EXTRACTING AND MODELING APPROXIMATED PINNA-RELATED TRANSFER FUNCTIONS FROM HRTF DATA

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

This paper addresses the extraction and modeling processes of the pinna-related transfer function (PRTF) from a public HRTF database. These processes are essential for HRTF automatic individualization from an anatomically structural point of view. Given that the PRTF is formed by the main resonances and notches of the HRTF magnitude, we approximate it in three steps: The six main resonances are tracked by a HRTF autoregressive model. Next, central frequencies of the main notches are located through a recently published algorithm slightly modified for our purposes. Finally, the notch contours are reconstructed and added to the HRTF envelope using the original HRTF values in the region of the central notch frequencies and their vicinity. Finally, principal component analysis of the PRTF magnitudes for the same spatial position is performed for all the subjects in the database.

1. INTRODUCTION

The head-related transfer function (HRTF) has been the focus of various research efforts in spatial hearing for several years. Being fundamental for spatial audio reproduction and, specifically, for virtual auditory spaces (VASs) involving binaural reproduction, the complete characterization of HRTFs for particular individuals remains relatively inaccessible. It is well know that the use of generic HRTFs produces important sound quality degradation in VASs. Even worst, when the requirement is high fidelity (HI FI), the use of generic HRTFs is unthinkable. The main identified problems are poor elevation characterization of the virtual sound, front-back confusions and in-head localization when playback is performed by headphones. In HI FI VASs it is essential to overcome these problems. Thus, two tasks must be accomplished: To individualize the HRTFs used; and to compensate individually the transducers used for the reproduction. Here we will take care of the first problem.

The conventional way of obtaining individualized HRTFs is by measuring them through a very tedious, expensive and time-consuming process. Therefore, the search for methods allowing automatic HRTF individualization from non-acoustic data is, nowadays, an open problem. On this respect, since the middle nineties, several new approaches have appeared like structural modeling [1], user-defined spectral manipulation [2], scaling frequency methods [3], numerical computation [4], morphological estimation [5], database matching [6] and external ear parameterization [7]. Among these, numerical calculation promises the most accurate results in the future. It involves the calculation of the exact HRTF for a particular individual, solving the wave equation for his body surface. However, at the moment available computation power is not enough for this purpose, and simpler approaches that just estimate the HRTF, prevail.

In this paper we address the first stage of a method combining structural modeling and morphological estimation approaches. The goal is to extract and to model an approximated pinna related transfer function (PRTF) from HRTF data. The PRTF is the acoustic response of the external ear if it would be isolated, without head and torso attached. As shown in several works [8] [9], the PRTF is formed by the main resonances and notches of the HRTF magnitude, and becomes particularly essential for elevation localization in frequencies above 3 kHz. In this context, individualized PRTF can reproduce the frequency location and relative magnitude of resonances and notches for particular individuals.

2. PROPOSED SOLUTION

Our solution is based on the structural approach. It assumes that different anatomical parts such as pinnae, head, torso and shoulders contribute to different temporal and spectral features in the HRTF [9]. While lateral localization is accomplished by processing interaural differences, elevation localization is a more subject-dependent process since it depends on finer anatomical details. The pinnae are said to be responsible for elevation localization above 3 kHz [8]. On the other hand, it has already been shown that the head and torso effect can provide auditory cues for elevation localization below this frequency [10]. We take here, as a case study, the median plane, where there are neither interaural differences nor important head diffraction effects. Figure 1 shows a typical bidimensional HRTF representation, with elevation and frequency axes, and grayscale levels for the frequency response magnitude. Darker regions represent notches in the frequency response magnitude and brighter regions represent resonances. Some structural characteristics are marked in the figure, such as torso response, recognized there as the arch-shaped pattern; pinna notches as diagonal dark regions starting at 6 kHz; and pinna resonances as horizontal brighter regions starting at 3 kHz. As it can be observed, the variation of the notches with elevation is much noticeable than the one due to the resonances. Beside, there are psychoacoustical, behavioral, and neurophysiological evidence to support the hypothesis that the pinna notch frequencies are important cues for elevation localization, as it has been outlined in [11].

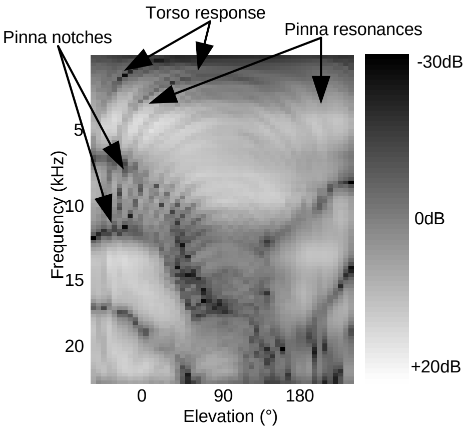

Figure 1.*Typical HRTF bidimensional representation. Structural features are marked.*

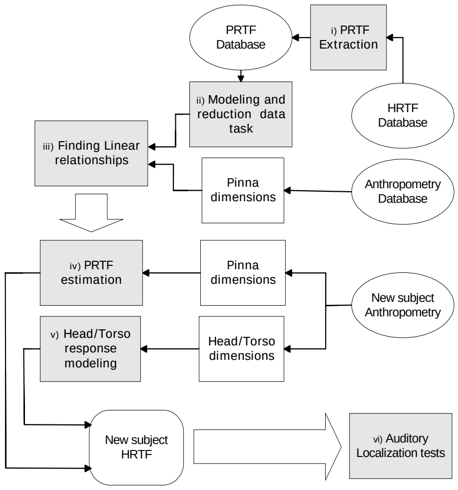

Figure 2.*Proposed solution for individualizing the HRTF.*

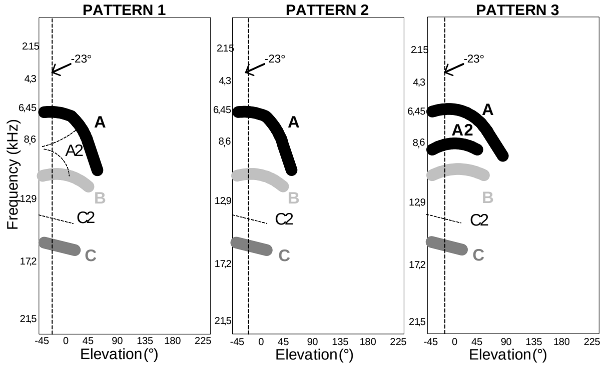

Figure 3.*Patterns found in median plane HRTF bidimensional representations. Vertical dashed lines indicate the spatial position of -23° elevation.*

Obviously, the pinna response or PRTF depends on pinna anthropometric dimensions, which are strictly individual. Thus, we propose a simple solution based on data from a HRTF and anthropometry experimental measuring [12]: i) First, we have to extract an approximated PRTF from the HRTF for all the subjects in the database. ii) Second, we need to model the data in order to reduce it and to separate inter-subject likeness from inter-subject variance. iii) Next, linear relationships between inter-subject PRTF variance and pinna anthropometry are found. iv) Then, we are able to represent an approximated PRTF as a function of certain anthropometric dimensions. These dimensions could be collected from new subjects, and their corresponding PRTF estimated. v) As a complement, the head/torso response can be modeled from anthropometry too [10]. vi) Finally, the method synthesizes the complete HRTF from anthropometry, and it must be validated through auditory localization tests. The whole process is illustrated in figure 2. The present work describes the first two tasks.

3. HRTF PATTERNS IN THE MEDIAN PLANE

Previous to the extraction task we will make important observations about some patterns found in HRTF data, presented in relation to the dimensional representations described before. Taking into account just the diagonal notches caused by the pinna, we were able to identify 3 main elevation–dependent notch contours (called here as *A*, *B* and *C*) and 2 secondary ones (*A2* and *C2*) in the median plane, frontal hemisphere. The main notch contours are already registered in the literature [13]. The presence or absence, and orientation of those notch contours in the image are the factors guiding the classification task. We classify the HRTFs in 3 patterns characterized by the number of notches and the regions where their frequencies may lie. Figure 3 shows the 3 patterns. Carefully observing the whole database of HRTF images, we can say that the first two main notch contours, *A* and *B*, are always present on the HRTFs, and after their appearance in the images at -45° elevation, they evolve until approximately 30° and then disappear. Notch contour *A* is the most diagonal notch contour and it is the longest one too; it appears relatively horizontal with respect to the image, until 0° of elevation, and then falls. The two other main notch contours are more horizontal, always with respect to the image. At -45° elevation, notch contour *A* always appears between 4.5 and 7 kHz. Notch contour *B* appears between 8 and 12 kHz but at the most cases it appears between 10 and 11 kHz. Notch contour *C* is the notch contour which is most variable in vertical localization. It appears between 13 and 17 kHz. But sometimes two notch contours appear in this range and we call the weakest one as *C2*. When *C* and *C2* appear, the first one emerges before 15 kHz and the other one after this frequency. In this case, one of the notches is indistinctly very much weaker than the other. In almost all cases, *C* is the shortest notch contour; often, it disappears after 10°. However, in a few cases notch contour *C* crosses the 30° elevation very slantingly, as well as *A*. It is common that all the notch contours fall with elevation, being *B* the only one that can rise. Anyway, this is not too often. Between notch contours *A* and *B*, almost always an intermediate notch appears. We call it as *A2*. Notch contour *A2* can be very close to *A* or *B* indistinctly and its evolution ends before 0°, and very rarely at 10°.

Pattern 1 has the main notch contours *A*, *B*, *C* and the secondary notch contour *A2* reaching *A* or *B* before the 30° elevation. Pattern 2 is the same but without *A2*. Finally, pattern 3 has the three main notch contours and *A2*, but this time, it reaches neither *A* nor *B*, it evolves independently. Notch contour *C2* can appear anytime in any pattern.

We found that 68.9% of the median plane HRTF images belong to pattern 1; 16.7% to pattern 2, and 12.2% to pattern 3. 2. Just 2% of the images don't belong to any pattern class. All these observations may serve for future works in automatic pattern classification of HRTF images.

Given that the extraction task is rather one-dimensional, knowing the pattern of any particular HRTF allows us to know how many notch contours we must expect, and which is which. This is totally necessary for finding the linear relationships we are looking for, since we will compare the "anthropometric origin" of every notch contour among all the subjects in the database, using linear algebra.

Finally, based on further observation of the entire HRTF database, for spatial positions outside the median plane it would be convenient to model ipsilateral HRTFs. They don't present any head diffraction effect, then, the image is "cleaner", since just the pinna notches are present at medium and high frequencies. The problem appears to be the fusion of several notch contours at different azimuths, depending on the individual. Therefore, to establish a new pattern classification scheme seems to be necessary.

4. EXTRACTING APPROXIMATED PRTF MAGNITUDES

4.1 Extracting pinna resonances and central frequencies of pinna notches

The extraction process is based on a very recently published algorithm for extracting central frequencies of pinna notches [11], along with some algorithms we have developed in order to completely reconstruct the PRTF magnitude. First, we apply linear prediction analysis to the HRTF database.

The autoregressive model tracks well the HRTF envelope, and thus, the resonances. For our purposes, the order of the model, p, has to be 12, since we are looking for 6 resonances; all of which are caused by the pinna as it was shown in [14]. If p is higher, spurious resonances may appear. For example, analyzing spatial position -23° in the median plane (marked by vertical dashed lines in the patterns of figure 3), the autoregressive model tracks acceptably well the resonances in 87 % of the cases, considering 90 HRTFs, corresponding to 45 subjects' left and right ears. The remaining 13% had to be discarded for next stages. The algorithm in [11] first computes the linear prediction (LP) residual of the HRTFs, The LP residual obviates the predictable part of the HRTF, i.e. the resonances, and its frequency response magnitude preserves just the original HRTF notches. Next, it is necessary to eliminate the notches caused by non-pinna parts of the body, like the torso. Then, a windowing operation is performed. Here we use half Kaiser Window. For our example spatial position we use a length of 1ms and the extra parameter of the Kaiser Window, α , equals to 5. These parameters are slightly variable between individuals, and using higher lengths than necessary may produce the preservation of spurious notches. Given that we know how many notches we must expect, we can reduce progressively the length of the window if more notches appear. Anyway this process is sometimes tricky, and we had to supervise the correct extraction. Beside that, the windowing operation reduces the frequency domain resolution, and it may mask or alter the frequencies of the spectral notches [11]. Then, the autocorrelation function of the windowed LP residual is calculated.

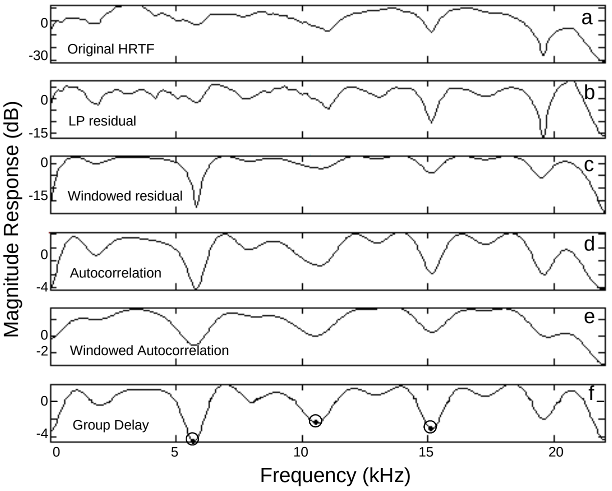

Figure 4.*The whole process for extracting central frequencies of pinna notches for subject 131 of the CIPIC database, right ear HRTF correspondent to -23° elevation, and 0° azimuth. Algorithm was developed in* [11] *and slightly modified by us. In this case, the window length for the LP residual was 1ms and for the autocorrelation function 0.7ms. The α parameter for the Kaiser window was 5 and the Group Delay function threshold, 0dB. Subject 131 right ear HRTF belongs to pattern 2 (notch contours: A, B, and C). Circles indicate a notch central frequency.*

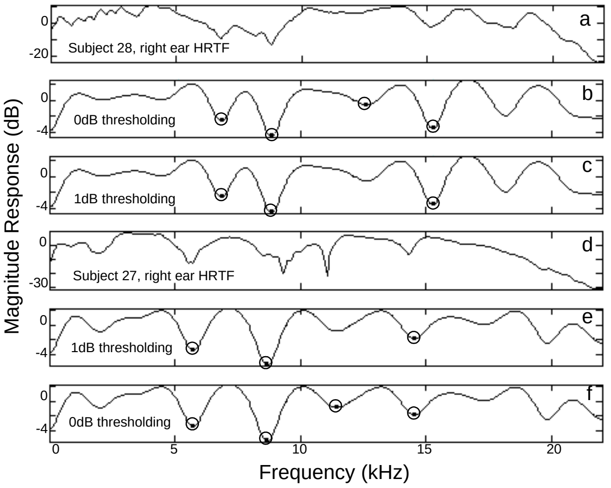

Figure 5.*Pattern classification influence on the extraction process. HRTFs and Group Delay functions corresponding to -23° elevation, and 0° azimuth. While subject 28 right ear HRTF belongs to pattern 2 (notch contours A, B and C), subject 27 right ear HRTF belongs to pattern 1 (notch contours A, A2, B and C). For subject 28 HRTF, using 0dB thresholding for the GD, introduces one spurious notch (between B and C). On the other hand, for subject 27, using 1dB thresholding for the GD, doesn't find the third notch (B). Threshold value is determined by pattern class. Circles indicate a notch central frequency. In the second central frequency of subfigures e and f, some adjustment is required. See figure 6.*

The effect of the size of the window is significantly attenuated when applied to the autocorrelation sequence rather than to the LP residual directly, as it is shown in [11]. Then, again, half Kaiser Window is applied with the same parameters. Next, the group delay (GD) function is calculated for the windowed autocorrelation of the windowed LP residual, since it emphasizes notches and resonances, making it easier to find local minima. Finally we seek local minima, checking slope shifts throughout the frequency range where pinna notches appear, that is from 5 kHz to 17 kHz. Sometimes, we find more minima than necessary, and then we apply a thresholding operation. The threshold depends on the individual and his or her pattern class.

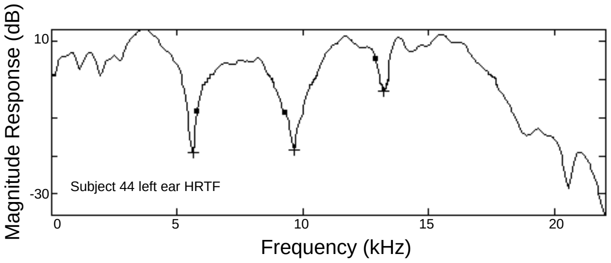

Figure 6.*Ajustment of pinna notch central frequencies based on HRTF local minima and extracted central frequencies of pinna notches. Crosses indicate HRTF minima and black points indicate extracted central frequencies.*

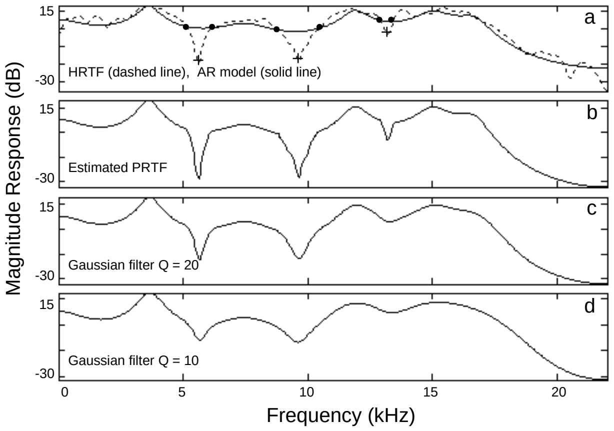

Figure 7.*Approximattion of the PRTF magnitude using the original HRTF, the HRTF AR model and the extracted and adjusted central frequencies of pinna notches. Process corresponds to subject 44 left ear HRTF, at spatial position of -23° elevation and 0° azimuth. In subfigure* a, *crosses indicates a central frequency and black points the coincidences between the HRTF and the AR model.*

As a final step, another improvement is introduced: it is convenient to adjust the central frequencies since sometimes the processing of the signal slightly alters the exact central frequency of the pinna notch. This is done by looking for the local minima in the original HRTF within a close vicinity of the extracted central frequency. Sometimes it coincides with the central frequency and sometimes it is a few frequency points further. Figure 4 shows the whole extraction process. Figure 5 illustrates the improvements made in the algorithm by classifying the HRTFs. Figure 6 shows the adjustment of the central frequencies. At the end, central frequencies of spectral notches caused by the pinna are found.

4.2 Forming approximated PRTF magnitudes

Once we determine the HRTF autoregressive model, preserving the six main HRTF resonances, and the central frequencies of spectral notches caused by the pinna, we are able to combine these two elements for forming an approximated PRTF magnitude. Our algorithm takes into account three elements: the magnitude response of the autoregressive model, the central frequencies of pinna notches, and the original HRTF magnitude. At the beginning of the algorithm, we assume the PRTF magnitude as the magnitude response of the autoregressive model, but at the frequencies corresponding to central frequencies of pinna notches, we constrain it to the values of the original HRTF.

Then, starting from the central frequencies, the algorithm seeks, forward and backward, the coincidence between the HRTF magnitude and the magnitude response of the autoregressive model. Sometimes they never coincide, thus, we give a minimum tolerance for the arithmetical difference between them. Our algorithm uses a tolerance of 0.5dB. If this value is reached, the algorithm stops seeking. Anyway, sometimes the algorithm never reaches this value, and then we put another tolerance for the frequency range within which the algorithm seeks. This tolerance is 500 Hz about the central frequency. This means a total tolerance of 1 kHz for each pinna notch. Thus, the algorithm gives the HRTF magnitude values to the PRTF magnitude at the frequency points surrounding the central frequencies until the searching process stops. As a complement, we apply a Gaussian filter with Q = 20 for smoothing the PRTF magnitude. Figure 7 shows the whole process. Minimum phase reconstruction may be applied for obtaining the complex frequency response, and thus the pinna-related impulse response.

5. MODELING THE PRTF WITH PRINCIPAL COMPONENT ANALYSIS

An extra stage treated here is the modeling process of the PRTF database extracted before. We use principal component analysis since it does exactly what we need. The idea is to eliminate the common part of the PRTFs. But, contrary to other works, we don't perform the PCA between all the spatial positions and all subjects in the database. We work with each spatial position individually. Our goal is to highlight the differences between subjects for the same spatial position. Then, two PRTFs in the median plane per subject serve as input data to the PCA. Considering our case study, i.e. -23° elevation, the first three principal components describe 22.9%, 17.96% and 13.6% of the total variance, respectively. The first ten components describe 92.5% of the total variances. In order to closely approach perfect reconstruction, we take the 20 first components, describing 99.28% of the total variance.

All the PCAs were applied with linear scales for magnitude and frequency. We tried logarithmic scales, and warped scales for frequency, and the decibel scale for magnitude. Besides that, we tried several modeling techniques in order to smooth the magnitude response. These were AR modeling, ARMA modeling, warped AR modeling and the response filtered by Gaussians filters whose Q equals to 8, 10 and 12. We were looking for fewer components to explain a greater percentage of the variance. In general, all these attempts have not made a significant difference with the original case. The HRTFs are sampled at 44.1 kHz, and for the whole process we chose 512 points for the frequency response until the half sampling frequency. As we can see in figure 8 the 20 principal components explain the variance in a very proportional way. This means that there is not a great percentage of variance explained by just one component. The differences between percentages are approximately equidistant until the ninth component. At the end, we already have a reduced representation of each subject PRTF: 20 principal component weights, and several pinna notch central frequencies.

6. SUMMARY AND NEXT STAGES

Classification of pinna notch contours was introduced. These notches are already presented in the literature since several years. The extraction process of the central frequencies of pinna notches was improved by simple manual HRTF pattern classification. However, this process is not totally automatic yet. Some supervision of the result is needed in order to get the correct number and type of pinna notch contours (*A, A2, B* or *C*).

Automatic pattern classification of bidimensional representations of HRTFs may be an interesting task to be accomplished. On the other hand, our hypothesis that the same pinna structures of different subjects produce equivalent notches in the frequency response magnitude remains to be proved. The next stage of the work is to look for linear relationships between anthropometric dimensions of the pinna and reduced data describing the extracted PRTFs. Thus, principal component weights and central frequencies of pinna notches appear to be good options.

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.

7. REFERENCES

[1] C.P. Brown, R.O. Duda, “A Structural Model for Binaural Sound Synthesis”, *IEEE Transactions on Speech and Audio Processing*, vol. 6, no. 5, pp.476-488, Sep. 1998.

[2] Ch. J. Tan, W. S. Gan, “ User-defined spectral manipularon of HRTF for improved localisation in 3D sound systems”, *Electronic Letters*, vol. 34, no. 25, pp. 2387-2389, Dec. 1998.

[3] 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, Sep. 1999.

[4] Y. Kahana, “Numerical modeling of the head-related transfer function”, *PhD.Thesis, University of Southampton*, 2000.

[5] S. Carlile, C. Jin, V. Harvey, “The generation and validation of high fidelity virtual auditory space”, *IEEE Engineering in Medicine and Biology Society*, pp. 1090-1095, 1998.

[6] 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 NY, pp. 157-160, 2003.

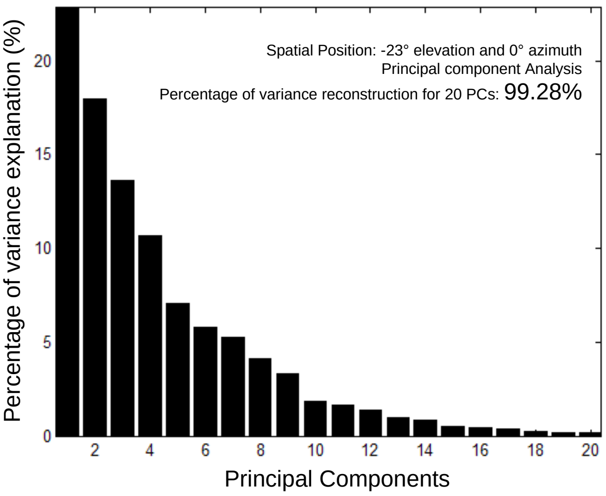

Figure 8.*Resultant Percentage of variance explanation for 20 principal components. The PCA was performed for the approximated PRTFs extracted from 64 HRTFs corresponding to 32 subject’s right and left ears, and to the spatial position of -23° elevation and 0°azimuth, CIPIC Database.*

[7] C. Hetherington, A.I. Tew, T. Yufei, “Three-dimensional elliptic Fourier methods for the parameterization of human pinna shape”, Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 5, pp. 612-615, Apr. 2003.

[8] J. P. Blauert, “Spatial Hearing [Revised edition]”, *MIT Press, Cambridge*, 1997.

[9] 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.

[10] V.R. Algazi, C. Avendano, R.O. Duda, “Elevation localization and head-related transfer function analysis at low frequencies”, *Journal of the Acoustical Society of America*, vol. 109, no. 3, pp.1110–1122, Mar. 2001.

[11] 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.

[12] V.R. Algazi, R.O. Duda, D.M. Thompson, C. Avendano, “The CIPIC HRTF Database”, *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New York, pp.99-102, Oct. 2001.

[13] 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, Nov.1996.

[14] E. A. G. Shaw, R. Teranishi, “Sound Pressure generated in an external-ear replica and real human ears by a nearby sound source”, *Journal of the Acoustical Society of America,* vol. 44, pp.240–249, 1968.