PINNA MORPHOLOGICAL PARAMETERS INFLUENCING HRTF SETS

Slim Ghorbal

3D Sound Labs
Paris, France
s.ghorbal@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 l_1-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 μ and standard deviations σ.

2.1.2. Selection and values

Based on the sets of ear parameters selected in [11, 1, 2, 8, 10], we retain the set of parameter $\mathscr{P} = \{d_1, d_2, d_3, d_4, d_5, d_6, \theta_1, \theta_2\}$ - 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 $p \in \mathscr{P}$, we target the following set of values:

$$\mathscr{V}_p = \{\mu_p + k * \sigma_p, k \in [\![-2, 2]\!]\} \tag{1}$$

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 $\mu_p + k * \sigma_p$.

Table 1: *Anthropometric statistics for parameters in $\mathscr{P}$. Distances are in cm and angles in degrees.*

Var	Measurement	μ	σ
d_1	cavum concha height	1.91	0.18
d_2	cymba concha height	0.68	0.12
d_3	cavum concha width	1.58	0.28
d_4	fossa height	1.51	0.33
d_5	pinna height	6.41	0.51
d_6	pinna width	2.92	0.27
θ_1	pinna rotation angle	24.01	6.59
θ_2	pinna flare angle	28.53	6.70

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.

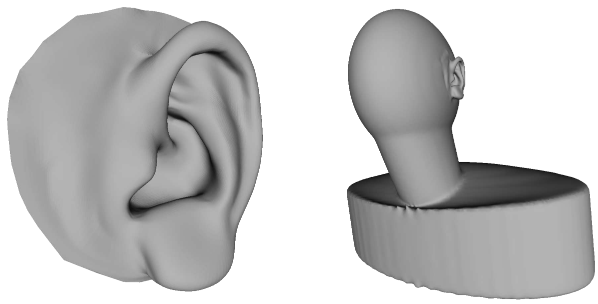

Figure 1: *Mean ear alone (left) and after merge (right)*

Although the CIPIC parameters definition may seem simple at first sight on a 2D drawing - see fig 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 appropriated.

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^o 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.

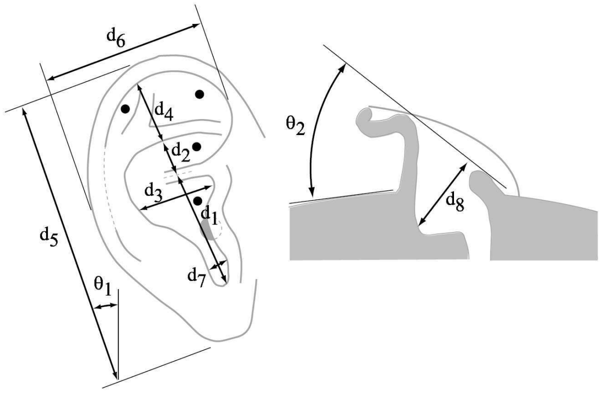

Figure 2: *Pinna measurements definition*

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:

$$SD = \sqrt{\frac{1}{N}\sum_{k=0}^{N-1}\left(20 * \log_{10}\left(\frac{H_1(f_k)}{H_2(f_k)}\right)\right)^2} [dB] \quad (2)$$

  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] kHz is split into 4 subbands of 4 kHz width each.

3. EXPERIMENTAL RESULTS

3.1. General observations

As an introductory example, the ipsilateral DTFs of the simulations $d_3^{k\sigma}, k \in \{-2,-1,1,2\}$ 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 $d_1^{k\sigma}, d_2^{k\sigma}, ..., \theta_2^{k\sigma}$ are studied together. Figure 5 presents the impact of a deviation of -2σ on each parameter, frequency band by frequency band and for each retained metric.

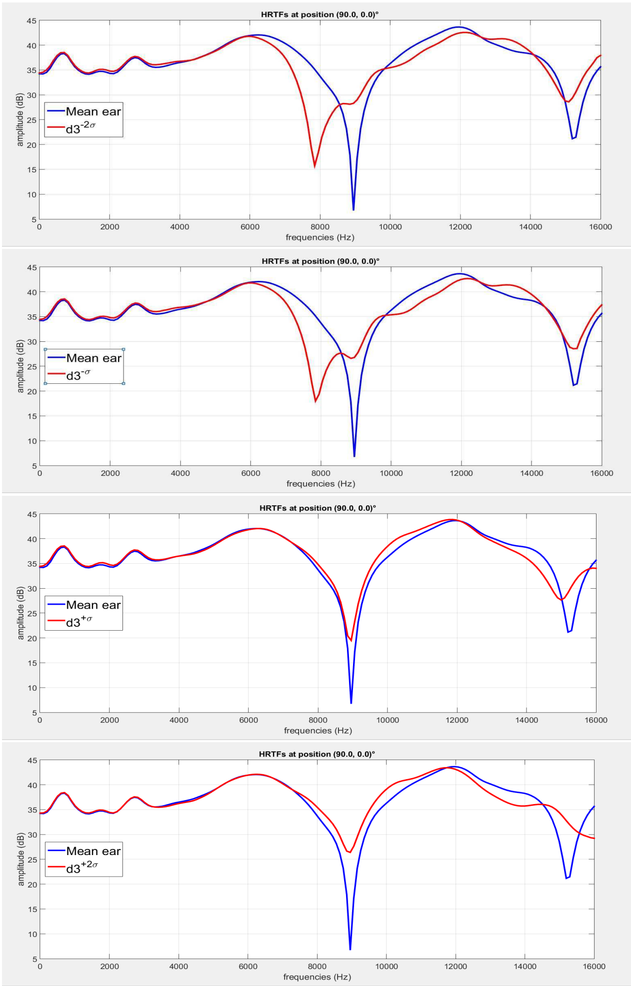

Figure 3: *DTFs in the ipsilateral direction for the mean ear (blue) and for the deformed ear (red).*

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 3). The ear has almost no effect on the low part of the spectrum, the real gap occurring in the $[4-8]kHz$ 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 -2σ group, we observe indeed in each case a major influence of parameters d_3, d_4 and θ_2 while d_1, d_2 and particularly d_6 have

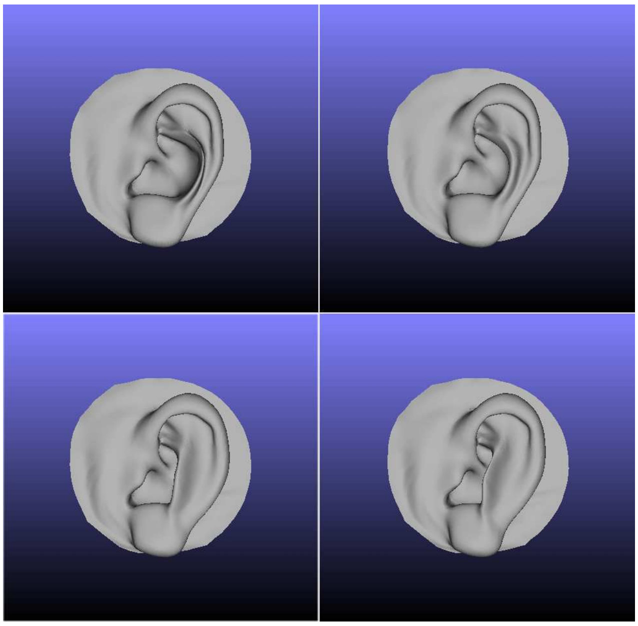

Figure 4: *The 4 ears generated for the d_3 simulations.*

little impact in comparison.

3.3. Deformation impact

The figure 6 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 -2σ (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 $d_5^{+2\sigma}$ 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.

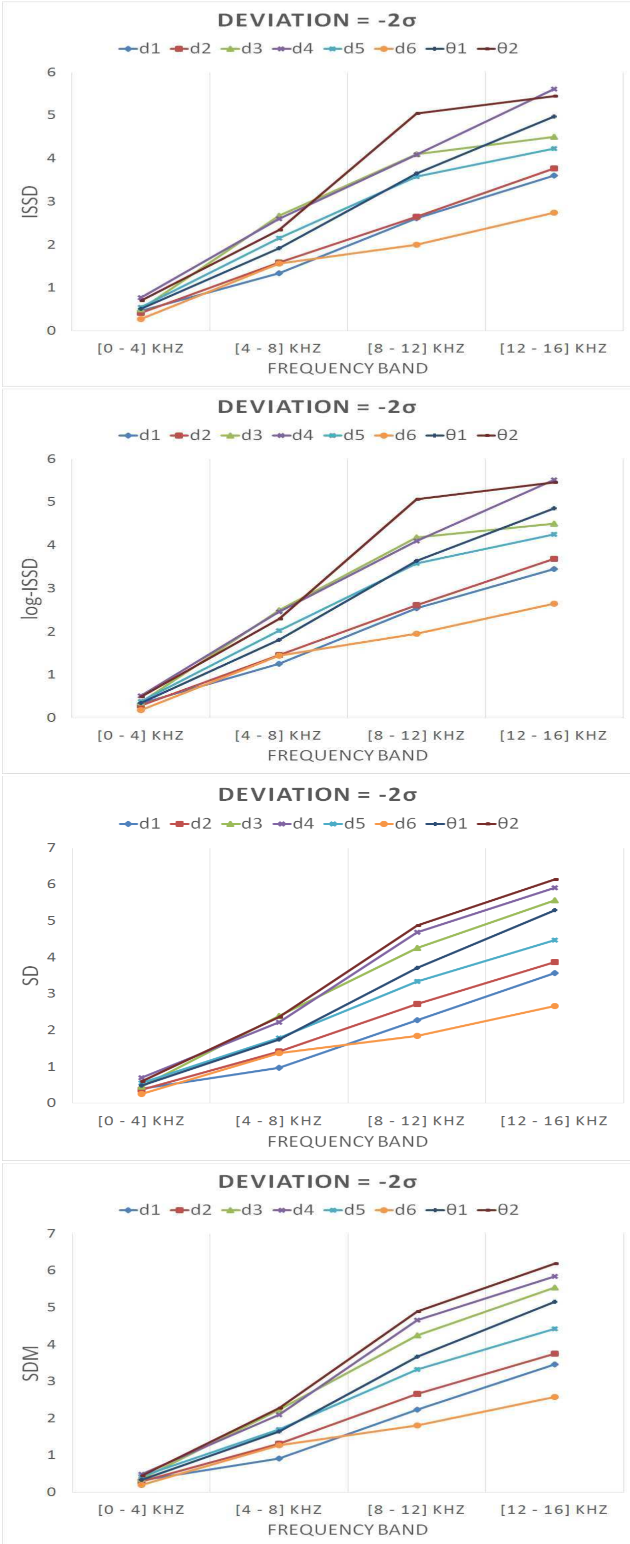

Figure 5: *Parameters' influence - deviation = -2σ. From top to bottom, metrics ISSD, log-ISSD, SD and SDM.*

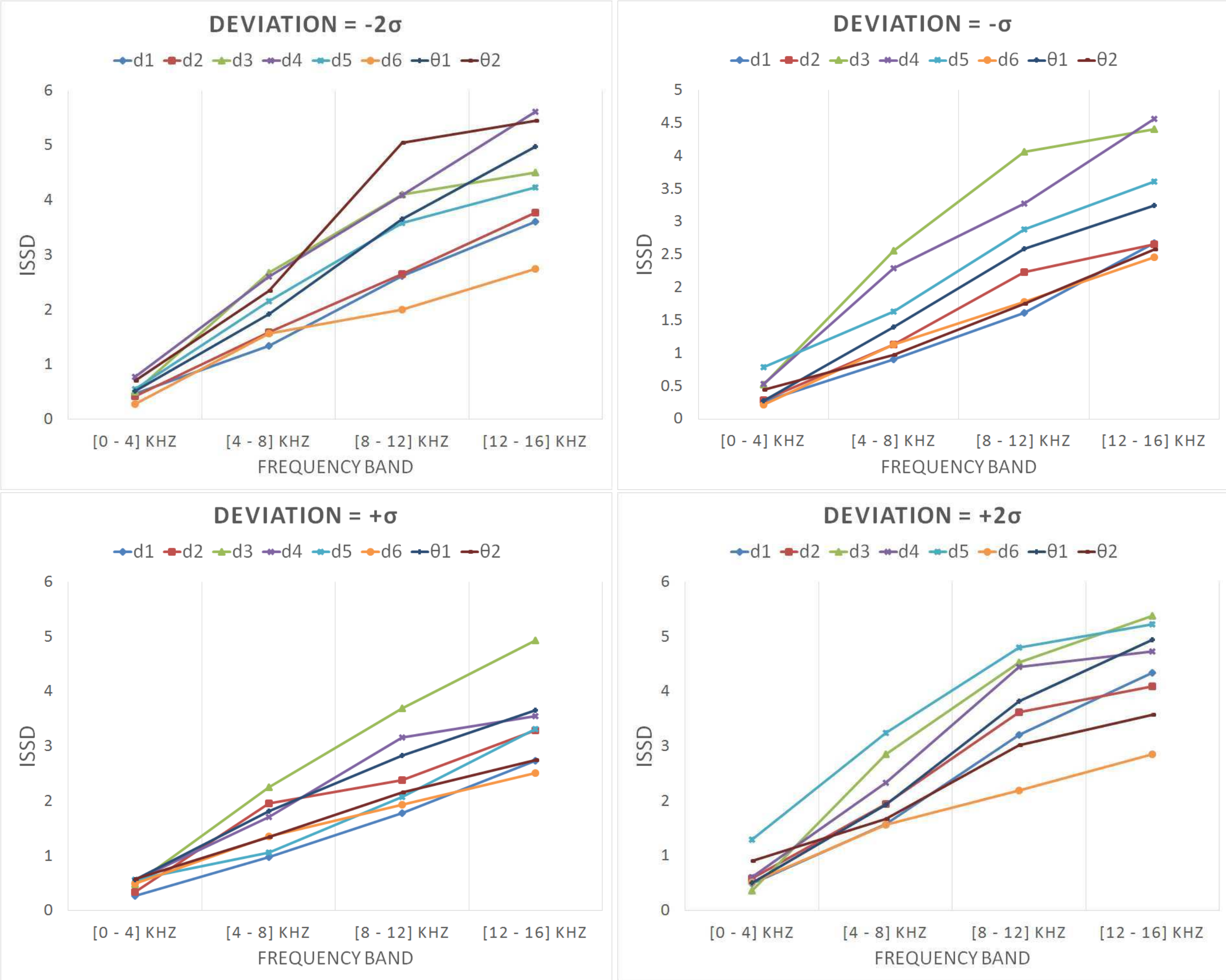

Figure 6: *Parameters influence for deviations* -2σ *(top left),* $-\sigma$ *(top right),* $+\sigma$ *(bottom left) and* $+2\sigma$ *(bottom right).*

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 θ_1 and θ_2 as significant measures. Our results suggest that θs 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, θ_1 and θ_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

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

Authors	Method	Data	Major parameters	Minor parameters
Xu et al.	Correlation analysis	CIPIC database (statistics and HRTFs)	θ_1, θ_2	all others [1]
Hu et al.	Multiple Regressions analysis	CIPIC database (statistics and HRTFs)	d_1, d_3, d_4, d_5, d_6	d_2, θ_1, θ_2
Hugeng et al.	Correlation analysis	CIPIC database (statistics and HRTFs)	d_1, d_3, d_5, d_6	all others [1]
Zhang et al.	Correlation analysis	CIPIC database (statistics and HRTFs)	$d_3, d_4, d_5, d_6, \theta_1, \theta_2$	all others [1]
Fels et al.	Numerical simulations	Own statistics and Numerical HRTFs	d_3, d_8	d_5
this work	Numerical simulations	CIPIC statistics and Numerical HRTFs	d_3, d_4	d_1, d_2, d_6

[1] No intermediate category available in these cases. Parameters could only be significant or not.

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 θ_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_8 (out of the scope of this study) and rejected d_5, while it was retained by the 2 previous studies. Here, d_5 appears to be a good example of non-linearities, as $d_5^{+2\sigma}$ proves to have a strong effect whereas $d_5^{+\sigma}$ 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.

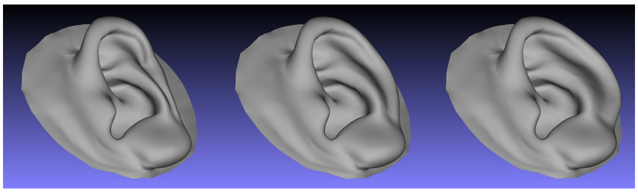

Figure 7: *From left to right, $d_6^{-2\sigma}$, the mean ear and $d_6^{+2\sigma}$.*

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 θ_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.

6. REFERENCES

[1] Naoya Inoue, Toshiyuki Kimura, Takanori Nishino, Katsunobu Itou, and Kazuya Takeda, "Evaluation of hrtfs estimated using physical features," *Acoustical science and technology*, vol. 26, no. 5, pp. 453–455, 2005.

[2] DYN Zotkin, Jane Hwang, R Duraiswaini, and Larry S Davis, "Hrtf personalization using anthropometric measurements," in *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.* Ieee, 2003, pp. 157–160.

[3] Piotr Bilinski, Jens Ahrens, Mark RP Thomas, Ivan J Tashev, and John C Platt, "Hrtf magnitude synthesis via sparse representation of anthropometric features," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2014, pp. 4468–4472.

[4] Hongmei Hu, Lin Zhou, Jie Zhang, Hao Ma, and Zhenyang Wu, "Head related transfer function personalization based on multiple regression analysis," in *2006 International Conference on Computational Intelligence and Security*. IEEE, 2006, vol. 2, pp. 1829–1832.

[5] Hongmei Hu, Lin Zhou, Hao Ma, and Zhenyang Wu, "Hrtf personalization based on artificial neural network in individual virtual auditory space," *Applied Acoustics*, vol. 69, no. 2, pp. 163–172, 2008.

[6] S Xu, ZZ Li, L Zeng, and G Salvendy, "A study of morphological influence on head-related transfer functions," in *2007 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, 2007, pp. 472–476.

[7] Song Xu, Zhizhong Li, and Gavriel Salvendy, "Improved method to individualize head-related transfer function using anthropometric measurements," *Acoustical science and technology*, vol. 29, no. 6, pp. 388–390, 2008.

[8] W Wahab Hugeng and Dadang Gunawan, "Improved method for individualization of head-related transfer functions on horizontal plane using reduced number of anthropometric measurements," *arXiv preprint arXiv:1005.5137*, 2010.

[9] Felipe Grijalva, Luiz Martini, Siome Goldenstein, and Dinei Florencio, "Anthropometric-based customization of head-related transfer functions using isomap in the horizontal plane," in *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 4473–4477.

[10] M Zhang, RA Kennedy, TD Abhayapala, and Wen Zhang, "Statistical method to identify key anthropometric parameters in hrtf individualization," in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on*. IEEE, 2011, pp. 213–218.

[11] Janina Fels and Michael Vorländer, "Anthropometric parameters influencing head-related transfer functions," *Acta Acustica united with Acustica*, vol. 95, no. 2, pp. 331–342, 2009.

[12] Simone Spagnol, Michele Geronazzo, and Federico Avanzini, "Fitting pinna-related transfer functions to anthropometry for binaural sound rendering," in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*. IEEE, 2010, pp. 194–199.

[13] Robert B King and Simon R Oldfield, "The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 39, no. 2, pp. 287–295, 1997.

[14] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano, "The cipic hrtf database," in *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*. IEEE, 2001, pp. 99–102.

[15] V Ralph Algazi, Richard O Duda, Ramani Duraiswami, Nail A Gumerov, and Zhihui Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2053–2064, 2002.

[16] Harald Ziegelwanger, Wolfgang Kreuzer, and Piotr Majdak, "Mesh2hrtf: Open-source software package for the numerical calculation of head-related transfer functions," in *22st International Congress on Sound and Vibration*, 2015.

[17] Harald Ziegelwanger, Piotr Majdak, and Wolfgang Kreuzer, "Numerical calculation of listener-specific head-related transfer functions and sound localization: Microphone model and mesh discretization," *The Journal of the Acoustical Society of America*, vol. 138, no. 1, pp. 208–222, 2015.

[18] Jyri Huopaniemi and Matti Karjalainen, "Review of digital filter design and implementation methods for 3-d sound," in *Audio Engineering Society Convention 102*. Audio Engineering Society, 1997.

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

[20] F Rugeles, M Emerit, and B Katz, "Évaluation objective et subjective de différentes méthodes de lissage des hrtf," in *Cong Français d'Acoustique (CFA)*. CFA, 2014, pp. 2213–2219.