Perceptual assessment of binaural decoding of first-order ambisonics

Julian Palacino, Rozenn Nicol, Marc Emerit, Laetitia Gros

ACOUSTICS 2012

Perceptual assessment of binaural decoding of first-order ambisonics

J. Palacino, R. Nicol, M. Emerit and L. Gros

France Télécom - Orange Labs, FT/OLNC/RD/TECH/OPERA/TPS, 2 Av. Pierre Marzin, 22300 Lannion, France
julian.palacino@orange.com

The first-order Ambisonics microphone (e.g. Soundfield®) is a both compact and efficient set-up for spatial audio recording with the benefit of a full 3D spatialization. Another advantage is that the signals delivered by this microphone (i.e. B-Format) can be rendered over headphones by applying appropriate processing, while ensuring that the 3D spatial information is preserved. With the growing use of personal devices, it should be considered that most audio content is listened to over headphones. Thus first order Ambisonics recording provides an attractive solution to pick-up 3D audio content compatible with headphone reproduction. "Binaural decoding" refers to the processing to adapt B-Format for headphone rendering (i.e. "binaural format"). One solution is based on binaural synthesis of virtual loudspeakers. One promising way to improve the decoding is active processing which takes information from a pre-analysis of the sound scene, particularly in terms of spatial information. This paper will compare various binaural decoders. Starting from a listening test which assesses existing solutions and which shows that the perceived quality may strongly vary from one decoder to another, the processing is analyzed step by step. The performances are measured by a set of objective criteria derived from localization cues.

1 Introduction

3D audio recording provides immersive rendering of sound scene. Today Ambisonics (and its generalization Higher Order Ambisonics or HOA) proposes promising tools for spatial sound recording with the advantage of both full 3D spatialization and compact microphone set-up. On the contrary, 3D sound rendering remains a tricky issue, mainly in terms of equipment requirement. An attractive solution is therefore binaural processing of Ambisonics recordings, which means that the Ambisonics multichannel stream is adapted to headphone listening. In the following, this processing will be referred to as "binaural decoding" [1].

The objective of this paper is to assess the quality of binaural decoding. Various decoders are available today and the first step is a benchmark test to assess their performances, focusing on two strategies of decoding and, in addition, comparing them to other spatialization technologies. This test is presented in Section 2. Then the binaural decoding is analyzed step by step in Section 3, in order to identify where potential improvement may be found. Section 4 concludes the study by assessing the reconstruction of the signals delivered to the listener's ears for different options of processing.

2 Preliminary listening test

2.1 Objective

Our concern here is to provide spatial sound for headphone listening. Among the tools to record spatial sound, dummy-head is the most straightforward since binaural spatialization is precisely dedicated to headphone rendering. Stereophonic recording is another reference, as a conventional practice of sound engineer to record sound scene. Ambisonics proposes an attractive alternative. It should be highlighted that, in comparison to stereo, Ambisonics provides full 3D spatialization. However, it requires pos-processing, namely binaural decoding, to adapt the Ambisonics signals to headphone listening.

Thus, in a preliminary experiment, a listening test is performed in order to compare the perceived quality of various recordings of a spatial sound scene for the context of headphone listening. Three recording set-ups are considered: a dummy-head (KU100 Neumann acoustic head), an AB Stereo pair (i.e. a pair of 103 V 4003 DPA omni-directional microphones separated by 0.30 m) and a Soundfield® microphone. The test is based on excerpts taken from a live recording of the opera "Die Entführung aus dem Serail" of Mozart at the opera hall of the city of Rennes [2]. All the recording set-ups were placed above one seat and approximately at the potential location of the spectator's head, which allows the listener to be surrounded by the audience as he would be if he was really present in the hall. Two successive post-processing are applied to the output signals of the Soundfield® microphone: first conventional Soundfield® decoding to get the B-format signals [3], and second binaural decoding to adapt to headphone rendering. Two types of binaural decoders (which will be referred to as "SF dec1" and "SF dec2" in the following) are considered in our experiment, to contrast a "basic" decoder (SF dec1), i.e. mainly based only on the emulation of virtual loudspeakers, with an "active" decoder (SF dec2) in which the decoding is enhanced by sound scene analysis. On the contrary, the binaural recording obtained from the dummy-head is only equalized and the stereo signals are left untouched.

2.2 Experimental set-up

As a result, the listening test compares 4 types of sound spatialization, namely: dummy-head ("KU100"), stereo pair ("Stereo"), and 2 binaural decodings of the SoundField® signals ("SF dec1", "SF dec2"). The objective is to assess the overall quality (including both the audio and the spatial aspects) of the rendering of the sound scene over headphones. The experimental paradigm is based on a modified version of a MUSHRA test [4]. Since it is difficult to choose a priori one technology as a reference, no reference is proposed. Only one low anchor is added. This anchor includes both timbre and spatial degradations. Thus, for one trial, the subject is asked to judge a set of 5 pairs of signals ("KU100", "Stereo", "SF dec1", "SF dec2", "Anchor"). The assessment uses a multi criteria grid composed of 3 items: "quality", "space" and "timbre", following the methodology proposed in [5], except that a common anchor is used for each criterion. The overall test is based on ten audio excerpts covering various contents taken from the opera recording.

Twelve subjects (5 experts and 7 naive listeners) took part into the experiment. The overall test lasted around 2 hours and was divided into 2 parts separated by a break. The test interface was developed in Matlab. The experiment was carried out in an acoustically isolated room. The audio signals were presented to the subjects over HFI-580 Ultrasone closed headphones through a Terratec Phase 26 sound card configured for 48 kHz sampling rate and 24 bits resolution.

2.3 Results

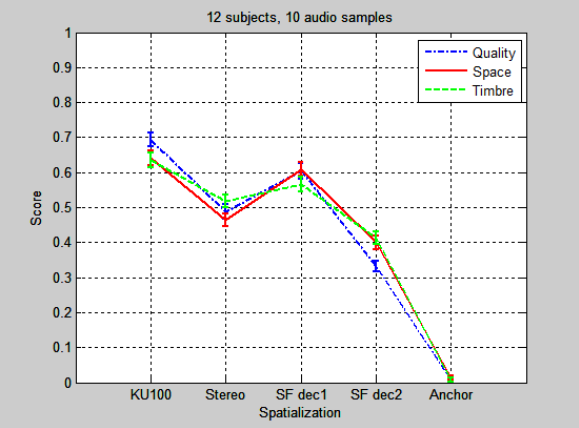

Figure 1: Mean score (and associated 95% confidence interval) in terms of quality, space and timbre.

For each sound spatialization, a total of 120 (10 audio excerpts x 12 subjects) scores were collected for the 3 criteria (quality, space, timbre). Figure 1 depicts the main score achieved by each technology. First, it should be noticed that the anchor of low quality was well identified. Second, in terms of global quality, the binaural recording is significantly preferred, followed by "SF dec 1" and the stereo pair. However, concerning space and timbre, the perception of the binaural recording and "SF dec 1" are very close. The stereo recording always stays in third position suggesting that this technology is not suited for headphone rendering. An ANOVA (ANalysis Of VAriance), considering 3 experimental factors (sound spatialization, audio excerpt, subject), confirms that the effect of sound spatialization is highly significant (p=0, F=233.89 for the "quality" criterion).

The same trend is observed for all subjects and all excerpts, except for the "applauses" excerpt. Indeed, for this latter, the binaural recording exhibits the worst score in terms of the "space" criterion. It may be due to the compression used to avoid the overload during the applauses. For all excerpts "SF dec 2" is judged significantly worse than "SF dec 1" in terms of "quality". It appears that binaural decoding of a first-order Ambisonic recording requires careful attention.

2.4 Conclusion

The results show that each system is clearly discriminated by the subjects and that the binaural rendering is significantly preferred. It is however striking that the Ambisonic recording is able to achieve a score close to the binaural sound, provided that a "proper" binaural decoding is applied. Indeed, it is also observed that the score of Ambisonics recording is highly dependant on the type of binaural decoding, which leads us to investigate the details of the processing in order to understand which element contributes to the perceived quality and where optimization can be expected.

3 Binaural decoding in question(s)

This part analyses step by step the overall processing from Ambisonics recording to headphone rendering.

3.1 HOA encoding

Ambisonics recording uses compact sensor arrays. The spatial encoding is based on the expansion of acoustical wave over spherical harmonics. Spherical harmonics Y_{mn}^{σ} define an Eigen base on the surface of a sphere of radius R defined by θ (azimuth) and ϕ $(elevation)$. Each element of this base is given by:

$$Y_{mn}^{\sigma} = \sqrt{(2m+1)\epsilon_n \frac{(m-n)!}{(m+n)!}} P_{mn}(sin\phi) \times \begin{cases} \cos n\theta \; si \; \sigma = 1 \\ \sin n\theta \;\; si \; \sigma = -1 \end{cases} \quad (1)$$

where

$$\begin{cases} m, n \in \mathcal{N}, n \leq m, \\ \sigma \in \{-1,1\}, \\ \epsilon_0 = 1, and \; \epsilon_n = 2 \; if \; n > 0 \end{cases} \quad (2)$$

m is the harmonic order and P_{mn} are the associated Legendre functions defined in $x \in [-1,1]$ by:

$$P_{mn}(x) = (1-x^2)^{\frac{n}{2}} \frac{d^n}{dx^n} P_m(x) \quad (3)$$

Under the assumption that sound sources are outside of the sphere of radius R, the expression of the acoustic wave inside is done by the following expansion:

$$p(kr,\theta,\phi) = \sum_{m=0}^{+\infty} i^m j_m(kr) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi) \quad (4)$$

where $j_m(kr)$ are spherical Bessel functions and B_{mn}^{σ} are obtained form the orthogonal projection of the acoustic pressure p to the corresponding spherical harmonic Y_{mn}^{σ}:

$$i^m j_m(kr) B_{mn}^{\sigma} = \langle p, Y_{mn}^{\sigma} \rangle \quad (5)$$

An approximation of pressure p can be done by truncating the expression (4) at the order $M \in \mathcal{N}$. This approximation gives K B_{mn}^{σ} coefficients defined by:

$$K = (M+1)^2 \quad (6)$$

The acoustic pressure field can be completely defined by the coefficients B_{mn}^{σ}. Those coefficients have been expressed here in the frequency domain. They can also be given in time domain using the inverse Fourier Transform.

$$b_{mn}^{\sigma}(t) = \int_{t=-\infty}^{\infty} B_{mn}^{\sigma}(\omega) e^{i\omega t} \, dt \quad (7)$$

In the case of a plane wave arriving from (θ_p, ϕ_p), which defines an angle γ with (θ, ϕ), the pressure is:

$$p(kr,\theta,\phi) = S \, e^{ikrcos\gamma} \quad (8)$$

its coefficients:

$$B_{mn}^{\sigma} = S(\omega) Y_{mn}^{\sigma}(\theta_p, \phi_p) \quad (9)$$

are the Ambisonics signals representing the whole acoustic field by the matrix relation:

$$\boldsymbol{b}_M = \boldsymbol{g}_M S(\omega) \quad (10)$$

where

$$\boldsymbol{b}_M = (B_{00}^1 \ldots B_{mn}^{\sigma} \ldots B_{MM-1}^{-1}) \quad (11)$$

and the gain array

$$\boldsymbol{g}_M = (Y_{00}^1 \ldots Y_{mn}^{\sigma} \ldots Y_{MM-1}^{-1}). \quad (12)$$

3.2 HOA decoding

The acoustic field can be reconstructed by a set L of loudspeakers located on the boundaries of a sphere. The coefficients B_{mn}^{σ} of the reconstructed soundfield are expressed as:

$$\boldsymbol{b} = \boldsymbol{C} \cdot \boldsymbol{s} \tag{13}$$

where

$$\boldsymbol{C} = \begin{bmatrix} Y_{00}^{1}(\theta_1,\phi_1) & \cdots & Y_{00}^{1}(\theta_l,\phi_l) & \cdots & Y_{00}^{1}(\theta_L,\phi_L) \\ & & \vdots & & \\ \vdots & & Y_{mn}^{\sigma}(\theta_l,\phi_l) & & \vdots \\ & & \vdots & & \\ Y_{M0}^{1}(\theta_1,\phi_1) & \cdots & Y_{M0}^{1}(\theta_l,\phi_l) & \cdots & Y_{M0}^{1}(\theta_L,\phi_L) \end{bmatrix} \tag{14}$$

and

$$\boldsymbol{s} = \begin{bmatrix} S_1 \\ \vdots \\ S_l \\ \vdots \\ S_L \end{bmatrix}, \boldsymbol{b} = \begin{bmatrix} B_{00}^{1} \\ \vdots \\ B_{mn}^{\sigma} \\ \vdots \\ B_{M0}^{1} \end{bmatrix} \tag{15}$$

S_l is the signal emitted by the loudspeaker l located at (θ_l, ϕ_l). The matrix $\boldsymbol{C}$ contains the spherical harmonics associated to the loudspeaker positions.

The objective is that the B_{mn}^{σ} coefficients of the reconstructed soundfield match those of the primary acoustic wave (Eq. (4)). An exact solution can be found when the number of loudspeakers L is higher than the order of truncation M. The loudspeaker signals are then derived from the B_{mn}^{σ} signals by applying a decoding matrix $\boldsymbol{D}$:

$$\boldsymbol{s} = \boldsymbol{D} \cdot \boldsymbol{b} \tag{16}$$

To find $\boldsymbol{D}$, which is in fact the inverse of the matrix $\boldsymbol{C}$, the Moore-Penrose pseudo-inverse can be used [6].

$$\boldsymbol{D} = \boldsymbol{C}^t \cdot (\boldsymbol{C} \cdot \boldsymbol{C}^t)^{-1} \tag{17}$$

If the loudspeaker array is uniformly distributed on the sphere, the relation (17) becomes [7]:

$$\boldsymbol{D} = \frac{1}{L}\boldsymbol{C}^t \tag{18}$$

because

$$\boldsymbol{C} \cdot \boldsymbol{C}^t = \frac{1}{L}\boldsymbol{I}_L \tag{19}$$

where $\boldsymbol{I}_L$ is the identity matrix of size $L \times L$.

3.3 Basic binaural decoding

Binaural synthesis uses a set of pair of binaural filters to create a virtual sound source for each position in space (r, θ, ϕ). Those filters are named Head Related Transfer Functions (HRTF) and can be obtained by measurement or modeling [8]. For Ambisonics decoding purposes, the reconstruction of acoustic field can be done by synthesizing virtual loudspeakers at the positions of available HRTFs. This method allows recreating an acoustic field at the entrance of the listener's ears.

The binaural signals $\boldsymbol{F_{bin}}$ of left and right channels are obtained as

$$\boldsymbol{F_{bin}} = \boldsymbol{H_{bin}} \cdot \boldsymbol{s} \tag{20}$$

$$\boldsymbol{F_{bin}}(\omega) = \begin{bmatrix} \boldsymbol{F_{bin,L}}(\omega) \\ \boldsymbol{F_{bin,R}}(\omega) \end{bmatrix} \tag{21}$$

$$\boldsymbol{H_{bin}}(\omega) = \begin{bmatrix} \boldsymbol{H_L}(\omega,\theta_1,\phi_1) & \cdots & \boldsymbol{H_L}(\omega,\theta_l,\phi_l) & \cdots & \boldsymbol{H_L}(\omega,\theta_L,\phi_L) \\ \boldsymbol{H_R}(\omega,\theta_1,\phi_1) & \cdots & \boldsymbol{H_R}(\omega,\theta_l,\phi_l) & \cdots & \boldsymbol{H_R}(\omega,\theta_L,\phi_L) \end{bmatrix} \tag{22}$$

where $\boldsymbol{H_{bin}}(\omega)$ is a matrix which defines the set of HRTFs measured for L directions. Substituting $\boldsymbol{s}$ for the loudspeaker signals in Eq.(20) leads to:

$$\boldsymbol{F_{bin}} = \boldsymbol{H_{bin}} \cdot \boldsymbol{D} \cdot \boldsymbol{b} \tag{23}$$

The binaural decoding matrix $\boldsymbol{E}$ is thus:

$$\boldsymbol{E} = \boldsymbol{H_{bin}} \cdot \boldsymbol{D} \tag{24}$$

which comes down to project the set of HRTFs on spherical harmonics [16]. Matrix $\boldsymbol{E}$ is $2 \times M$, which is a relatively small matrix for computation.

3.4 Pre-processing of HRTF

When implementing HRTF for binaural synthesis, some pre-processings are commonly used. It is intended to examine their potential impact on binaural decoding. First, modeling the HRTF by a minimum phase filter and a pure delay is considered. The delay is computed by the new method proposed by proposed by Nam [17] and validated by Nicol [18]. It consists in looking for the maximum of inter-correlation function between the HRIR and its minimum phase filter. Second, frequency smoothing is assessed. It is performed by critical band filter as described by Smith [19] and based on Hanning window.

In available databases (J.M. Pernaux, IRCAM[1], CIPIC[2], University of Maryland[3], Tohoku University[4] and Nagoya University[5]), HRTFs have been measured on the upper hemisphere of the sphere and in some cases also on part of the bottom hemisphere. The measurements are generally uniformly distributed over a single coordinate (azimuth or elevation) but not uniformly distributed over the whole sphere. However, for HOA decoding purposes, the projection of HRTF on spherical harmonics requires uniform sampling, in order to get an invertible decoding matrix in Eq.(19). Therefore HRTF interpolation is needed. The method based on Spherical Thin Plate Spline (STPS) derived from Wahba spherical spline [23] is chosen. Indeed Hartung et al [21] showed that this latter achieves the best performance.

4 Instrumental assessment of binaural decoding

Following the analysis of the processing, it is now intended to assess the performances of binaural decoding and to determine the effect of HRTF pre-processing over the resulting decoding. As a preliminary step, prior to a listening test, the assessment is based on a set of criteria, which are introduced in the next subsection.

4.1 Criteria

The assessment is focused on the rendering of spatial information. Therefore it is examined how the localization cues are reproduced in the signals delivered to the listener's ears. Sound localization uses mainly 2 kinds of cues: inter-aural cues (namely the Interaural Time Difference or ITD and the Inter-aural Level Difference or ILD, as described by the Lord Rayleigh's duplex theory [24]) and monaural cues [22]. ITD and ILD can be directly compared direction by direction. ILD is calculated using the method proposed by Larcher [16]:

$$ILD(\theta,\phi) = 10\log_{10}\frac{\int_{1.5kHz}^{10\ kHz}|H_L(\theta,\phi,f)|^2 df}{\int_{1.5kHz}^{10\ kHz}|H_R(\theta,\phi,f)|^2 df} \tag{25}$$

where H_L and H_R are the pressures at the left and right ear respectively. ITD is calculated using the method presented in subsection 3.4. Monaural cues rely essentially on spectral features. Therefore, the Inter-Subject Spectral Difference or

[1] http://recherche.ircam.fr/equipes/salles/listen/
[2] http://interface.cipic.ucdavis.edu/sound/hrtf.html
[3] http://www.isr.umd.edu/Labs/NSL/
[4] http://www.ais.riec.tohoku.ac.jp/lab/db-hrtf/
[5] http://www.sp.m.is.nagoya-u.ac.jp/HRTF/database.html

ISSD [25], which is derived from the variance of the difference between the original and reconstructed spectrum, is used to assess how spectral information (i.e. the frequency pattern involved in HRTFs) is correctly reproduced at the listener's ear:

$$\mathrm{ISSD}(\theta,\phi) = \left[\frac{1}{9\ kHz}\int_{4\ kHz}^{13\ kHz} 10\log_{10}\frac{\hat{H}(\theta,\phi,f)}{H(\theta,\phi,f)} - \Psi(\theta,\phi)df\right]^2 \quad (26)$$

$$\Psi(\theta,\phi) = \frac{1}{9\ kHz}\int_{4\ kHz}^{13\ kHz} 10\log_{10}\frac{\hat{H}(\theta,\phi,f)}{H(\theta,\phi,f)}\ df \quad (27)$$

For a set of HRTFs, a single value of ISSD can be calculated as the mean of ISSD of all considered directions. Middlebrooks points out the value of 6.18 dB as the optimum ISSD value [25].

4.2 Experimental protocol

For the evaluation of the different pre-processings, the private HRTF database *J.M. Pernaux* is used [20]. This database has a regular distribution on the upper part of sphere from an elevation of -56.25° and it contains 965 measured directions. The sampling frequency is 48 kHz and each HRTF is composed of 512 samples. All the assessment is performed over the subject labeled n° 1.

The various pre-processings are described in Table 1.

Table 1: List of applied pre-processings.

Description / Name	Minimum phase filter + ITD	Frequency smoothness	Interpolation
000			
100	X		
120	X	X	
130	X		X
123	X	X	X

Ambisonics encoding-decoding is applied over the entire set of HRTFs (965 original directions and 1026 interpolated directions). Spherical harmonic truncation used is 1, 4 and 30. The 1st and 4th orders corresponds to commercial Ambisonics microphones: 1st is Soundfield® [3] and 4th is Eigenmike®.The 30th order is calculated in order to get the best Ambisonics reconstruction as shown in Eq.(6) where optimum order M is chosen:

$$M = \sqrt{K} - 1 \quad (28)$$

K is the number of measured directions.

4.3 Experimental Results

Monaural cues (ISSD)

As shown in Figure 2, decreasing reconstruction order corrupts principally high-frequencies. Table 2 lets appear that Ambisonics reconstruction is quite precise at 30th order. In addition, processing "100" improves its quality in comparison to direct Ambisonics reconstruction for 30th and 4th order. But ISSD is degraded for smoothed HRTF and this occurs before Ambisonics processing. Nevertheless ISSD is lower for an Ambisonics reconstruction of any order of a smoothed HRTF. This property can be used to simplify HRTF spectrum for computational saving.

In the present case, interpolation over non-measured directions adds a negligible improvement of Ambisonics reconstruction. The interest of interpolation can be discarded taking into account that this pre-processing is computational expensive. However the current HRTF database is regularly distributed. Further test must be done over HRTFs sets that don't have this property.

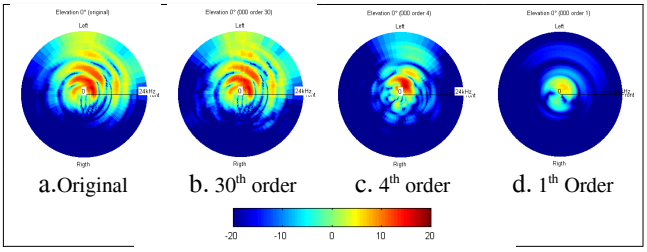

Figure 2: Horizontal cut at elevation 0° of magnitude spectrum of HRTF, Original (a.) and 1st ,4th, 30th order Ambisonics reconstruction without pre-processing

Table 2: ISSD values for different orders and pre-processing methods (in dB^2).

Pre-processing Name	Ambisonics order 1 ISSD (dB^2)	1 ISSD Degradation (dB^2)	4 ISSD (dB^2)	4 ISSD Degradation (dB^2)	30 ISSD (dB^2)	30 ISSD Degradation (dB^2)	Pre-processing only (Before Ambisonics decoding)
0	39.04	*39.04*	45.24	*45.24*	9.69	*9.69*	0
100	39.91	*39.38*	33.26	*32.73*	3.68	*3.15*	0.53
120	39.55	*30.82*	33.70	*24.97*	10.94	*2.21*	8.73
130	39.55	*39.02*	32.98	*32.45*	4.81	*4.28*	0.53
123	39.15	*30.42*	33.39	*24.66*	11.7	*2.97*	8.73

Binaural cues (ITD, ILD)

Table 3: ITD error mean and uncertainty values for 30th order encoding-decoding (in μs).

Name	000	100	120	130	123	mean
Mean ITD Error (μs)	18.8	14.9	14.6	15.0	14.8	15.6
ITD Error std. deviation (μs)	*2.5*	*3.6*	*3.9*	*3.7*	*3.9*	*3.5*

ITD varies commonly between 0 and 700 μs. As shown in Table 3, the ITD is well reconstructed at 30th order. Error is Gaussian distributed over a mean value varying around 15 μs with a standard deviation of 3 μs. For lower orders (1st and 4th) and for all the pre-processings, the ITD is completely lost and the resulting value oscillates around 0 s.

Like ISSD, ILD is really well reconstructed at 30th order. The achieved error is always less than 1 dB. At 4th order reconstruction of non pre-processed ILD values is Gaussian distributed over a mean value varying around 0 dB with a standard deviation of 3 dB and for some directions the maximum error reaches 8 dB. Anyway the spatial variation is coherent with natural ILD. All pre-processings increase the ILD error mainly for directions at the north hemisphere where ILD is over estimated. The mean of absolute value ITD error is then 7 dB and its standard deviation is 3 dB.

Generally 1st order reconstruction of non pre-processed HRTF gives good results in terms of ILD. Only some values at spots situated at (90°,-20°) and (270°,-20°) are over estimated. All pre-processings increase the number of maximum areas of ILD error. This happens because energy is focused principally over the Eigen vectors of 1st order spherical harmonics.

5 Conclusion

In the present paper, we studied the impact of HRTF database pre-processing for Ambisonics encoding-decoding purposes. Pre-processing considered are: modeling the HRTF by a minimum phase filter and a pure delay, frequency smoothing and HRTF interpolation over a regular distributed HRTF set over the whole sphere. HRTF reconstruction was assessed in terms of monaural cues using ISSD and in terms of interaural cues using ILD and ITD.

Modeling the HRTF by a minimum phase filter and a pure delay gives best results in terms of ISSD. Smoothness deteriorates ISSD before Ambisonics reconstruction but the reconstruction in any order of a smoothed HRTF is better than of non smoothed HRTF. Interpolation doesn't provide any improvement for the current HRTF database. Further studies must be done over less regular distributed HRTF sets.

ILD and ITD are well reconstructed over 30^{th} order and for all studied pre-processings. For lower orders, ILD is over estimated for some directions buts its general behavior remains. On the contrary, ITD is completely lost for 1^{st} and 4^{th} orders.

Future work will investigate the evolution of the different criteria as a function of Ambisonics order and the perceptual links by a listening test. Binaural active decoding is another issue to examine with the same criteria.

References

[1] S. MOREAU, « Étude et réalisation d'outils avancés d'encodage spatial pour la technique de spatialisation sonore Higher Order Ambisonics : microphone 3D et contrôle de distance », École doctorale de l'université du Maine, Le Mans, France, 2006.

[2] http://www.mozartsurecrans.com/ (consulted on September 2011)

[3] M.A. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound." *50th AES International conference*, 1975.

[4] ITU-R Recommendation BS.1534, "Method for the subjective assessment of intermediate quality level of coding systems," *International Telecommunications Union, Radio-communication Assembly*, 2003

[5] S. Le Bagousse et al., "Sound Quality Evaluation based on Attributes - Application to Binaural Contents ",131th AES International conference. New York, USA, 2011

[6] Golub et al, "Matrix computations". 3th edition. JHU Press, 1996.

[7] D. B. Ward et al. "Reproduction of a plane-wave sound field using an array of loudspeakers ", *Speech and Audio Processing, IEEE Transactions on*, vol. 9, n°. 6, p. 697–707, 2001.

[8] R. Nicol, "Binaural technology". *Audio Engineering Society*, 2010.

[9] Kulkarni, A., et al. "Sensitivity of human subjects to head-related transfer-function phase spectra". J. *Acoust. Soc. Am.* 105 (1999): 2821.

[10] Mehrgardt (S.) et al, "Transformation characteristics of the external human ear", *J. Acoust. Soc. Am.*, 61(6), 1977, p. 1567–1576.

[11] Glasberg (B. R.) et al, "Derivation of auditory filter shapes from notched-noise data", *Hearing Research*, 47, 1990, p. 103–138.

[12] Asano (F.) et al, "Role of spectral cues in median plane localization", *J. Acoust. Soc. Am.*, 88(1), 1990, p. 159–168.

[13] Kulkarni (A.) et al, "Variability in the characterization of the headphone transfer-function", *J. Acoust. Soc. Am.*, 107(2), 2000, p. 1071–1074.

[14] Guillon, Pierre. « Individualisation des indices spectraux pour la synthèse binaurale ». *Ph.D, Université du Maine*, 2009.

[15] P. Minnaar, et al, "The interaural time difference in binaural synthesis", presented at the *AES 108th convention*, Paris, 2000.

[16] V. Larcher, « Techniques de spatialisation des sons pour la réalité virtuelle », *PhD, Paris VI*, Paris, 2001.

[17] J. Nam, et al, "A method for estimating interaural time difference for binaural synthesis", in *125th Audio Engineering Society Convention*, San Francisco, 2008, vol. 21.

[18] R. Nicol, « Représentation et perception des espaces auditifs virtuels », *HDR, Université du Maine*, Le Mans, France, 2010.

[19] J. O. Smith, "Techniques for Digital Filter Design and System Identification with Application to the Violin", *PhD thesis, Elec. Engineering Department., Stanford University* (CCRMA), June 1983.

[20] J.-M. PERNAUX, « Spatialisation du son par les techniques binaurales : application aux services de télécommunications », PhD, I.N.P.G, Grenoble, 2003.

[21] K. Hartung et B. Jonas, "Comparison of different methods for the interpolation of head-related transfer functions", in *AES 16th*.

[22] J. Blauert, "Spatial hearing: the psychophysics of human sound localization", 2^{e} éd. Cambridge: MIT Press, 1983.

[23] G. Wahba, "Spline Interpolation and smoothing on the sphere", *SIAM J. Sci. Stat. Comput.*, vol. 2, mars 1981.

[24] L. Rayleigh, "On our perception of sound direction", *philosophical Magazine*, 13, 1907, p. 214-232

[25] J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency", *The Journal of the Acoustical Society of America*, vol. 106, p. 1480, 1999.