinna-related transfer function (PRTF). We use a set of PRTF magnitudes presented in the companion study. They were extracted from a HRTF database, and modeled by means of principal component analysis. First, we calculate the correlation between several spectral features and anthropometric dimensions. These features are principal component weights and central frequencies of pinna notches. Then, we derive other anthropometric pinna parameters in order to improve the correlation. Next, multidimensional linear regression is performed. As a result, linear transformations are calculated by solving the least square problem through QR decomposition. Finally, we are able to collect anthropometric data from new subjects and to estimate approximated individualized PRTFs.

1. INTRODUCTION

Nowadays, individualized HRTFs are becoming an important and critical issue in high fidelity virtual auditory spaces (VAS) reproduced by means of binaural technology. Although applications like acoustic simulation for room acoustics do not always require the exact location of the sound source being replicated for particular individuals, there exist many cases where neuroscience research, virtual reality systems and novel communication applications do need rendering 3D sound in a way that the most real aural experience can be recreated. This means the exact location of the sound sources must be reproduced, without mattering who is using the system. Obtaining individual HRTFs is not a simple task. Conceptually it can be done by acoustical or computational means. The former is the conventional way, and even when it delivers precise results, it demands time, expensive equipment and a lot of knowledge and expertise in acoustical measurements. Although in some cases the task of measuring HRTFs can be accomplished, many applications can not cover these demands, especially when the requirement is real-time or consumer product technology. Here, the computational approach appears to be more suitable. The basic idea is to synthesize the HRTF through computational means starting from non-acoustic data, i.e. anthropometric data. On this respect, besides the numerical calculation of acoustic fields for the body surface, which is currently unavailable, methods for estimating an approximated HRTF or improving a generic one for specific subjects have been proposed since several years [1] [2]. The HRTF describes the scattering of acoustic waves on the body surface; therefore, anthropometric features are intimately related with it. Synthesizing an approximated HRTF from anthropometric data seems to be adequate.

The HRTF presents several features produced by different parts of the body as shown in previous works [3] [4]. While lateral localization is derived from interaural differences as it has been known for a century, elevation localization depends on pinnae (above 3 kHz) and on head/torso response (below 3 kHz). In this paper we present the linear relationships found between anthropometric data and spectral characteristics of the pinna. These spectral characteristics are: principal component weights correspondent to PRTFs extracted from HRTF data, and central frequencies of spectral notches caused by the pinna in the HRTF. The extraction process of these two elements is presented in the companion study [5]. From now on, the former will be called as PCW and the latter as NCF. We use the calculated linear transformations for estimating an approximated PRTF.

2. INDIVIDUAL CORRELATION COEFICIENTS

Figure 1 shows pinna anthropometric dimensions registered in the CIPIC database. As said before, we already count with spectral features representing 64 different ear PRTF magnitudes by elevation angle in the median plane [5]. At the moment we are working at the frontal hemisphere, with 25 elevation angles from -45° to 90°, from which the nine most representative angles are already modeled. We have 20 PCWs per ear PRTF magnitude and 10 anthropometric dimensions per ear. For every single elevation angle in the median plane, we have a 20 x 64 matrix W for the PCWs; and on the other hand we have the constant 10 x 64 matrix D for the pinna anthropometric data of all the ears. We will show results for a case study characterized by a -23° elevation angle and 0° azimuth angle. At this point, it is important to observe that even when right and left ears of the same subject present similar characteristics, they are not the same: anthropometric dimensions vary as well as spectral features of their corresponding PRTF. From previous works [6] we know that the most important dimensions belong to the main pinna cavity: the concha. It seems to be responsible for a great percentage of HRTF variation, especially for the variation of the spectral notches. Then, it is possible that concha-related dimensions present better correlation with spectral features of PRTFs. The correlation coefficients between each PCW and its best anthropometric predictor dimension are shown in Table 1. As we can see in the table, d4 appears 4 times; d10 and d6 3 times; and d3, d1 and d7 2 times each. But this is not a reliable factor

for judging which dimensions are more important. It is necessary to take into account the significance of the PCW, namely the eigenvalue related to the corresponding principal component. Then, we introduce a measure of importance, denoted by,

$$I_i = \frac{\sum_{j=1}^{20} eig_j \left|\rho_{ji}\right|}{\sum_{j=1}^{512} eig_j}, 1 \le i \le 10 \qquad (1)$$

where eig_j is the eigenvalue corresponding to the *j*th principal component, and ρ_{ji} the correlation coefficient between the *j*th PCW and the pinna dimension denoted by the subindex *i*. Subindex *i* goes from 1 to 10 since we have 10 pinna dimensions in total. This time, the importance of each dimension is computed along the 20 PCWs used for modeling the PRTFs. Table 2 shows the results for that computation.

TABLE 1

PCW	Best predictor	Correlation ρ
w1	d5	-0.35
w2	d4	-0.43
w3	d3	-0.31
w4	d10	-0.35
w5	d4	0.15
w6	d1	-0.22
w7	d7	0.31
w8	d3	0.36
w9	d2	0.34
w10	d6	-0.25

Table 1.*Correlation coefficients between the 10 first PCWs of a set of approximated PRTFs and their best predictors among a set of pinna dimensions registered in the CIPIC Database.*

TABLE 2

d1	d2	d3	d4	d5	d6	d7	d8	d9	d10
0.18	0.12	0.15	0.14	0.19	0.11	0.16	0.13	0.19	0.11

Table 2.*Importance measure for the 10 pinna dimensions.*

From table 2 we can see that the most important dimensions are d1, d5 and d9, and just d1 is related with the concha. However, even when correlation coefficients are not so bad for the first four PCWs (which explain 65% of the PRTF magnitude variance), they are still insufficient for attempting a PCW estimation from linear regression.

3. DERIVING NEW ANTHROPOMETRIC PARAMETERS

The problem appears to be the dimensions. Simple dimensions like these may not represent with accuracy the anatomical structures of the pinna. Then, we derive 12 extra parameters from the existenting dimensions. They are listed in table 3. Some of these parameters are bidimensional in nature, like d13 or d18, and others are tridimensional in nature, as d14 and d16. Also, some parameters do not have much physical meaning, like d15 and d17. Anyway, they characterize well the pinna they represent, as we will see later. With these new parameters, we calculate again the correlation coefficients calculated before.

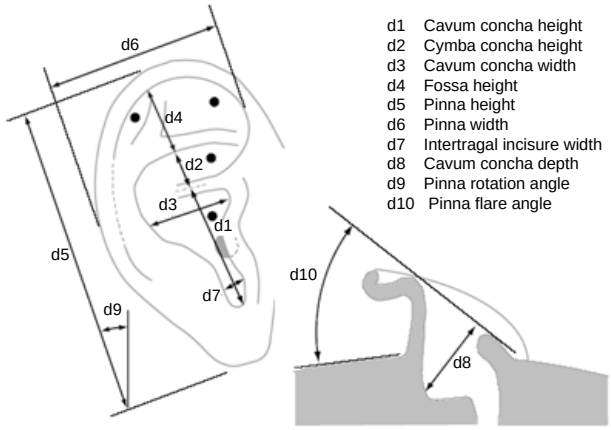

Figure 1.*Pinna dimension registered in the CIPIC database for each subject's ear.*

TABLE 3

Notation	New derived pinna parameters
d11	d1 + d2
d12	d1 + d2 + d4
d13	(d1 + d2).d3
d14	(d1+d2).d3.d8
d15	(d1+d2).d3.d10
d16	(d1+d2).d7.d8
d17	(d1+d2).d7. d10
d18	d1.d3
d19	d5.d6
d20	d5.d6.d8
d21	d5.d6. d10
d22	d4.d6

Table 3. *New derived anthropometric parameters of the pinna.*

TABLE 4

PCW	Best predictor	Correlation ρ
w1	d13	-0.37
w2	d12	-0.45
w3	d18	-0.31
w4	d10	-0.35
w5	d4	0.15
w6	d1	-0.22
w7	d16	0.34
w8	d13	0.4
w9	d2	0.35
w10	d6	-0.25

Table 4.*Correlation coefficients between PCWs of a set of approximated PRTFs and their best predictors among a set of pinna parameters.*

Table 4 shows the results just for the first ten PCWs. As it is obvious, only little improvement is attained. However, now the best predictors, d12 and d13, are better related with the concha, especially d13. Beside of that, we calculate correlation coefficients between pinna parameters and NCFs. The latter are labeled as A, A2, B and C, as it was done in the companion study [5]. Table 5 shows the results. The correlation between NCFs and pinna parameters are stronger, but still poor for performing a linear regression. Then, it seems to be more suitable to sum the contribution of each pinna parameter to the estimation of each spectral feature of the pinna.

TABLE 5

NCF	Best predictor	Correlation ρ
A	d15	-0.46
A2	d8	-0.76
B	d14	-0.57
C	d11	-0.44

Table 5.*Correlation coefficients between central frequencies of pinna notches and their best predictors among a set of pinna parameters.*

TABLE 6

d11	d12	d13	d14	d15	d16	d17	d18	d19	d20	d21	d22
0.2	0.17	0.2	0.2	0.18	0.14	0.18	0.2	0.16	0.16	0.16	0.13

Table 6.*Importance measure for 12 new pinna parameters.*

4. MULTIDIMENSIONAL LINEAR REGRESSION

We have to redefine matrix D, formulated in the second section. Of course, we add the 12 new parameters, making up a 22 x 64 matrix. But another step is needed. We must eliminate columns containing dimensions d2 and d4 from the matrix. The reason is that they can not coexist in the same matrix analysis with d1, d11 and d12 due to the linear dependence involved in the origin of d11 and d12. Due to this linear dependence the resulting matrix is singular. Beside of that, they are the least important in that subgroup (d1, d2, d4, d11 and d12), as it can be observed from table 2 and table 6. Table 6 shows the importance measure for the new parameters, calculated according to equation (1). Now, the most important parameters are d18, d13 and d14, all of which are considerably well related with the concha cavity. Finally, we are ready to perform the muldimensional linear regression by solving the following equation:

$$\mathrm{D}^{\mathrm{T}}.\mathrm{T} + \varepsilon = \mathrm{W}^{\mathrm{T}}, \tag{2}$$

where the 21 x 64 matrix D corresponds to the pinna parameters (a row of 1's had to be added in order to get a constant term in the model), the 10 x 64 matrix W corresponds to the PCWs, ε is a residual matrix and the 21 x 20 matrix T corresponds to the linear transformation to be calculated. Each column of T is calculated by solving the least square problem for the matrix D^{T} and each column of W^{T}, as follows:

$$\mathrm{D}^{\mathrm{T}}.t_i + \varepsilon_i = w_i, \tag{3}$$

where t_i and ε_i denote the *i*th column of matrix T and ε respectively, and w_i the *i*th column of matrix W^{T}, which corresponds to the vector formed by the *i*th PCW of all the ear's PRTF magnitude. QR decomposition is applied in order to solve each least square problem. At the end, the complete matrix T is found and used for estimating PCWs, by solving,

$$\mathrm{D}^{\mathrm{T}}.\mathrm{T} = \hat{\mathrm{W}}^{\mathrm{T}}, \tag{4}$$

where $\hat{\mathrm{W}}^{\mathrm{T}}$ is the estimated matrix of PCWs. Table 7 shows the correlation coefficients between the 20 original PCWs and their estimated versions.

TABLE 7

PC weight	Estimated PC weight	Correlation ρ
w1	ŵ1	0.7237
w2	ŵ2	0.7118
w3	ŵ3	0.6064
w4	ŵ4	0.6975
w5	ŵ5	0.6005
w6	ŵ6	0.5969
w7	ŵ7	0.5738
w8	ŵ8	0.7247
w9	ŵ9	0.6612
w10	ŵ10	0.6150
w11	ŵ11	0.6358
w12	ŵ12	0.5445
w13	ŵ13	0.7565
w14	ŵ14	0.4795
w15	ŵ15	0.6230
w16	ŵ16	0.6915
w17	ŵ17	0.5982
w18	ŵ18	0.5482
w19	ŵ19	0.5220
w20	ŵ20	0.5851

Table 7.*Correlation coefficients between original principal component weights and estimated ones*

Exactly the same process is performed for each NCF. Since just the A and B notches are found in almost all the HRTFs, we estimate just these two NCFs. In this case, given the scalar nature of the NCFs, just an estimation vector is calculated. Thus, vectors of length 21 are found, calling them as u_{A} for the A notch, and u_{B} for the B notch. The correlation coefficients between the NCFs of the A and B notches and their estimated versions are 0.74 and 0.8 respectively. As it is shown, the multidimensional linear regression performs a great improvement on the initial result.

Finally, an accuracy measure for the PCW estimation is done by calculating an estimation index, similar to the importance measure of equation (1):

$$E = \frac{\sum_{j=1}^{20} eig_j \left|\rho_j\right|}{\sum_{j=1}^{512} eig_j}, \tag{5}$$

where ρ_j is the correlation coefficient between the *i*th PCW and its estimated version and eig_j the eigenvalue corresponding to that principal component. For our case study, *E* results in 66% of total estimation. However, there is one more way for improving this result. Since the estimated NCFs present better correlation with their original versions, than PCWs do, we combine these two estimations in just one magnitude response.

5. THE MOVING-NOTCH ALGORITHM

Once we estimate the PCWs for a new subject, we reconstruct his corresponding PRTF magnitude, through the principal component reconstruction process. But in order to improve the result, we use the estimated NCFs for adjusting the notch positions. The idea is to find the closest local minimum of the estimated PRTF magnitude to each of the two estimated NCFs considered, A and B. Now, we are going to assume that the estimated central frequency is on the left of the local

minimum, with respect to the frequency axis, as it is shown in figure 2. Then we calculate the difference in frequency points between the estimated central frequency and the local minimum, and denote it as *x*. Thus, *x* frequency points to the left of the estimated central frequency and *x* to the right of the local minimum, we mark our zone of influence with 1 and 2 respectively. Now, from point 1 to the local minimun we perform an operation of decimation of the PRTF magnitude in order to convert those 2*x* frequency points in just *x* frequency points. On the same way, from the local minimum to point 2, we perform a task of linear interpolation in order to convert those *x* frequency points in 2*x* frequency points. This process is performed for A and B. Figure 3 shows the PRTF magnitudes involved in all the process: extracted original version, modeled version (by 20 principal components), estimated version using just the PCWs estimation and notches-adjusted version. As a final step, minimum phase reconstruction is applied to the estimated PRTF magnitude.

6. CONCLUSION AND FUTURE WORKS

It is possible to estimate with an encouraging degree of precision an individualized PRTF. However, the method can be largely improved with extended available data. On this respect, it is obvious that better correlated pinna parameters can be calculated, for example, through calculation of determined areas extracted from a pinna photography. Image processing can be very usefull through automatic segmentation processes. On the other hand, auditory localization tests are needed in order to validate the method. The estimated PRTFs should, at least, improve the performance of generic HRTFs. One more field of analysis is the error estimation matrix, ε, which may be mathematically modeled.

ACKNOWLEDGEMENTS

This work is partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) under grant No 134310/03-9, and by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) under grant No. 04/13206-7.

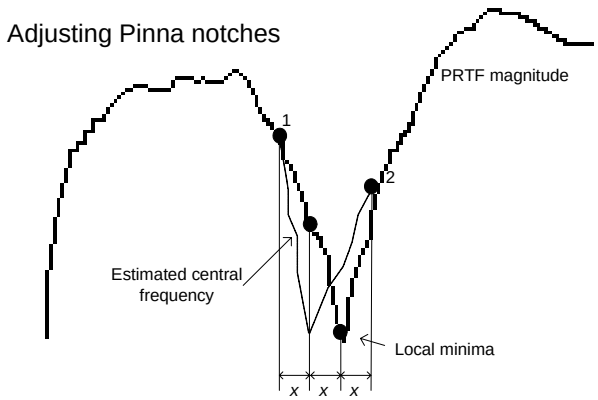

Figure 2.*Adjustment pinna notches process through the moving-notch algorithm.*

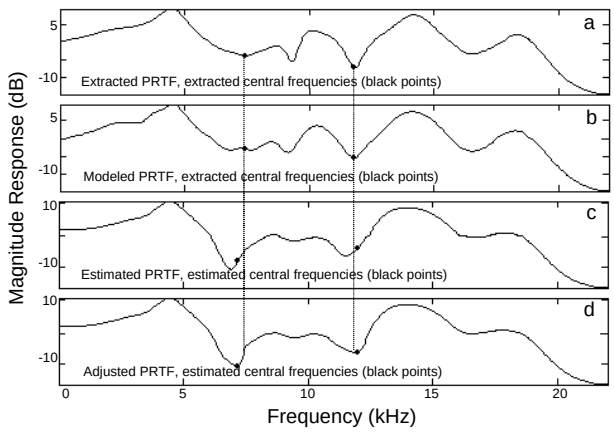

Figure 3.*Extracted, modeled (PCA), estimated and notches-adjusted PRTF corresponding to subject 156, left ear, CIPIC Database.*

7. REFERENCES

[1] J.C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency", *Journal of Acoustical Society of America*, vol. 106, no. 3, pp.1480-1492, 1999.

[2] D. Zotkin, J. Hwang, R. Duraiswami, L. Davis, "HRTF personalization using anthropomentric measurements", *Proceedings of the 2003 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, pp. 157-160, 2003.

[3] V.R. Algazi, R.O. Duda, R.P. Morrison, D.M. Thompson, "Structural composition and decomposition of HRTFs", *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, pp.103–106, 2001.

[4] V.C. Raykar, R. Duraiswami, B. Yegnanarayana, "Extracting frequencies of the pinna spectral notches in measured head related impulse responses", *Technical report CS-TR-4609, Perceptual Interfaces and Reality Laboratory, University of Maryland*, 2004.

[5] S.G. Rodriguez, M.A. Ramirez, "Estimating and modeling approximated pinna-related transfer functions from HRTF data", to appear in *Proceedings of the 11th Meeting of the International Conference of Auditory Display, Limerick, 2005*

[6] E.A. Lopez-Poveda, R. Meddis, "A physical model of sound diffraction and reflections in the human concha", *Journal of the Acoustical Society of America*, vol.100, no.5, pp.3248-3259, 1996.