PINNA MORPHOLOGICAL PARAMETERS INFLUENCING HRTF SETS

Slim Ghorbal
3D Sound Labs
Paris, France
s.ghoral@3dsoundlabs.com

Théo Auclair
3D Sound Labs
Paris, France
t.auclair@3dsoundlabs.com

Catherine Soladié
FAST
CentraleSupélec
Rennes, France
catherine.soladier@centralesupelec.fr

Renaud Séguier
FAST
CentraleSupélec
Rennes, France
renaud.seguier@centralesupelec.fr

ABSTRACT

Head-Related Transfer Functions (HRTFs) are one of the main aspects of binaural rendering. By definition, these functions express the deep linkage that exists between hearing and morphology - especially of the torso, head and ears. Although the perceptive effects of HRTFs is undeniable, the exact influence of the human morphology is still unclear. Its reduction into few anthropometric measurements have led to numerous studies aiming at establishing a ranking of these parameters. However, no consensus has yet been set. In this paper, we study the influence of the anthropometric measurements of the ear, as defined by the CIPIC database, on the HRTFs. This is done through the computation of HRTFs by Fast Multipole Boundary Element Method (FM-BEM) from a parametric model of torso, head and ears. Their variations are measured with 4 different spectral metrics over 4 frequency bands spanning from 0 to 16kHz. Our contribution is the establishment of a ranking of the selected parameters and a comparison to what has already been obtained by the community. Additionally, a discussion over the relevance of each approach is conducted, especially when it relies on the CIPIC data, as well as a discussion over the CIPIC database limitations.

1. INTRODUCTION

The HRTFs of a listener are intimately related to his morphology. Thus, a good knowledge of his shape should be a sufficient condition for inferring his HRTFs. Following that idea, many efforts have been done to personalise HRTF sets using anthropometric data. Therefore, the literature is rich of articles exploring this path.

Inoue at al. [1] measured the HRTFs and nine physical features of the head and ears of 86 Japanese subjects. Then, they studied their relationship through multiple regression analysis and used it as an estimation method.

For their part, Zotkin et al. [2] proposed an HRTF personalisation algorithm based on digital images of the ear taken by a video camera. They perform 7 measurements on it and compute out of them a distance between subjects of a given database. The closest match is selected and his HRTFs are used as raw material for the individualisation experiment.

Bilinski et al. [3] propose a method for the synthesis of the magnitude of HRTFs using a sparse representation of anthropometric features. They use a super-set of the features defined by the CIPIC database and learn the sparse representation of subjects of a training database. Then a l1-minimisation problem is solved for finding the best sparse representation of a new subject. This representation is then used for the synthesis of his HRTF set.

However, in these studies, all the parameters are not necessarily independent nor even decisive. Hence, several researchers proposed new means for refining their selection.

Among them, Hu et al. introduced a correlation analysis in two steps [4, 5] prior to the personalisation process. They used the CIPIC database and highlighted significant correlations leading to the selection of only 8 parameters out of the available 27. 5 of them were related to the ear.

Xu et al. [6, 7] retained ten measurements after performing a correlation analysis between the CIPIC HRTFs and anthropometric parameters. It is worth noting that the analysis was restrained to 7 directions and 4 frequencies and that only two of the retained measurements were related to the ear.

Hugeng et al. [8] also realised a correlation analysis over the CIPIC data but ended up in retaining 8 measurements. 4 of them were ear measurements.

Grijalva et al. [9] applied a customisation process of HRTFs using Isomap and Artificial Neural Networks on the CIPIC database. Prior to this, they used the results of [8] as the appropriate morphological parameters to focus on.

While the previous studies added the measurements selection into a wider personalisation process, others exclusively focused on the relative influence of each parameter. This is what did Zhang et al. [10], who concluded after a correlation analysis that 7 ear measurements were among the 8 most significant ones, and Fels et al. [11] who use a parametric model of head, torso and ear for generating new HRTF sets by Boundary Element Method (BEM).

The latter study compared separately the influence of 6 parameters describing the head and 6 others describing the pinna based on the modifications introduced in the HRTFs. The evaluation took into account the spectral distance, the interaural time difference (ITD) and the interaural level difference (ILD) variations. One parameter at a time was modified, ranged in limits derived from anthropometric statistics of adults and children.

Although it provided insights on the relative weights of the parameters in each group, a clear ranking was not established. Moreover, the simulations were limited to frequencies below 8 kHz. This is a major limitation as the human hearing ranges up to 20
kHz and that localisation information due to the pinna is classically comprised between 3-4 kHz and 14-15 kHz or more [12,13].

This multiplicity of works and conclusions reveals an absence of clear consensus about the way each anthropometric parameter modifies - or not - the HRTFs. In the present paper, we take an approach similar as Fels et al. but extended to frequencies up to 16 kHz and establish a categorisation of the pinna parameters.

More in detail, section 2 goes through the process of selection of parameters, the generation of meshes, the computation of HRTFs and the choice and definition of the metrics. Section 3 presents the results themselves, discusses the impact of the chosen metric and establishes a ranking between the retained pinna parameters based on their relative influence over the DTFs. In section 4, we effectively compare our results to the conclusions proposed up to now by the community and lead a discussion over the convergences and points of disagreement. Finally, section 5 sums up the results and conclusions of the present paper and gather the opened questions and perspective of future works.

2. PROTOCOL

2.1. Parameters

2.1.1. CIPIC database

As it is widely used in the community, we have chosen to consider the morphological parameters defined by CIPIC [14]. This database consists in sets of HRTF and morphological parameters measured on 45 subjects. These parameters - 27 in total - are intended to describe the torso, head and ears of the human body with a focus on what could likely impact the HRTF. In particular, 12 parameters describe the position and shape of the ear, the remaining ones describe the head and the body.

It is worth noting that only 35 subjects have been fully measured, meaning that each parameter comes with a set of measures comprised between 35 and 45 values. The database also comes with their mean values $\mu$ and standard deviations $\sigma$.

2.1.2. Selection and values

Based on the sets of ear parameters selected in [11,12,13,10], we retain the set of parameters $\mathcal{P} = \{d_1,d_2,d_3,d_4,d_5,d_6,\theta_1,\theta_2\}$ - defined in figure 2, namely the cavum concha height, the cymba concha height, the cavum concha width, the fossa height, the pinna height, the pinna width, the pinna rotation angle and the pinna flare angle (See table 1). This choice is a result of the number of occurrences of each parameter in these studies, their selection or not as major parameters and our ability to set them with precision in our 3D model (as a reminder, the CIPIC parameters are defined in 2D).

To stick to plausible deformations, for each $p \in \mathcal{P}$, we target the following set of values:

$$\mathcal{P}_p = \{\mu_p + k \ast \sigma_p, k \in [-2,2]\} \quad (1)$$

Moreover, as we intend to study the influence of each $p$ independently, only one at a time can be different from its mean value. When possible - or when it makes sense -, the other CIPIC parameters are set to their mean value. In what follows, we will denote $p^{k}\sigma$ the simulation where parameter $p$ is set to $\mu_p + k \ast \sigma_p$.

<table>
<thead>
<tr>
<th>Var</th>
<th>Measurement</th>
<th>$\mu$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_1$</td>
<td>cavum concha height</td>
<td>1.91</td>
<td>0.18</td>
</tr>
<tr>
<td>$d_2$</td>
<td>cymba concha height</td>
<td>0.68</td>
<td>0.12</td>
</tr>
<tr>
<td>$d_3$</td>
<td>cavum concha width</td>
<td>1.58</td>
<td>0.28</td>
</tr>
<tr>
<td>$d_4$</td>
<td>fossa height</td>
<td>1.51</td>
<td>0.33</td>
</tr>
<tr>
<td>$d_5$</td>
<td>pinna height</td>
<td>6.41</td>
<td>0.51</td>
</tr>
<tr>
<td>$d_6$</td>
<td>pinna width</td>
<td>2.92</td>
<td>0.27</td>
</tr>
<tr>
<td>$\theta_1$</td>
<td>pinna rotation angle</td>
<td>24.01</td>
<td>6.59</td>
</tr>
<tr>
<td>$\theta_2$</td>
<td>pinna flare angle</td>
<td>28.53</td>
<td>6.70</td>
</tr>
</tbody>
</table>

2.2. Morphological model

The model used in this paper is a parametric one, result of a merge between a schematic representation of head and torso - to which we will refer as snowman - although it does not strictly match the one introduced by Algazi & Duda [15] - and an ear model realised thanks to Blender and deformable through multiple blend shapes. The snowman is used in order to add more realism to the generated HRTFs and get closer to what could actually be measured on a real person. The ear model is the true source of interest. It is designed to be as close as possible to real ears, there again for realism. Figure 1 represents the mean ear before and after merge on the snowman - referred as the mean shape.

![Figure 1: Mean ear alone (left) and after merge (right)](image)

Although the CIPIC parameters definition may seem simple at first sight on a 2D drawing - see figure 2 - for them to be fully usable in our 3D environment we need whether to carry out projections of the model on well-chosen plans or to extend them to 3D. Although both options come with drawbacks, the latter seemed more appropriate.

2.3. HRTF generation

Each mesh obtained from the model is then used to feed the FM-BEM computation software mesh2hrtf [16,17]. A virtual source is placed inside the ear canal and virtual microphones are distributed on a sphere of radius 1.2 m whose centre coincides with the centre of the interaural axis. It is worth noting that this sphere is not strictly uniform but slightly denser on the pole than on the equator. Moreover, the directions of elevation inferior to $-60^\circ$ have been excluded from the computations.

The output is a set of 2141 HRTFs computed for every frequency between 100 Hz and 16 kHz by steps of 100 Hz.
2.4. Metrics

In order to compare the variations introduced by the ear deformations in the DTF sets, the following metrics are used:

- The widely-used [1, 8, 9] Spectral Distortion (SD) - also sometimes referred as Log Spectral Distortion - defined as:
  \[ SD = \sqrt{\frac{1}{N} \sum_{k=0}^{N-1} \left( 20 \times \log_{10} \left( \frac{H_1(f_k)}{H_2(f_k)} \right) \right)^2} \text{ [dB]} \]  
  where the frequencies \( f_k \) are regularly spaced on a linear scale.

- The Spectral Distance Measure (SDM), introduced by Huopaniemi [18], corresponding to the SD distance where the frequencies \( f_k \) are regularly spaced on a logarithmic scale.

- The Inter-Subject Spectral Distortion (ISSD) introduced by Middlebrooks [19] and defined as the mean over the directions of the variance of the difference between the DTFs to compare.

- The log-ISSG introduced by Rugeles [20] and corresponding to the ISSD distance where the frequencies \( f_k \) are regularly spaced on a logarithmic scale.

Additionally, the frequency band \([0, 16] \text{ kHz}\) is split into 4 sub-bands of 4 kHz width each.

3. EXPERIMENTAL RESULTS

3.1. General observations

As an introductory example, the ipsilateral DTFs of the simulations \( d_{k\sigma}^\alpha \), \( k \in \{-2, -1, 1, 2\}\) are compared to the ipsilateral DTFs of the mean ear simulation in figure 3. The corresponding ears are gathered in figure 4. As expected, only the high frequencies are affected by the change of the shape (frequencies above 6kHz in the present case).

For coherence between the outputs, the simulations have been gathered into 4 different groups corresponding to the deviation applied to the parameter under study. In practice, it means that for each \( k \in \{-2, -1, 1, 2\} \), the simulations \( d_1^\alpha \), \( d_2^\alpha \), ..., \( d_6^\alpha \) are studied together. Figure 5 presents the impact of a deviation of \(-2 \sigma\) on each parameter, frequency band by frequency band and for each retained metric.

3.2. Metrics choice

An immediate observation is that the log-ISSD (resp. SDM) yields almost the same results as the ISSD (resp. SD). In fact, their average absolute difference varies between 0.05 and 0.13 (resp. 0.05 and 0.11) for each frequency band, while their average values are around 2.2 (resp 2.3).

Another straightforward observation is that all the curves are monotonically increasing, with a low initial value. This is coherent with what has been seen in the introductory example (see fig 4). The ear has almost no effect on the low part of the spectrum, the real gap occurring in the \([4 - 8] \text{ kHz}\) band. Moreover, as expected, we find that the higher the frequencies, the greater the sensitiveness to pinna deformations. Finally, the ranking obtained through the ISSD or through the SD appear to be very similar. Focusing on the \(-2\sigma\) group, we observe indeed in each case a major influence of parameters \( d_3, d_4 \) and \( \theta_2 \) while \( d_1, d_2 \) and particularly \( d_6 \) have
little impact in comparison.

### 3.3. Deformation impact

The figure shows the parameters influence for each applied deviation with respect to the ISSD metric.

As expected, the greater the deviation, the greater the influence of the parameters. In fact, for each parameter, and excepted the band [0 - 4]kHz where the impact on the HRTFs is not significant enough, the norms computed for a deviation of $-2\sigma$ (resp. $+2\sigma$) are greater than the ones computed for a deviation of $-\sigma$ (resp. $+\sigma$).

However, it is worth noticing that these changes are not linear with respect to the deviation. In other words, doubling the deviation will not necessarily imply doubling the metric. In particular, the simulations $d_3^{2\sigma}$ and $\theta_2^{-2\sigma}$ appear to strongly change the HRTFs while the other simulations for these 2 parameters show a moderate or weak influence.

Nevertheless, some regularities exist. This is the case for parameters $d_3$ and $d_4$, which systematically rank among the most influential ones and for $d_1$, $d_2$ and $d_6$ which almost systematically rank among the least influential ones.

![Figure 4: The 4 ears generated for the $d_3$ simulations.](image)

![Figure 5: Parameters’ influence - deviation = $-2\sigma$. From top to bottom, metrics ISSD, log-ISSD, SD and SDM.](image)
4. DISCUSSION

The first conclusions we can draw out of these results have to do with the choice of metric. The use of a logarithmic scale does not bring any additional information to what could be extracted from the ISSD and the SD metrics and cannot be seen as a wiser option. On the contrary, its computation forces to interpolate the data, adding computational complexity and inducing a loss of precision. Moreover, the SD and the ISSD reveal similar trends in the data and can be considered here as equivalent. Nevertheless, it must not be forgotten that the 4 retained metrics derive all from HRTF amplitudes differences and are part of the same family of metrics, making the similarity of their behaviour less surprising.

Regarding the ranking previously established, a comparison to the other studies from the community is presented in table 2. Its analysis leads to the following facts: Xu et al. [6, 7] only retained $\theta_1$ and $\theta_2$ as significant measures. Our results suggest that $\theta$s can effectively be of a certain importance without being major terms. Such a divergence can partly be explained by the ambition behind Xu et al.’s studies. More in details, they performed a correlation analysis over the whole set of 27 CIPIC parameters using the CIPIC database, which only contains 45 subjects - among which 10 do not come with complete sets of measures - while the reliability of such a technique strongly depends on the size of the underlying database. Additionally, they only used a very small subset of the available HRTFs: 7 directions and 4 frequencies.

Hu et al. [4, 5] retained $d_1$, $d_3$, $d_4$, $d_5$ and $d_6$ as main factors. $d_3$ is indeed one of our main factor but $d_1$ and $d_6$ are not. It is worth noticing that their first regression analysis, performed to select the parameters with large correlations with the DTFs did not retain $d_2$, $\theta_1$ and $\theta_2$. This is at least mind confusing if compared to the previous conclusions. It must then be recalled that Xu et al. only used 7 directions and 4 frequencies. However, they also used the CIPIC database and a statistical analysis. Hence, the same remark as the previous one can be done here.

Hugeng et al. [8] retained $d_1$, $d_3$, $d_5$ and $d_6$ as main factors. As in the previous case, $d_3$ is indeed one of our main factors but $d_1$ and $d_6$ are not. However, their framework being very close to...
the one presented by Xu et al., they are also subject to the same remarks.

Zhang et al. [10] have exhibited 8 parameters, among which 6 describe the pinna shape. Namely, they are $d_3$, $d_4$, $d_5$, $d_6$, $\theta_1$ and $\theta_2$. The first 2 are the ones we have seen as the most important here and $d_1$ and $d_2$ are not in their selected set. Nevertheless, $d_6$ is here again presented as a prominent parameter while it completely fails to present this characteristic in our simulations.

Last, Fels et al. [11] retained $d_3$ as the most important factor as well as $d_6$ (out of the scope of this study) and rejected $d_5$, while it was retained by the 2 previous studies. Here, $d_6$ appears to be a good example of non-linearities, as $d_5^{2.52}$ proves to have a strong effect whereas $d_2^{2.52}$ does not.

As it can be observed, the only clear consensus that can be reached is for $d_3$. As it represents the cavum concha width, this conclusion is also coherent with the prior intuition one could have about it.

That being said, another point worthy of interest is the case of parameter $d_6$, twice retained as an important parameter, in total disagreement with our observations. In order to investigate it, the original ear meshes are presented in figure 7 hereafter.

![Figure 7: From left to right, $d_6^{2.52}$, the mean ear and $d_6^{1.25}$.](image)

As we can see, the introduced distortions are not visible from the ear pit, where lies the virtual sound source. Hence, the concha operates as a mask, considerably reducing the potential effect of $d_6$. An immediate consequence is that another value of $\theta_2$ could have yielded a totally different outcome. This fact reveals in particular that not only the parameters’ values are important but so are their combinations, making any statistical analysis more challenging.

5. CONCLUSIONS

In the present work, we have built a database of meshes and computed HRTFs fully dedicated to the study of pinna influence. The anthropometric data were carefully selected to be as relevant as possible. This starts with the use of the CIPIC parameters definition and statistics, known as a reference in the community. Moreover, the choice of the parameters themselves emerges from previous works published in the literature. Furthermore, the anthropometric values as well as the HRTFs generation parameters were set so as to correspond as much as possible to real life problematics and data. In particular, we have covered the whole bandwidth usually said to contain spectral cues. Finally, 4 different metrics have been used to perform the study and to compare the results to previous studies.

Regarding the metrics, it has been shown that they all led to the same conclusions. Thus, one can simply pick and choose the metric that best fits its use case. In our case, retaining the ISSD, parameters $d_3$ and $d_4$ showed a stronger effect on the HRTFs than the other ones while $d_1$, $d_2$ and $d_6$ had, comparatively, much less importance. These conclusions have been confronted to results issued from the community, unveiling a consensus about $d_3$.

In addition, it has been observed that non-linearities exist between the CIPIC parameters and the HRTFs. The specific study of $d_6$ has underscored the need for numerous different ear shapes, i.e. for a bigger database, especially when performing statistical analyses. It also raises the question of the relevance of the parameters choices as introduced by CIPIC, perhaps not perfectly suited for HRTFs analyses, and their definitions, not easily adaptable to 3D data.

Finally, the lack of reachable consensus between the studies aiming at defining a clear set of major parameters also question in general the validity of the studies that present a selection step prior to other treatments (as HRTF individualisation).

However, the current set of data has not delivered all of its information yet. More specifically, future works will investigate the directionality of the impact of each pinna parameter over the HRTFs.

**Table 2: Comparison of our results to the ones of the community.**

<table>
<thead>
<tr>
<th>Authors</th>
<th>Method</th>
<th>Data</th>
<th>Major parameters</th>
<th>Minor parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xu et al.</td>
<td>Correlation analysis</td>
<td>CIPIC database (statistics and HRTFs)</td>
<td>$\theta_1$, $\theta_2$</td>
<td>all others $^i$</td>
</tr>
<tr>
<td>Hu et al.</td>
<td>Multiple Regressions analysis</td>
<td>CIPIC database (statistics and HRTFs)</td>
<td>$d_1$, $d_3$, $d_4$, $d_5$, $d_6$</td>
<td>$d_2$, $\theta_1$, $\theta_2$</td>
</tr>
<tr>
<td>Hugeng et al.</td>
<td>Correlation analysis</td>
<td>CIPIC database (statistics and HRTFs)</td>
<td>$d_1$, $d_3$, $d_5$, $d_6$</td>
<td>all others $^i$</td>
</tr>
<tr>
<td>Zhang et al.</td>
<td>Correlation analysis</td>
<td>CIPIC database (statistics and HRTFs)</td>
<td>$d_3$, $d_4$, $d_5$, $d_6$, $\theta_1$, $\theta_2$</td>
<td>all others $^i$</td>
</tr>
<tr>
<td>Fels et al.</td>
<td>Numerical simulations</td>
<td>Own statistics and Numerical HRTFs</td>
<td>$d_3$, $d_8$</td>
<td>$d_5$</td>
</tr>
<tr>
<td>this work</td>
<td>Numerical simulations</td>
<td>CIPIC statistics and Numerical HRTFs</td>
<td>$d_3$, $d_4$</td>
<td>$d_1$, $d_2$, $d_6$</td>
</tr>
</tbody>
</table>

$^i$ No intermediate category available in these cases. Parameters could only be significant or not.
6. REFERENCES


