Audio Engineering Society

Convention Paper

Presented at the 120th Convention
2006 May 20–23 Paris, France

Looking for a relevant similarity criterion for HRTF clustering: a comparative study

Rozenn Nicol[1], Vincent Lemaire[2], Alexis Bondu[2], and Sylvain Busson[1]

[1] *France Telecom R&D, TECH/SSTP, Lannion, France*

[2] *France Telecom R&D, TECH/SUSI, Lannion, France*

Correspondence should be addressed to Rozenn Nicol (rozenn.nicol@francetelecom.com)

ABSTRACT

For high-fidelity Virtual Auditory Space (VAS), binaural synthesis requires individualized Head-Related Transfer Functions (HRTF). An alternative to exhaustive measurement of HRTF consists in measuring a set of representative HRTF in a few directions. These selected HRTF are considered as representative because they summarize all the necessary spatial and individual information. The goal is to deduce the HRTF in non-measured directions from the measured ones by appropriate modeling. Clustering is applied in order to identify the representative directions, but the first issue relies on the definition of a relevant distance criterion. The paper presents a comparative study of several criteria taken from literature. A new insight in HRTF (dis)similarity is proposed.

1. INTRODUCTION

Binaural synthesis is a powerful tool for rendering 3D audio scene. Sound spatialization is based on binaural filters derived from the Head-Related Transfer Function (HRTF), which describes the acoustic path between the sound source and the listener's ears. HRTF highly depends on the individual morphology, but acquiring individualized HRTF is still a key issue of current research in binaural technologies. One solution is HRTF measurement, which is quite long and uncomfortable for subjects [1] [2] [3] [4]. What's more individual HRTF measurement should be discarded for commercial use of binaural technologies on a massive scale. Another solution is BEM modeling [5] [6], but this method does not provide accurate modeling in high frequencies because of computational limitations.

A third approach is investigated in the present paper. The idea is to measure HRTF only in a few directions. It is based on data reduction performed by HRTF clustering [7] [8]. The HRTF database is analyzed according to a given criterion of HRTF

similarity focused on the magnitude spectrum[1] of HRTF. As a result, the HRTF are grouped into several clusters, which denotes the main features of HRTF. For each cluster, a representative HRTF is identified as the closest HRTF to all the HRTF contained in the cluster. Therefore, it is intended that one given HRTF in any direction can be deduced from the HRTF measured in the representative directions, which suggests a simplified protocol of HRTF measurement [9]. Several methods, such as HRTF interpolation [8] or neural network modeling [9], are available for deriving HRTF in any direction from the representative HRTFs. This issue will not be dealt with in the present paper, which will be focused on the first step of HRTF clustering.

However, HRTF clustering relies on a similarity or distance criterion, which should be carefully defined according to the data considered. Several distance criteria designed for HRTF are available from literature, however, they are not specific to HRTF clustering. It is intended to compare them when they are applied to HRTF clustering. First, an overview of HRTF (dis)similarity criteria[2] will be given. Five distance criteria are selected. They will be first examined only from the point of view of HRTF (dis)similarity (*A priori* assessment), disregarding clustering purposes. Then their performances for HRTF clustering will be assessed (*A posteriori* assessment), after a brief recall of clustering methodology. The paper will conclude by summarizing the main results of the two studies (*a priori* and *a posteriori* assessments). By merging the two points of view, it will be investigated whether one particular criterion stands out from the others or not.

2. OVERVIEW OF DISTANCE CRITERIA USED FOR HRTF SIMILARITY

The goal of a distance criterion for HRTFs is to quantify the (dis)similarity between two HRTFs. In the present paper, the HRTF (dis)similarity will be judged only from the point of view of signal processing. The HRTFs are compared according to their magnitude spectrum. The distance criterion will be used here for clustering purposes. It is intended to identify common features within the HRTFs of a whole database, in order to sort HRTFs by similarity. Apart for clustering, distance criteria are also required for HRTF modeling purposes, in order to compare the modeled HRTF with the original one [9]. For these various problems, several distance criteria have been defined. An exhaustive list of all the criteria available from literature is beyond the scope of our study. Only five "standard" criteria will be considered in the following.

2.1. Definition of the distance criteria

2.1.1. MSE Criteria

The first criteria, which is certainly the most obvious, is the well-known MSE (Mean Square Error) distance criterion. It is defined as:

$$C_{MSE} = \frac{1}{N}\sum_{i=1}^{N}[H_1(i) - H_2(i)]^2 \qquad (1)$$

where $H_1(i)$ is the magnitude spectrum of one HRTF and $H_2(i)$ that of another HRTF. The index i refers to the frequency index, and N is the number of FFT points.

The MSE criterion can be modified by taking into account the frequency selectivity of the auditory system [10]. Since the auditory ability of frequency analysis is poorer for high frequencies than for low frequencies, it is proposed to lower the high frequencies part by frequency weighting. The frequency selectivity is well described by the concept of the critical bands, the bandwidth of which follows the frequency resolution of the auditory system. The critical bandwidth is 100 Hz for low frequencies (frequencies below 500 Hz) and increases up to 3500 Hz for f=13500 Hz. Its value (in Hz) is given for frequency f (in kHz) by ("Munich" Formula [10]):

$$\Delta(f) = 25 + 75(1 + 1.4F^2)^{0.69}. \qquad (2)$$

Thus the frequency weights $\alpha(i)$ are computed as the inverse of the critical bandwidth:

$$\alpha(i) = \frac{1}{a_0\Delta(f_i)} \qquad (3)$$

[1] The phase spectrum, which is related to temporal cues such as ITD (Interaural Time Difference), is not considered here.

[2] It should be noticed that some criteria used in the following may be not considered as "pure" distance criteria according to a mathematical sense, insofar as they do not fulfill all the properties of a mathematical distance.

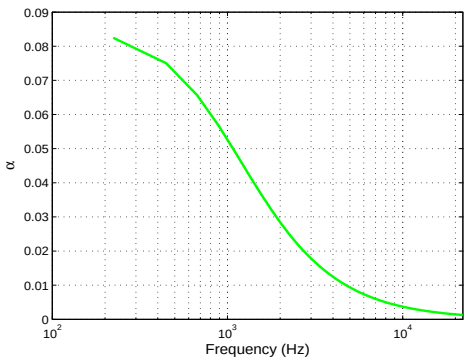

Fig. 1: Frequency weighting according to critical band

where a_0 is a normalization value:

$$a_0 = \sum_{i=1}^{N} \frac{1}{\Delta(f_i)}$$

ensuring that:

$$\sum_{i=1}^{N} \alpha(i) = 1.$$

Fig. 1 illustrates the frequency weights. The MSE criterion including frequency weighting according to the critical bands (which will be referred to as the *CB criterion*) is thus defined by:

$$C_{CB} = \frac{1}{N} \sum_{i=1}^{N} \{\alpha(i)[H_1(i) - H_2(i)]\}^2 \qquad (4)$$

2.1.2. Fahn Criterion

HRTF clustering has been already investigated by Fahn & al for HRTF interpolation purposes [8], a problem very close to our study. The memory cost of binaural synthesis is high if the HRTF measured for all the directions are stored. One solution is to interpolate HRTF in any direction from a limited number of HRTF stored in a few directions. But there are many ways to chose these "useful" HRTF. One of this method is clustering and Fahn & al showed that this latter gave better interpolation than uniform sampling. The performance evaluation was based on a "reconstruction error" defined as:

$$C_F = \frac{\sum_{i=1}^{N}[H_1(i) - H_2(i)]^2}{\sum_{i=1}^{N}[H_1(i)]^2} \qquad (5)$$

This criterion is the third distance criteria used in our study and will be called the *Fahn criterion*. It differs from the MSE criterion mainly by the fact that the MSE distance is weighted by the energy of one HRTF.

2.1.3. Avendano Criterion

The fourth criterion is due to Avendano & al, who have introduced a new "error measure" in a paper about the modeling of the contralateral HRTF [11]. This error is based on the MSE distance expressed on a dB scale:

$$C_A = 10log10 \left\{ \frac{\sum_{i=1}^{N}[H_1(i) - H_2(i)]^2}{\sum_{i=1}^{N}[H_1(i)]^2} + 1 \right\}. \qquad (6)$$

Another advantage of this error criterion is that zero error (i.e. perfect modeling) does not lead to infinity, but to 0 dB, which is more relevant.

2.1.4. Durant Criterion

The last criterion is given by Durant & al in a study about filter design based on Genetic Algorithm for HRTF approximation. The authors have proposed a modified error measure computed as:

$$C_D = \frac{1}{N} \sum_{i=1}^{N} \left\{ 20log10 \left[\frac{H_2(i)}{H_1(i)} \right] - \bar{d} \right\}^2 \qquad (7)$$

with:

$$\bar{d} = \frac{1}{N} \sum_{i=1}^{N} 20log10 \left[\frac{H_2(i)}{H_1(i)} \right].$$

In this criterion, the distance between the two HRTFs is evaluated by magnitude ratio instead of magnitude difference. As the Avendano criterion it is expressed on a dB scale. With the parameter $\bar{d}$, the authors intended to discard the effect of overall gain mismatch. This idea is clever for HRTF modeling, since the reproduction of the main features of the spectral magnitude (i.e. the peaks and notches) is the first goal. Often it is considered that the absolute level is secondary. From Equ. 7, it can also be noticed that the *Durant* criterion is similar to a variance.

2.2. A priori assessment

The previous criteria are first examined in order to assess how they account for HRTF (dis)similarity.

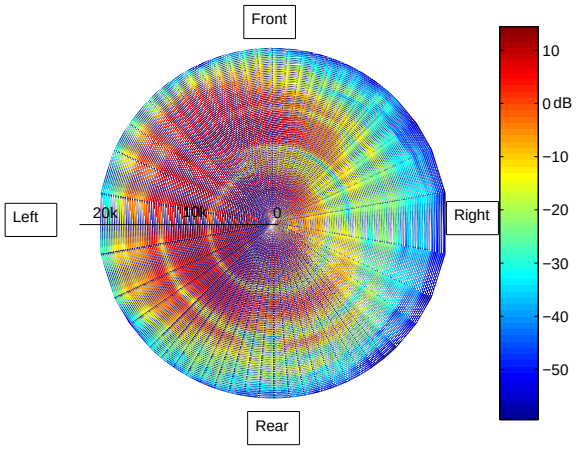

Fig. 2: HRTF magnitude spectrum (dB) in function of the azimuth angle in the horizontal plane (Left ear of subject 003 taken from the CIPIC database): "Polar" representation of HRTF magnitude spectrum, where the radius corresponds to the frequency axis and the angle to the azimuth angle.

The HRTF database of the CIPIC is considered [4]. Under the assumption that increasing angular difference between HRTF direction leads to increasing HRTF dissimilarity[3], the behavior of each criterion is checked for various pairs of HRTF corresponding to low and high angular mismatch.

2.2.1. Methodology

The goal is to collect a set of HRTF pairs with controlled dissimilarity, in order to assess the distance criteria. The HRTF database of the CIPIC [4], which provides a huge amount of HRTF data with 45 individuals (including one dummy head) and 1250 directions measured in the 3D sphere for each subject, will be used. Fig. 2 depicts the variation of the magnitude spectrum of HRTF in function of azimuth angle in the horizontal plane for one subject of the CIPIC database. Since the HRTFs illustrated in Fig. 2 are measured from the left ear, the HRTF magnitude is the highest on the left and decreases for right locations because of the acoustic diffraction induced by the head. Peaks and notches, which are mainly due to pinnae resonance, are also observed. Going from the left to the right, the magnitude spectrum varies quite continuously with the azimuth angle. Therefore it may be expected that when comparing two HRTFs located at azimuth angle θ_1 and θ_2 for instance in the front horizontal plane, their dissimilarity increases with their angular difference, which defines their *angle mismatch*:

$$d\theta = |\theta_2 - \theta_1|. \tag{8}$$

On the other hand, from a psychoacoustic point of view, we know that increasing angular mismatch leads to increasing error of localization. Thus it can be reasonably assumed that the HRTFs in the horizontal plane provides a wide range of HRTF dissimilarity. The HRTF dissimilarity relies both on signal processing (magnitude spectrum) and perception (localization error). However, before constituting the HRTF pair, it should be noticed that low dissimilarity may occur for strong angle mismatch because of the symmetry between front and rear HRTFs. For instance if the two HRTFs considered are taken from two locations which are symmetric with respect to the interaural axis, the HRTF are very similar despite a strong angular mismatch (cf. Fig. 2). Therefore it is preferred to consider separately the front and rear HRTFs in order to keep a confident link between the HRTF dissimilarity and the angle mismatch. On this condition, the HRTF dissimilarity varies in a monotone way with the angular mismatch.

In the CIPIC database, 25 directions are measured in the front horizontal plane, corresponding to azimuth angle varying from $-80°$ (on the left) to $80°$ (on the right), in the interaural polar coordinates. From these 25 HRTFs, 300 pairs are obtained with angular mismatch $d\theta$ varying from $5°$ to $160°$. Several pairs are associated to the same value of angular mismatch. In the same way, 300 pairs are also obtained from the rear horizontal plane. Thus a total of 1200 pairs, including 600 pairs both from the left and the right ear, is collected for each individual. The five dissimilarity criteria presented in Section 2 are then evaluated[4] for all the individuals available

[3] In this case, the HRTF dissimilarity is related to a localization error.

[4] When computing the criterion value, the HRTF considered as H_1 is always the HRTF with maximum energy, i.e. corresponding to the azimuth the more on the left for the left

in the CIPIC database [4]. The data obtained from all the individuals, the left/right and the front/rear sets are merged for each value of angle mismatch. The statistical analysis of the criteria values in function the angle mismatch is presented in the following Section.

2.2.2. Results

The behavior of the various criteria in function of the angular mismatch are displayed in Fig. 3. The results are plotted as blue boxes delimited by the lower (25%) and upper (75%) quartile. The median is depicted by a red line. In addition, green curves show the 5th (lower curve depicted by crosses) and 95th (upper curve depicted by circles) centiles, in order to illustrate the extent of the rest of the data. Since the range of values strongly varies from one criterion to another, the values displayed are all normalized by the maximum value of the 95th centile, in order to compare the criteria with the same scale. The values obtained before normalization for 5, 10 and 15° angle mismatch are given in Tab. 1.

A confident criterion should have monotone variation with increasing HRTF dissimilarity and low deviation for equivalent levels of dissimilarity, because it is intended to link criteria values with angular mismatches. From Section 2, it should be kept in mind that all the criteria are null or positive. The criterion value is null for perfect similarity (i.e. $H_1 = H_2$) and increases for increasing dissimilarity. In Fig. 3 the five criteria all exhibit almost linear increase with the angular mismatch. However, the values of the *Fahn* and *Avendano* criteria reach a ceiling for the highest angle mismatch (i.e. for mismatch greater than 100°). Except for *Durant* criterion, the deviation also increases with the angular mismatch. This phenomenon is particularly strong for the *MSE* and the *CB* criteria. Low deviation for small difference of azimuth is not surprising, since it can be observed from Fig. 2 that for small variation of azimuth, the HRTF variations are very close, whereas for greater variation of azimuth the HRTF variation are less consistent. The low deviation for small angular mismatch provides fine discrimination for low dissimilarity. From this point of view, the constant deviation of the *Durant* criterion is a drawback.

ear set and the more on the right for the right ear set. This choice has no influence for most of the criteria except for the *Fahn* and *Avendano* criteria, which include a normalization by the energy of the HRTF H_1.

Criterion	5°	10°	15°
C_{MSE}	0.0319	0.0677	0.116
C_{CB}	3.8310^{-6}	8.10^{-6}	1.3510^{-6}
C_F	0.0321	0.0691	0.108
C_A	0.137	0.29	0.446
C_D	11.8	16.3	20.3

Table 1: Median value of the 5 criteria for 5, 10, 15° angle mismatch.

Criterion	5°	10°	15°
C_{MSE}	0.0624	0.139	0.241
C_{CB}	6.10^{-6}	8.8110^{-6}	1.30^{-6}
C_F	0.0376	0.0609	0.0827
C_A	0.158	0.246	0.323
C_D	12.7	16.0	17.1

Table 2: Interquartile range of the 5 criteria for 5, 10, 15° angle mismatch.

It is also worth examining the extent of criteria values (i.e. the range delimited by the lower and upper green curves in Fig. 3) in function of the angular mismatch. It is striking that the range of criteria values for a given angular mismatch is wider for the *MSE* and the *CB* criteria than for the other criteria. Particularly, for the *MSE* criterion, the 5th-centile curve keeps very close to zero whatever the angular mismatch is, which means that this criterion may give low value although the HRTF dissimilarity is quite strong, which is not confident. The same defect is observed for the *CB* criterion. On the contrary, the *Fahn* and *Avendano* criteria show narrow extent of values, which suggests that these criteria provide a fine discrimination of HRTF dissimilarity. From these results, these two criteria can be considered as the most suited as distance criteria for HRTF dissimilarity. The influence of magnitude smoothing [13] of HRTF spectrum has been also studied, but no difference with the previous results has been pointed out.

2.3. Criterion calibration

When using dissimilarity criteria, one difficulty is to link the criterion values with dissimilarity level in

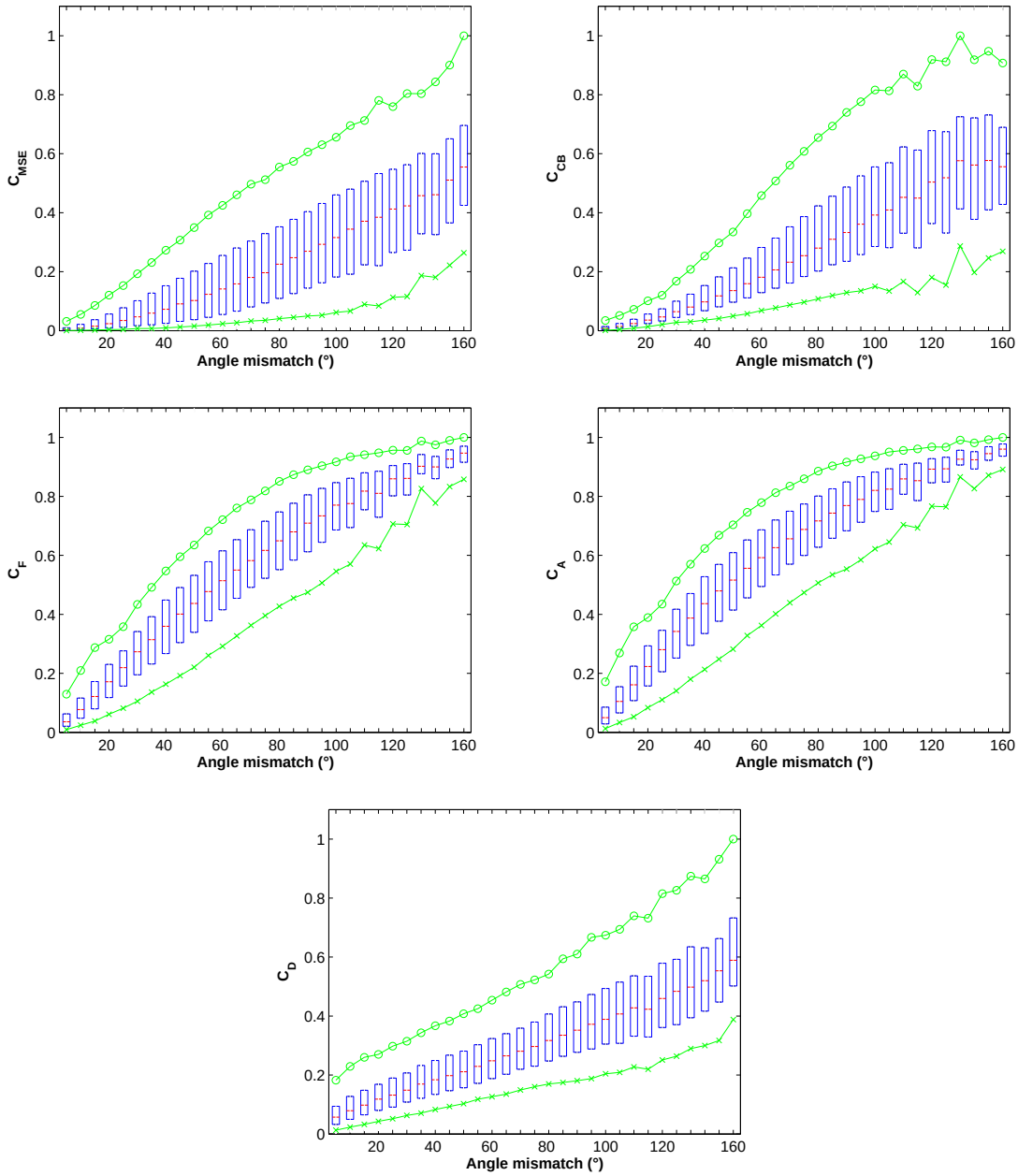

Fig. 3: A priori assessment of the dissimilarity criteria: criterion values in function of the angular mismatch. From left to right and top to bottom: *MSE* criterion, *CB* criterion, *Fahn* criterion, *Avendano* criterion, and *Durant* criterion. The blue boxes describe the lower and upper quartile. The median is depicted by a red line. The green curves show the 5th (lower curve depicted by crosses) and 95th (upper curve depicted by circles) centiles.

terms of the compared data. Considering two different pairs of HRTFs, if we suppose for instance that the *MSE* criterion gives a value of 0.03 for one pair and a value of 0.12 for the other, it is not obvious to know if these values denote low or high dissimilarity. From the previous analysis (cf. Section 2.2.2), we have now some knowledge about the physical and psychoacoutic meaning of the dissimilarity criteria. First, Tab. 1 shows that a *MSE* criteria value of 0.03 corresponds to an angular mismatch of 5°. The dissimilarity can be interpreted in two ways, by considering: either the difference between the HRTF magnitude spectrum or the localization mismatch between the two HRTFs. In Fig. 2, it can be observed that an angular mismatch of 5° leads to small variation of magnitude spectrum. In terms of auditory perception, a localization mismatch of 5° is very close to the lowest Minimum Audible Angle (MAA) [14] and so can be considered as hardly noticeable[5]. As a result, an angular mismatch of 5° is a low level of dissimilarity, whereas a mismatch greater than 10° corresponds to a noticeable level of dissimilarity, which allows us to calibrate each criterion. Tab. 1 gives the calibration values for the 5 criteria. Moreover, it is also interesting to know for a given step of decrease or increase of a criterion value whether this step is significant or not. The curves plotted in Fig. 3 can be used to interpret a given increase or decrease in terms of angular mismatch in order to assess its significance.

3. A POSTERIORI COMPARISON OF DISTANCE CRITERIA VIA HRTF CLUSTERING

After the previous *a priori* study, the present section will present an *a posteriori* study, where the five (dis)similarity criteria described in Section 2 will be assessed for clustering purpose.

Among clustering methods [20] the Self-Organizing Map (SOM) [15] is an excellent tool for data survey because it has prominent visualization properties. It creates a set of prototype vectors representing the data set and carries out a topology preserving projection of the prototypes from the N-dimensional input space onto a low-dimensional grid (two dimensions in the present paper). This ordered grid can be used as a convenient visualization surface for showing different features of the data[6] [16]. The SOM method is used in the following sections to compare the five criteria, by judging their ability to produce an homogeneous clustering and low quantification errors.

3.1. Methodology - Organization of the experiment

The HRTF data used for the clustering are first presented. Then it is described how the criteria are included in the training of a SOM. Thirdly, the three axis of the experiment are explained.

3.1.1. The data

Clustering of one or two HRTFs

In the CIPIC database (see Section 2.2.1), each individual is represented by his(her) HRTFs for various azimuths and elevations (θ,ϕ) described in the interaural polar coordinates. A total number of 1250 directions is available for each individual (see figure 4).

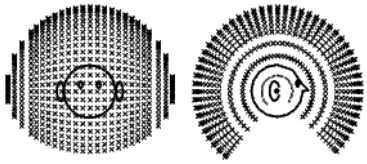

Fig. 4: Graphical description of the 1250 directions (CIPIC database of HRTF).

For each position (θ,ϕ), the HRTF is therefore represented by a vector of 100 components, one component per frequency. In the following study, the input vectors considered for HRTF clustering consists either of 100 components (if only one ear is considered) or of 200 components (if the ipsilateral and contralateral HRTF are considered) (see Fig. 5).

To cluster HRTFs one can considers input vectors which are represented with 100 components if only one ear is considered and represented as a vector with 200 components if the two ears are considered (see Fig. 5).

[5] However it should be noticed that the auditory perception of HRTF mismatch is not so simple and can not be considered only as a pure localization mismatch in a thorough analysis. Perception of spectrum difference should also be taken into account. The present paper provides only a first analysis.

[6] These visualization surfaces are not shown in this paper because the main interest here is to rank the criteria.

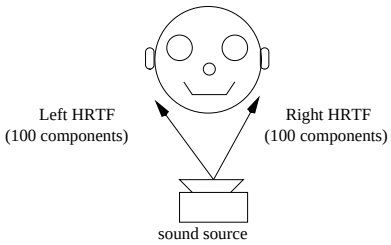

Fig. 5: Using one or two HRTFs.

HRTF preprocessing

The amplitude scale of the raw HRTFs is linear. It is transformed into a logarithmic scale closer to our auditory perception than linear scale (see [17] for instance).

In terms of the amplitude range, we consider that a lowest threshold of $-80dB$ (10^{-4} in the linear scale of amplitude) is sufficient from a psychoacoustic point of view. The input vectors (HRTF) are transformed as follows:

$$Hl_{(\lambda,\theta,\phi)}(i) = 20\log_{10}\left(max(H_{\lambda,\theta,\phi}(i), 10^{-4})\right) \quad (9)$$

where H denotes a HRTF in the linear scale, Hl a HRTF in the logarithmic scale and λ refers to the individual.

An example of Hls is given in Fig. 6 for the individual 003 of the database and for three positions: (θ=-80, ϕ=-45), (θ=0, ϕ=90) and (θ=80, ϕ =230). Even on a log-scale, the spectra exhibit strongly localized features (i.e. peaks and notches) which are critical for the sound localization. The accurate modeling of such features from a few measurements only is therefore a real challenge.

Statistical Learning Set

When clustering data, it is well known that it is necessary to split the data into several sets: a **training set** used to adjust the parameters of the model and a **test set** to estimate the generalization error of the modeling (in order to prevent from over-training).

In our case, the CIPIC database is composed of 45 individuals, each described by 1250 or 2500 HRTFs. The data have been split into two sets : 23 individuals for the training set and 22 individuals for the test set. Therefore the training set consists of 1250

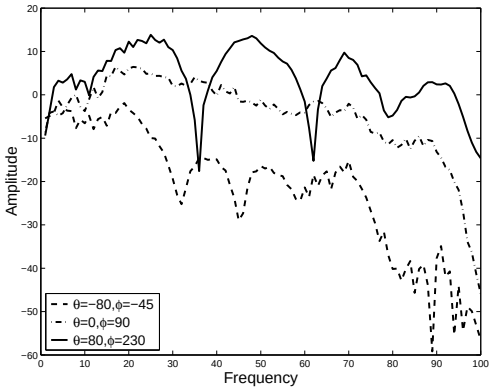

Fig. 6: HRTFs measured for three directions: (θ=-80, ϕ=-45), (θ=0, ϕ=90) and (θ=80, ϕ =230) - Individual 003 of the CIPIC database.

x 23 = 28750 vectors and the test set of 1250 x 22 = 27500 vectors.

3.1.2. Applying the five distance criteria into a SOM algorithm

The basic SOM algorithm

The basic SOM algorithm[7] is described below[8]. All the SOM in this article are square maps with hexagonal neighborhoods and are initialized with Principal Component Analysis (PCA). First the size of the Self-Organizing Map (SOM) [15] i.e k, the number of clusters and the topology of the SOM have to be fixed.

The basic SOM algorithm comprises five steps:
(A) choose the number, k, of clusters (H_1 prototypes);
(B) choose a topology of the map;
(C) initialization : choose random values for the k prototypes;
(D) for all iteration t

(D-1) random selection of an example H_2 from the training set,
(D-2) election of the nearest prototype (the "winner") using a distance criterion, for example if the mean squared error is used

$$\arg\min_k \|H_1^k - H_2\| \quad (10)$$

[7] All the experimentation on SOM have been done with the SOM Toolbox package for Matlab © [18]

[8] Here the algorithm is presented in a very simple way just to introduce the "winner" notion, for more details see [15]

$$\arg\min_k \left[\frac{1}{N} \sum_{i=1}^{N} [H_1^k(i) - H_2(i)]^2 \right] \qquad (11)$$

where k denote the kth prototype of the map and i refers to the frequency index.
(D-3) bring the winner near the example H_2 with a learning rate α;
(D-4) bring the neighbours of the winner, at this iteration t, near the example H_2
(D-5) go to step D-1 until convergence is not reached[9].

(E) end

Projecting the position information (this information is not used for the construction of the map) on the map allows to investigate the distinctive profiles of the clusters in terms of position and dispersion (see Fig. 7).

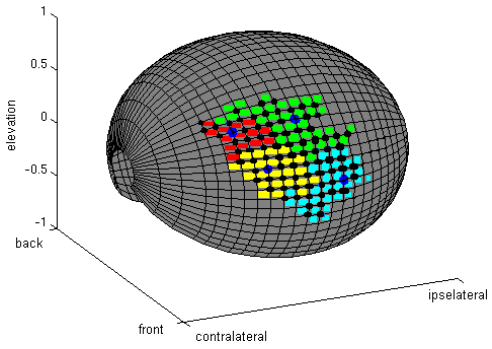

Fig. 7: Illustration of clusters of HRTFs.

At the end of the training the whole test set is presented to the map. For each example of this test set a winner is selected using the criterion used to build the map.

The modified SOM algorithm

The training procedure includes the notion of "winner" using a distance criteria. It is straightforward that the equation 11 can be changed by any of the criteria described in section 2 to elect the winner.

It should be remarked that there is only a scale difference between the *Fahn* and *Avendano* criteria (Section 2): the former is based on a linear scale, whereas the latter is logarithmic. In both cases, the election of the winner prototype gives the same result. The clustering results obtained with a SOM trained with either the *Fahn* or the *Avendano* criteria would be the same. Therefore only the *Avendano* results will be presented below.

[9]The convergence is obtained when winners do not move significantly.

3.1.3. Three axis of investigation

The study focuses on three issues : (1) the choice of the distance criterion, described in section 2; (2) the number of clusters (k); (3) the input data.

For (1) the four distance criteria have been described above.

For (2) : The number of clusters corresponds to the SOM size. For instance a SOM with a topology 4x4 contains 16 clusters. The final aim of this study is to find few representative HRTFs so only small value of k are considered. SOM larger than 8x8 ($k = 64$) have not been examined[10].

For (3) : These experiment includes two ways of considering the input data. Either the left and right HRTF are considered independently, i.e. the input vector is: H = H_{1L} or H_{1R} (vector of length N). Or the left and right HRTF are pulled together (see Fig. 5), in order to take advantage of shared information between the left and right HRTF about the overall diffraction by the listener's head. The input vector is then: H = [H_{1L} H_{1R}] (vector of length 2*N). It is examined whether it is useful to consider both the ipsilateral and contralateral HRTF for describing a direction and if taking into account this solution provides any advantage.

3.2. Clustering results

3.2.1. Introduction

Three errors are defined to compare the clustering performances of the distance criteria.

• The **global average quantification error** is defined as:

$$Eq = \sum_{t=1}^{T} \sum_{p=1}^{P} \sum_{i=1}^{N} |H_1^k(i) - H_2(i)|, \qquad (12)$$

[10]The following information may be useful for the readers who want to carry out the same experiments using Matlab Tool box [18]. The number of iterations for the rough tuning phase is 1500 for 2x2 and 4x4 SOMs, and 4000 iterations for 6x6 and 8x8 SOMs; the number of iterations for fine tuning phase is 500 for all SOM size. The time need to train all the SOM used in this article has been 30 days on a Pentium IV 3.8 GHz.

where H_2 denotes an input HRTF presented to the map, H_1 represents one of the k prototypes (the winner according to a given criterion), T is the number of individuals in the test set, P is the number of considered positions and i refers to the frequency index. This global error versus the SOM size and the criteria is shown on Fig. 8(a). This error is thus a very global error which merges all individuals of the test set (22), all positions (1250) and all frequencies (100).

• The **average quantification error per position** whatever individual λ is defined as:

$$Eq(\lambda, \theta, \phi) = \sum_{i=1}^{N} |H_1^k(i) - H_2(i)| \qquad (13)$$

The dispersion of this average quantification error per position is illustrated by Fig. 8(b) for each criterion, given a SOM size. This statistical analysis uses all the positions (1250) and all the individuals (22) of the test set. Therefore 27500 errors are aggregated inside each box plot.

• The **quantification error per frequency** is defined as:

$$Eq(i) = |H_1(i) - H_2(i)|, \qquad (14)$$

The dispersion of the quantification error per frequency is depicted for each criterion in Fig. 8(c), 8(d), 8(e) and 8(f). As previously the statistics include all the positions (1250) and all the individuals (22) of the test set, which leads to 27500 errors for each frequency.

These three errors are used below to compare the four criteria considering one or two HRTFs as input data.

3.2.2. Clustering the right ear HRTFs

In this first experiment, the input data consists only of the right ear HRTFs. Fig. 8(a) shows the influence of the number of clusters on the average error (Equ. 12) for each criterion. The common trend is the decrease of the average error when the size of the SOM increases. Of course the error will be null if the number of clusters is equal to the number of vectors constituting the training set[11]. It is intended to reach a good compromise between the number of clusters and the error. From Fig. 8(a) it can be seen that a SOM of size 6x6 (36 clusters) gives a reasonable error for each criterion. What's more a SOM size of 8x8 does not provide a great improvement. Based on this error, the ranking order of the four criteria is (beginning from the best): *MSE* (1), *Avendano* (1), *CB* (2) and *Durant* (3). The average quantification error obtained by the *MSE* and *Avendano* criteria for a SOM size of 6x6 is 3.8 dB. In terms of angular mismatch (cf. Section 2.2.2), this error value can be considered as equivalent to the level of dissimilarity observed in average between two HRTFs taken in the horizontal plane with an azimuth difference of 75°. This is a strong dissimilarity, but the level of data reduction is also considerable, since clustering by a 6x6 SOM means that 27500 vectors are described by only 36 representatives.

Now the SOM size is fixed to 6x6 (36 clusters):

- Fig. 8(b) describes the average error per position (Equ. 13) versus the criterion. The "best" criterion is the one which provides the smallest quantification error (i.e. the smallest median value) with low dispersion (i.e. narrow box plot). The same ranking order as in Fig. 8(a) is derived from Fig. 8(b), considering either the median value of the error or its dispersion. However it is still impossible to decide between the *MSE* and *Avendano* criteria.
- Fig. 8(c), 8(d), 8(e), 8(f) show the distribution of average errors (Equ. 14) in function of frequency for the four criteria. The criteria are judged according to the median value of the quantification error and the size of the box plot. The conclusions are the same as for the previous results (Fig. 8(a) and 8(b)): i.e. *MSE* (1), *Avendano* (1), *CB* (2) and *Durant* (3).

As a result, the *MSE* and *Avendano* criteria stands out as the best criteria from this experiment. They should be considered as equivalent without further information.

3.2.3. Clustering both the right and left ear HRTFs

The results obtained when using both the ipsilateral and contralateral HRTF (H_{1L} and H_{1R}) are pre-

[11] For instance the asymptotic result for the *MSE* criterion is close to 3 dB for 144 clusters [19].

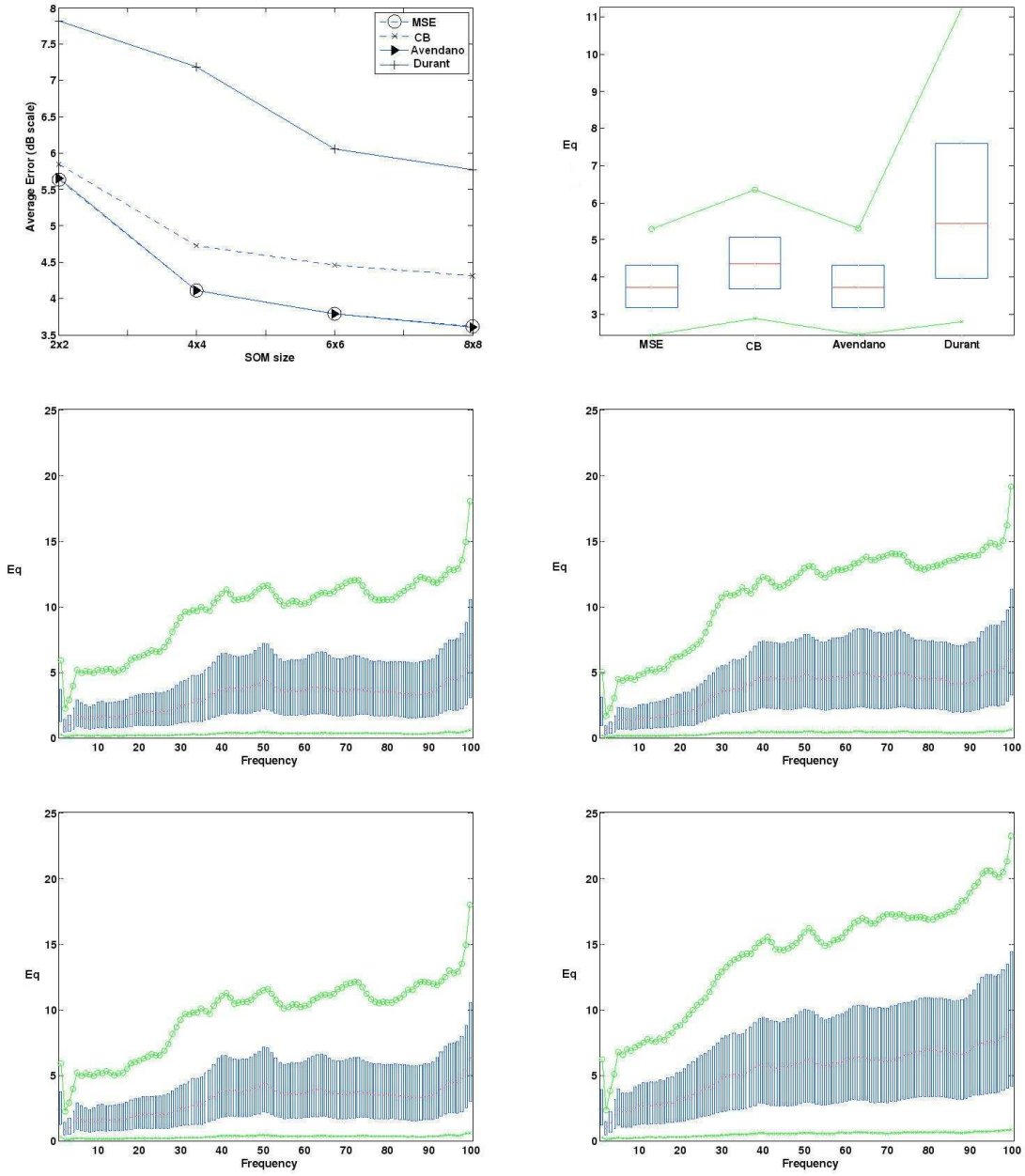

Fig. 8: From left to right and up to down : (a) Average quantification error (dB) (see Equ. 12) versus SOM size for each criterion (*MSE*, *CB*, *Avendano* and *Durant* criterion); (b) Average quantification error per position (see Equ. 13) for each criterion for a SOM which contains 36 clusters (6x6); (c)(d)(e)(f) Average quantification error per frequency (see Equ. 14) respectively for the *MSE*, *CB*, *Avendano* and *Durant* criterion, for a SOM which contains 36 clusters (6x6). (c)(d)(e)(f) : The blue boxes describe the lower and upper quartile. The median is depicted by a red line. The green curves show the 5th (lower curve depicted by crosses) and 95th (upper curve depicted by circles) centiles.

sented[12] in Fig. 9(a) to 9(f). The quantification error is slightly greater than when considering only the right ear HRTFs, but the difference is poorly significant[13]. A detailed analysis of the figures leads to the same ranking order of the distance criteria as previously.

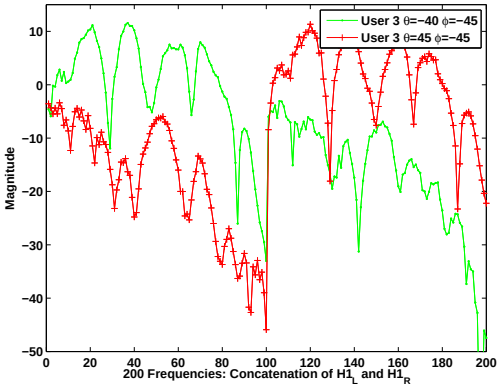

Fig. 10: Combining the left and right ear. Each plot depicts the concatenated HRTFs $H_1 = [H_{1L} H_{1R}]$ represented in log-scale (see equation 9).

At first sight it may be surprising that considering both the ipsilateral and contralateral HRTF for describing a direction does not provide any advantage. This result suggests that the information contained in $H = [H_{1L} H_{1R}]$ is not greater, in the sense of clustering, than the information provided only by H_{1L} or H_{1R}. More precisely the additional information conveyed by the HRTFs of the second ear is greater than the information of the first ear for certain positions (i.e. when the second ear is the ipsilateral one), but noisier for other positions (i.e. when the second ear is the contralateral one). This phenomenon is illustrated in Fig. 10.

In Fig. 10, the 100 first components on both curve represent the signal perceived by the left ear and the 100 following components are the signal perceived by the right ear. The right ear is illuminated by the sound source for location ($\theta = -40, \phi = -45$), but is shadowed for location ($\theta = 45, \phi = -45$). This is the opposite for the left ear. It is obvious that including the second ear HRTFs in the clustering algorithm adds information for the first location, whereas it adds only noise for the second location. Therefore considering all the database using two HRTFs for each position does not give any improvement.

[12] The 200 frequencies are used to elect the winner (see section 3.1.2) but only the 100th frequencies are used to compute the errors presented here since one wants to compare to the results presented in section 3.2.2

[13] Except for Durant which is realy improved using the two HRTFs on a position

4. CONCLUSION

HRTF (dis)similarity has been investigated through five distance criteria taken from literature. The criteria were assessed and compared in two ways: first by examining their behavior towards a sample of HRTFs with "controlled" dissimilarities, which are linked to various levels of localization mismatch, second by evaluating their performances for HRTF clustering. It is striking that the two studies point out the same criterion, namely the *Avendano* criterion. In addition, it has been shown how to link the value of a distance criterion to a physical scale of HRTF dissimilarity, in order to know whether a given value means either a low or a high dissimilarity, which is of prime interest when using distance criteria.

HRTF clustering has been used successfully for reducing the size of a HRTF database. Input data, which consists of 27500 HRTFs, can be described by only 36 representatives. The study first considered only one HRTF by direction. It was also examined whether it is useful to consider both the ipsilateral and contralateral HRTF for describing a direction, but the results suggest that this solution provides no advantage.

From the HRTF representatives it is expected to derive clever modeling of individualized HRTF, which will be the next step. Preliminary studies have given promising results [9]. Beyond data reduction, HRTF clustering is also a powerful tool for investigating the spatial and individual dependence of HRTF, which could be analyzed in the light of auditory perception.

5. ACKNOWLEDGMENT

The authors are very grateful to P. Guillon for many fruitful discussion during the writing of this paper.

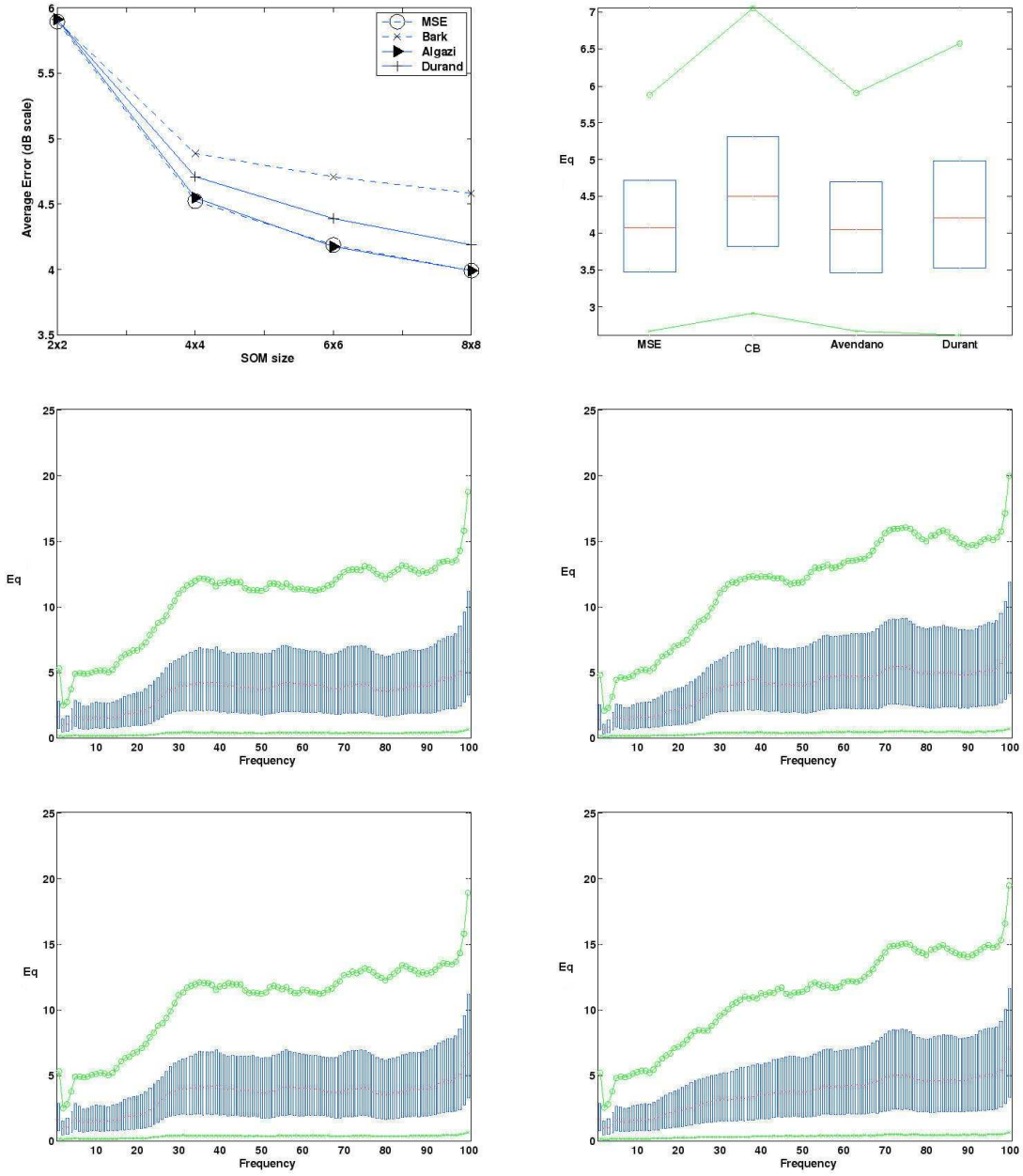

Fig. 9: From left to right and up to down: (a) Average quantification error (Equ. 12) versus the SOM size for each criterion (*MSE*, *CB*, *Avendano* and *Durant*); (b) Average quantification error per position (Equ. 13) for each criterion for a SOM which contains 36 clusters (6x6); (c)(d)(e)(f) Average quantification error per frequency (Equ. 14) respectively for the *MSE*, *CB*, *Avendano* and *Durant* criterion, for a SOM which contains 36 clusters (6x6). (c)(d)(e)(f) : The blue boxes describe the lower and upper quartile. The median is depicted by a red line. The green curves show the 5th (lower curve depicted by crosses) and 95th (upper curve depicted by circles) centiles.

6. REFERENCES

[1] F.L. Wightman and D.J. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis", JASA 85(2), pp. 858-867, 1989.

[2] A.W. Bronkhorst, "Localization of real and virtual sound sources", JASA 98(5), pp. 2542-2553.

[3] H. Moller, M.F. Sorensen, D. Hammershoi, C.B. Jensen, "Head related transfer functions of human subjects", JAES vol. 43, pp. 300-321, 1995 may.

[4] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, "The CIPIC HRTF Database", Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.

[5] Y. Kahana, "Numerical modelling of the head related transfer function", PhD thesis, University of Southampton, 2000.

[6] B.F.G. Katz, "Measurement and calculation of individual head related transfer functions using a boundary element model including the measurement and effect of skin and hair impedance", PhD thesis, Pennsylvania State University, 1998.

[7] S. Shimada, N. Hayashi, and S. Hayashi, "A clutering method for sound localization transfer functions", JAES vol. 42 n7/8, pp. 577-584, 1994.

[8] C.S. Fahn, Y.C. Lo, "On the Clustering of Head-Related Transfer Functions Used for 3-D Sound Localization", Journal of Information and Engineering, 19, pp. 141-157 (2003).

[9] V. Lemaire, F. Clrot, S. Busson, R. Nicol and V. Choqueuse, "Individualized HRTFs from few measurements: a statistical learning approach", International Joint Conference on Neural Networks IJCNN 2005.

[10] W.M. Hartmann, "Signals, Sound, and Sensation", Springer, 1998.

[11] C. Avendano, R.O. Duda, V.R. Algazi, "Modeling the contralateral HRTF", AES 16th International Conference on Spatial Sound Reproduction (1999).

[12] E.A. Durant, G.H. Wakefield, "Efficient model fitting using a Genetic algorithm: Pole-zero approximations of HRTFs", IEEE Transactions on speech and audio processing, Vol. 10 (1), (January 2002).

[13] J. O. Smith, "Techniques for Digital Filter Design and System Identification with Application to the Violin", PhD thesis, Elec. Engineering Dept., Stanford University (CCRMA), June 1983, CCRMA Technical Report STAN-M-14.

[14] J. Blauert, "Spatial Hearing, the Psychophysics of human sound localization", MIT Press, 1983.

[15] T. Kohonen, "Self-organizing maps", In Springer Series in Information Sciences, volume 30, Springer, Berlin, Heildelberg, 1995.

[16] "The many faces of Kohonen Map", V. Lemaire, Fabrice Clrot in "Classification and Clustering for Knowledge Discover", Series: Studies in Computational Intelligence, Vol. 4, Editeurs : Halgamuge, Saman K.; Wang, Lipo (Eds.) 2005, Approx. 300 p., Hardcover, ISBN: 3-540-26073-0.

[17] J.O Smith. "Techniques for digital filtering design and system identification with the violin". PhD thesis, CCRMA, Stanford, 1983.

[18] "SOM Toolbox for Matlab 5". TEchnical Report by Juha Vesanto and Johan Himberg and Esa Alhoniemi and Juha Parhankangas. Helsinki University of Technology, Neural Networks Research Centre. Report A57, April 2000,

[19] "Utilisation d'outils statistiques pour l'individualisation des HRTF" Report Research - France Telecom R&D - Vincent Choqueuse

[20] "Finding groups in data: An introduction to cluster analysis". Leonard Kaufman & Peter J. Rousseeuw. Wiley Series In Probability And Mathematical Statistics. John Wiley and Sons, Inc., 1989.