Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Objective evaluation of a three-dimensional sound field reproduction system

Nicolas Epain, Pierre Guillon, Alan Kan, Roman Kosobrodov, David Sun, Craig Jin and André van Schaik

Computing and Audio Research Laboratory, School of Electrical and Information Engineering,
The University of Sydney, Sydney, Australia

PACS: 43.38.Md,43.66.Pn

ABSTRACT

We present the results of an empirical evaluation of a three-dimensional sound field reproduction system consisting of 32 loudspeakers installed in a hemi-anechoic room at the University of Sydney. This loudspeaker arrangement allows up to third-order, two-dimensional, and fourth-order, three-dimensional Higher Order Ambisonic (HOA) reproduction of sound fields. The ability of this system to recreate a known sound field at the ears of a listener is evaluated using measurements with an acoustic manikin in the optimal listening position. In particular, we compare the Interaural Time Delay (ITD) and the Interaural Level Difference (ILD) generated by HOA for different sound source angles against reference values measured in an anechoic room. In addition, the influence of a listener's position on the quality of the reproduction is investigated based on measurements performed for different positions of the manikin around the "sweet spot".

INTRODUCTION

A good sound field reproduction system should accurately reproduce the spatial and timbral information of sound sources within a sound scene. For a human listener, the head-related impulse responses (HRIR) captures the direction-dependent acoustic transformation of a sound from its source position to the listener's ear. These HRIRs capture the interaural time difference (ITD), interaural level difference (ILD), and monaural spectral cues due to the ear, head and torso, which are known to be important cues for sound source localization (Blauert 1997). Hence, it is important for a reproduction system to be able to faithfully recreate these cues at the ears of the listener.

A three-dimensional sound field reproduction system has been built at the Computer and Audio Research Laboratory (CARLab), consisting of an array of 32 loudspeakers installed in a hemi-anechoic room at the University of Sydney. This system is intended for the reproduction of sound fields recorded by a spherical microphone array (Parthy et al. 2006), using the Higher Order Ambisonics (HOA) method. In the HOA framework, the sound field is described as a series of spherical harmonic functions up to a given order. This order has an effect on the size of the sweet spot and the frequency range at which sound source cues can be accurately reconstructed. The maximum order that our loudspeaker array can achieve is 4: according to the theory (Gumerov and Duraiswami 2005), this allows for a perfect reconstruction of the sound field in an area large enough to surround a listener up to approximately 2 kHz only. Also, the number and positioning of loudspeakers can affect the quality of the reproduced sound field (Bertet et al. 2009, Gerzon 1980).

In order to evaluate the performance of our three-dimensional sound field reproduction system, we have conducted an objective evaluation by measurements with an acoustic manikin. In particular, we assess the ability of the system to reproduce the ITD and ILD cues at the ears of the acoustic manikin, in the centre of the loudspeaker array - the “sweet spot” - and for locations up to 32 cm away from this point. The results of this evaluation are presented in this paper.

3D SOUND FIELD REPRODUCTION SYSTEM

HOA sound field synthesis

Our 3D sound field reconstruction system is based on Higher Order Ambisonics (HOA). In the HOA framework, the speakers are driven so that they reconstruct a sound field having a given order-L spherical harmonic expansion. Assuming the speakers act on the sound field as plane wave sources, the order-L spherical harmonic expansion of the sound field emitted by the loudspeakers is given by:

$$\mathbf{b}(f) = \mathbf{Y}_{\mathrm{spk}}\,\mathbf{g}(f)\,, \tag{1}$$

where $\mathbf{b}(f)$ denotes the vector of the order-L spherical harmonic expansion coefficients of the sound field at frequency f, $\mathbf{g}(f)$ denotes the vector of the speaker gains and $\mathbf{Y}_{\mathrm{spk}}$ denotes the matrix whose elements are the values of the spherical harmonic functions up to order L in the speaker directions, *e.g.*:

$$\mathbf{Y}_{\mathrm{spk}} = \begin{bmatrix} Y_0^0(\theta_1,\varphi_1) & Y_0^0(\theta_2,\varphi_2) & \dots & Y_0^0(\theta_{32},\varphi_{32}) \\ Y_1^{-1}(\theta_1,\varphi_1) & Y_1^{-1}(\theta_2,\varphi_2) & \dots & Y_1^{-1}(\theta_{32},\varphi_{32}) \\ \vdots & \vdots & \ddots & \vdots \\ Y_L^L(\theta_1,\varphi_1) & Y_L^L(\theta_2,\varphi_2) & \dots & Y_L^L(\theta_{32},\varphi_{32}) \end{bmatrix}, \tag{2}$$

where Y_l^m is the spherical harmonic function of order l and degree m and (θ_i,φ_i) denotes the angular direction of the ith speaker.

In order to play back a sound scene described by a given reference spherical harmonic expansion vector $\mathbf{b}_{\mathrm{REF}}(f)$, the speaker gains are calculated by multiplying $\mathbf{b}_{\mathrm{REF}}(f)$ with a decoding matrix $\mathbf{D}$, as follows:

$$\mathbf{g}(f) = \mathbf{D}\,\mathbf{b}_{\mathrm{REF}}(f)\,. \tag{3}$$

From Eq. 1, it is clear that a minimum sound field reconstruction error is obtained when D is calculated as the pseudo-inverse of $\mathbf{Y}_{\mathrm{spk}}$, *e.g.*:

$$\mathbf{D} = \mathrm{pinv}\left(\mathbf{Y}_{\mathrm{spk}}\right). \tag{4}$$

D is denoted as the *basic* decoding matrix.

Although the basic decoding matrix ensures a minimum sound field reconstruction error, the order-L spherical harmonic expansion of the sound field accurately describes the sound field only in the area defined by (Gumerov and Duraiswami 2005):

$$r \leq \frac{2L+1}{ek}, \tag{5}$$

where r denotes the distance to the origin, k is the wavenumber and e is the base of the natural logarithm, also known as Euler's number. Assuming an order-4 HOA sound field reconstruction is used, this means that above approximately 2 kHz, reconstructing the right spherical harmonic expansion does not imply that the pressure sound field is accurately reconstructed around a human head. In order to improve the perceived quality of the reproduced sound field, another decoding matrix is used at high frequencies. This second decoding matrix is denoted as the *maxrE* decoding matrix and is obtained by multiplying **D** with a weighting matrix, as follows:

$$\mathbf{D}_{\mathrm{maxrE}} = \mathbf{D}\,\mathbf{W}_{\mathrm{maxrE}}, \tag{6}$$

where $\mathbf{W}_{\mathrm{maxrE}}$ is a diagonal matrix whose non-zero elements depend on the spherical harmonic order only. In his PhD thesis, Daniel shows that this alternate decoding improves the perceived quality of the reconstructed sound field at high frequencies (Daniel 2000). In the case of a three-dimensional order-L decoding, the maxrE weights are given by:

$$w(l) = P_l(\gamma_L) \quad \text{for } \; l = 0, 1, ..., L, \tag{7}$$

where P_l denotes the Legendre function of degree l and γ_L is the largest root of P_{L+1}, *e.g.*:

$$\gamma_L = \max\left\{x \mid P_{L+1}(x) = 0\right\}. \tag{8}$$

Assuming the sound field to be reproduced consists of a single plane wave in direction (θ, φ), the corresponding time-domain spherical harmonic expansion signals are given by:

$$\mathbf{b}_{\mathrm{REF}}(t) = \mathbf{y}(\theta, \varphi)\; s(t), \tag{9}$$

where $s(t)$ denotes the time-domain source signal and $\mathbf{y}(\theta, \varphi)$ denotes the vector of the spherical harmonic functions in the source angular direction, *e.g*:

$$\mathbf{y}(\theta, \varphi) = \left[Y_0^0(\theta, \varphi) \;\; Y_1^{-1}(\theta, \varphi) \;\; \ldots \;\; Y_L^L(\theta, \varphi)\right]^T. \tag{10}$$

The corresponding speaker signals are then calculated by convolving the spherical harmonic expansion signals with a matrix of decoding filters:

$$\mathbf{g}(t) = \mathbf{D}(t) \circledast \mathbf{b}_{\mathrm{REF}}(t), \tag{11}$$

where $\mathbf{g}(t)$ denotes the vector of the time domain speaker signals and $\mathbf{D}(t)$ denotes the matrix of the Finite Impulse Response (FIR) decoding filters, whose frequency responses at low and high frequencies are the basic and maxrE decoding matrix, respectively. In the following, we are using a 4th-order three-dimensional HOA sound field reconstruction.

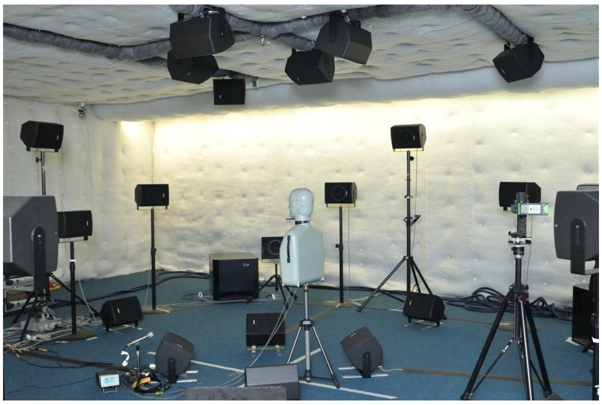

Figure 1: The sound field reproduction system.

Description of the loudspeaker array

The sound field reproduction system consists of a 32-channel array of loudspeakers installed in a hemi-anechoic room at the University of Sydney (Sun et al. 2009). Due to the physical dimensions of the room, the loudspeakers are arranged in a novel, five-layer configuration whereby 8 of the loudspeakers are equally spaced on the horizontal plane and the remaining 24 loudspeakers are placed in the directions of the vertices of a snub cube. The arrangement of loudspeakers in the hemi-anechoic room is shown on Fig. 1 and the loudspeaker spherical coordinates are shown in Tab. 1. Note that, due to the height of the room, the loudspeakers located on the ceiling and the floor are closer to the listener than the others.

Table 1: Spherical coordinates of the sound field reproduction system 32 loudspeakers.

♯	$\theta(^\circ)$	$\varphi(^\circ)$	ρ(m)	♯	$\theta(^\circ)$	$\varphi(^\circ)$	ρ(m)
1	16.5	0	2.7	17	0	14.5	2.7
2	61.5	0	2.7	18	90	14.5	2.7
3	106.5	0	2.7	19	180	14.5	2.7
4	151.5	0	2.7	20	-90	14.5	2.7
5	-163.5	0	2.7	21	33	-14.5	2.7
6	-118.5	0	2.7	22	123	-14.5	2.7
7	-73.5	0	2.7	23	-147	-14.5	2.7
8	-28.5	0	2.7	24	-57	-14.5	2.7
9	0	58	1.22	25	78	-27.5	2.34
10	90	58	1.22	26	168	-27.5	2.34
11	180	58	1.22	27	-102	-27.5	2.34
12	-90	58	1.22	28	-12	-27.5	2.34
13	45	27.5	2.34	29	33	-58	1.22
14	135	27.5	2.34	30	123	-58	1.22
15	-135	27.5	2.34	31	-147	-58	1.22
16	-45	27.5	2.34	32	-57	-58	1.22

This loudspeaker arrangement allows up to third-order, two-dimensional, and fourth-order, three-dimensional Higher Order Ambisonic (HOA) reproduction of sound fields. Audio signals for the loudspeakers are played from a computer equipped with an RME HDSP MADI sound card. The MADI output is converted to ADAT by an RME ADI 648 converter and then to analog signals by 3 Apogee DA-16 digital-to-analog converters. The analog signals are amplified by Lab Gruppen C Series amplifiers which are connected to the 32 Tannoy V6 loudspeakers.

Loudspeaker inverse filters

In order to compensate for individual differences in the loudspeaker transfer functions, including the different distances from the speakers to the centre of the array, inverse filters were calculated for each of the 32 loudspeakers in the array. The inverse filters were calculated from impulse response measurements made from each loudspeaker to a Brüel and Kjær Type 4165 calibration microphone placed in the centre of the loudspeaker array. The microphone was powered and amplified by a Brüel and Kjær Type 2610 measurement amplifier and its output digitized using an Apogee AD-16 analog-to-digital convertor and recorded by the computer providing audio signals to the loudspeakers. A 10-second long logarithmic sinusoidal sweep from 10 Hz to 23 kHz was used as a stimulus and appropriate processing was applied to recover the impulse response (Farina 2000). Since the room is not fully anechoic, the impulse responses were truncated so that only the direct sound component was used to calculate the inverse filters. The frequency response $E(k)$ of the inverse filters was calculated as:

$$E(k) = \min\left\{\frac{1}{|C(k)|}, \beta(k)\right\} e^{-i\angle C(k)}, \tag{12}$$

where k is the frequency index, C(k) is the N-point Fast Fourier Transform (FFT) of the measured loudspeaker impulse response, N is the number of coefficients of the inverse filter, and $\beta(k)$ is the maximum amplitude of the inverse filter at a particular frequency. For our inverse filters, β values were chosen for particular frequencies (shown in Table 2) and linearly interpolated to obtain the $\beta(k)$ values corresponding to every FFT bin.

Table 2: The β values at different frequencies.

Frequency (kHz)	0	0.05	0.1	0.2	0.5	20	24
β (dB)	0	0	6	10	20	20	0

The obtained inverse filters are applied to the speaker signals prior to the play back. Hence, the equalised speaker signals are given by:

$$\hat{\mathbf{g}}(t) = \mathbf{e}(t) \circledast \mathbf{g}(t), \tag{13}$$

where $\hat{\mathbf{g}}(t)$ denotes the vector of the equalised speaker signals and $\mathbf{e}(t)$ denotes the vector of the inverse filters. Replacing $\mathbf{g}(t)$ by the expression in Eq. 11, the equalised speaker signals can be expressed as a function of the spherical harmonic expansion of the sound field to be reconstructed:

$$\hat{\mathbf{g}}(t) = \mathbf{e}(t) \circledast \mathbf{D}(t) \circledast \mathbf{b}_{\mathrm{REF}}(t). \tag{14}$$

MEASUREMENTS

Loudspeaker array HRIR measurement

HRIRs for each of the loudspeaker positions were recorded using a Brüel and Kjær Head and Torso Simulator (HATS) manikin (Type 4128). Again, a 10-second long logarithmic sinusoidal sweep from 10 Hz to 23 kHz was used as a stimulus and the HRIR recovered from the recorded signal. HRIR recordings were made using HATS in the centre of the loudspeaker array (the "sweet spot"). Additional measurements were also made at displacements of ±1, ±2, ±4, ±8, ±16, and ±32 cm relative to the sweet spot in the front-back and left-right directions, and ±1, ±2, and ±4 cm in the up-down direction.

Effect of the inverse filtering

In order to assess the effect of the loudspeaker inverse filtering, as well as the accuracy of our measurements, we compare the equalised HRIRs with the data provided by Brüel and Kjær for the manikin. The equalised speaker HRIRs are obtained by convolving each of the raw HRIRs with the corresponding speaker inverse filter, *e.g.*:

$$\begin{bmatrix} \hat{h}_{L,i}(t) \\ \hat{h}_{R,i}(t) \end{bmatrix} = \begin{bmatrix} e_i(t) \circledast h_{L,i}(t) \\ e_i(t) \circledast h_{R,i}(t) \end{bmatrix}, \tag{15}$$

where $h_{L,i}(t)$ and $h_{R,i}(t)$ denote the left and right HRIRs corresponding to the ith speaker, respectively, $e_i(t)$ denotes the impulse response of the ith speaker inverse filter and $\hat{h}_{L,i}(t)$ and $\hat{h}_{R,i}(t)$ denote the corresponding equalized HRIRs.

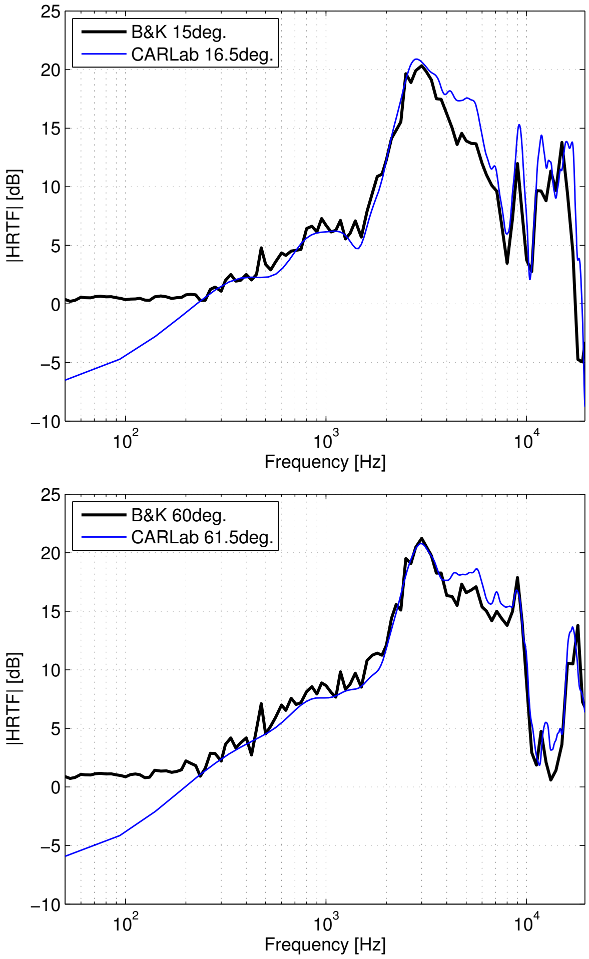

Figure 2: Comparison of the equalized loudspeaker HRTF magnitude with the data provided by Brüel and Kjær. On the top, the magnitude of the equalized left-ear HRTF obtained for the speaker located in direction (16,5 °,0 °) is compared with the frequency response magnitude provided by B&K for direction (15 °,0 °). On the bottom, the magnitude of the equalized left-ear HRTF obtained for the speaker located in direction (61,5 °,0 °) is compared with the frequency response magnitude provided by B&K for direction (60 °,0 °)

We then truncate the resulting impulse response to keep only the direct sound field part, and calculate the magnitude of the manikin Head Related Transfer Functions (HRTFs). Finally we compare the HRTF magnitudes with the frequency responses provided by Brüel and Kjær for directions close to some of our loudspeaker array directions. The result is shown on Fig. 2 for the left ear and two loudspeakers located in the horizontal plane. The equalized HRIRs match the reference very accurately for every frequency above 200 Hz. Below this frequency, however, the inverse filters do not sufficiently compensate the transfer functions of the loudspeakers.

Reference HRIR measurement

Reference HRIR measurements on HATS were also made in the anechoic chamber of the Auditory Neuroscience Laboratory at the University of Sydney. The anechoic chamber is equiped with a single loudspeaker (VIFA-D26TG-35) mounted on a robotic arm. The robotic arm can accurately position the loudspeaker to within a fraction of a degree, at any point on the surface of an imaginary sphere of 1 m radius. HATS was placed in the centre of the measurement sphere and aligned to the axes of the measurement system with the aid of a laser-alignment system. Golay codes were used as stimuli for the HRIR recording and the HRIRs recovered using the steps described by (Zhou et al. 1992). In order to improve the signal-to-noise ratio, 16 repetitions of the Golay codes, each 1024 samples in length, were recorded at each position. Tucker Davis Technology (TDT) system II hardware, interfaced with customized MATLAB software, was used to play and record the codes at an 80 kHz sampling rate. HRIRs were recorded for 393 sound source directions, upwards of 45° below the audio-visual horizon and equally distributed around HATS. The transfer function of the HRIR recording system to the centre of the measurement sphere was also recorded and an inverse filter calculated to deconvolve the recording system transfer function from the recorded HRIRs. Additionally, since the recorded HRIRs are at the limits of the noise floor for frequencies below 500 Hz, the magnitude response below 500 Hz of each HRIR were replaced with that calculated from a rigid sphere model (Duda and Martens 1998).

EVALUATION OF THE 3D SOUND FIELD REPRODUCTION SYSTEM

HOA-reconstructed HRIRs

In order to evaluate the accuracy of the sound field reconstruction at the ears of a listener, we compare the set of HRIRs measured in the anechoic room with HRIRs reconstructed using our sound field reconstruction system for the same source directions.

The sound field reconstructed at the ears of the manikin is given by:

$$\mathbf{p}(t) = \mathbf{H}_{\mathrm{spk}}(t) \circledast \hat{\mathbf{g}}(t)\,, \qquad (16)$$

where $\mathbf{x}(t)$ denotes the vector of the left and right ear pressure signals, and $\mathbf{H}_{\mathrm{spk}}(t)$ denotes the matrix of the recorded speaker HRIRs. Replacing $\hat{\mathbf{g}}(t)$ by the expression in Eq. 14, we can express the reconstructed ear pressure signals as a function of the spherical harmonic expansion of the sound field to be reconstructed:

$$\mathbf{p}(t) = \mathbf{H}_{\mathrm{spk}}(t) \circledast \mathbf{e}(t) \circledast \mathbf{D}(t) \circledast \mathbf{b}_{\mathrm{REF}}(t)\,. \qquad (17)$$

Using Eq. 9, we obtain the expression of the reconstructed ear pressure signals in the case of a single plane-wave sound field:

$$\mathbf{p}(t,\theta,\varphi) = \mathbf{H}_{\mathrm{spk}}(t) \circledast \mathbf{e}(t) \circledast [\mathbf{D}(t)\,\mathbf{y}(\theta,\varphi)] \circledast s(t)\,. \qquad (18)$$

Therefore, in the case of a single plane-wave sound field, the effect of the 3D sound field reconstruction system is equivalent to filtering the source signal with reconstructed HRIRs, *e.g.*:

$$\mathbf{p}(t,\theta,\varphi) = \mathbf{h}_{\mathrm{HOA}}(t,\theta,\varphi) \circledast s(t)\,, \qquad (19)$$

where the vector of the reconstructed HRIRs for direction (θ,φ), $\mathbf{h}_{\mathrm{HOA}}(t,\theta,\varphi)$, is given by:

$$\mathbf{h}_{\mathrm{HOA}}(t,\theta,\varphi) = \mathbf{H}_{\mathrm{spk}}(t) \circledast \mathbf{e}(t) \circledast [\mathbf{D}(t)\,\mathbf{y}(\theta,\varphi)]\,. \qquad (20)$$

Finally, we can evaluate the performance of the sound field reproduction system by comparing the set of HRIRs measured in the anechoic room with the HRIRs reconstructed for the anechoic measurement directions.

HOA-reconstructed binaural cues

In addition to the HRIRs and HRTFs, the interaural cues provide additional information regarding the sound field reconstruction quality. Reconstructing the interaural cues accurately is critical for the localisation of sound sources in azimuth. The first and most important interaural cue is the Interaural Time Delay (ITD). We calculate the ITD using the maximum interaural cross-correlation method (Kistler and Wightman 1992): for each source direction, the ITD is calculated as the time shift maximising the cross-correlation between the left and right-ear HRIR low-pass filtered at 2 kHz.

The second important cue for localising sources in azimuth is the Interaural Level Difference (ILD). The ILD is usually defined by the HRTF amplitude ratio for a particular frequency and source direction:

$$\mathrm{ILD}(f,\theta,\varphi) = 20\,\log_{10}\left|\frac{h_L(f,\theta,\varphi)}{h_R(f,\theta,\varphi)}\right|, \qquad (21)$$

where $h_L(f,\theta,\varphi)$ and $h_R(f,\theta,\varphi)$ denote the left and right ear HRTF at frequency f and for source direction (θ,φ), respectively. However, as suggested by Larchet, we calculate the ILD by averaging the HRTF energy over a given frequency range (Larchet 2001), *e.g.*:

$$\mathrm{ILD}(\theta,\varphi) = 10\,\log_{10}\left(\frac{\sum_{i=1}^{F}|h_L(f_i,\theta,\varphi)|^2}{\sum_{i=1}^{F}|h_R(f_i,\theta,\varphi)|^2}\right), \qquad (22)$$

where F is the number of frequency bins in the chosen frequency range. In our calculation of the ILD, we chose a frequency range of 1 to 3.5 kHz.

Finally, in order to evaluate the sound field reconstruction accuracy when the manikin is moved away from the center of the speaker array, we define two overall reconstruction errors for the interaural cues. First, we define the overall ILD reconstruction error as the absolute ILD error averaged over every source direction, *e.g.*:

$$\Delta_{\mathrm{ILD}} = \frac{1}{M}\sum_{i=1}^{M}|\mathrm{ILD}_{\mathrm{HOA}}(\theta_i,\varphi_i) - \mathrm{ILD}_{\mathrm{REF}}(\theta_i,\varphi_i)|\,, \qquad (23)$$

where $\mathrm{ILD}_{\mathrm{REF}}$ and $\mathrm{ILD}_{\mathrm{HOA}}$ denote the reference and HOA-reconstructed ILD values, respectively, and M is the total number of source directions. Similarly, we define the overall ITD reconstruction error as the absolute ITD error averaged over every source direction:

$$\Delta_{\mathrm{ITD}} = \frac{1}{M}\sum_{i=1}^{M}|\mathrm{ITD}_{\mathrm{HOA}}(\theta_i,\varphi_i) - \mathrm{ITD}_{\mathrm{REF}}(\theta_i,\varphi_i)|\,. \qquad (24)$$

The overall ITD and ILD errors provide two global estimations of the sound field reconstruction performance for every manikin position around the centre of the loudspeaker array.

RESULTS

Results obtained with the manikin in the centre of the loudspeaker array

We now present the results of the sound field reconstruction system evaluation. Fig. 3 shows the amplitude of the reference and HOA-reconstructed left-ear HRTFs for sources in the horizontal plane, with HATS located at the exact centre of the loudspeaker array. The general shape of the reconstructed HRTFs match the reference, the HRTF magnitude being clearly greater for sources located in the left hemisphere (positive azimuth values). However, the HRTF spectrum is accurately reconstructed only for frequencies below 3 kHz. This was expected as we are using an order-4 HOA system which, according to Eq. 5, is able to physically reconstruct the sound field around a listener up to 2 kHz only. On the other hand, note that the reconstructed HRTFs show a surprisingly good agreement with the reference ones between 12 and 16 kHz.

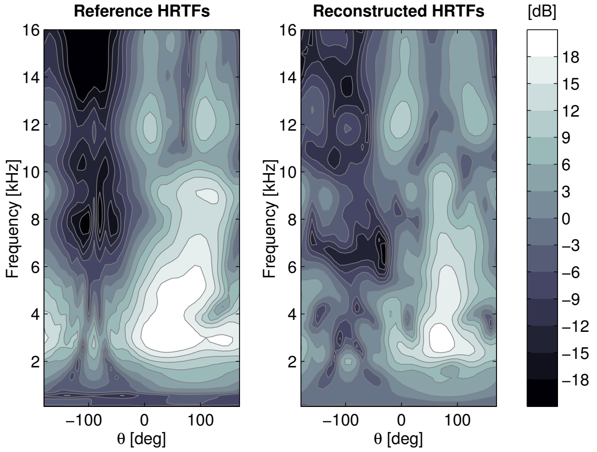

Figure 3: Comparison of the reference and reconstructed HRTF magnitudes for sources located in the horizontal plane, as a function of the frequency and the source azimuth. Left: magnitude of HATS left-ear HRTF measured in an anechoic room. Right: magnitude of HATS left-ear HRTF reconstructed by the HOA system.

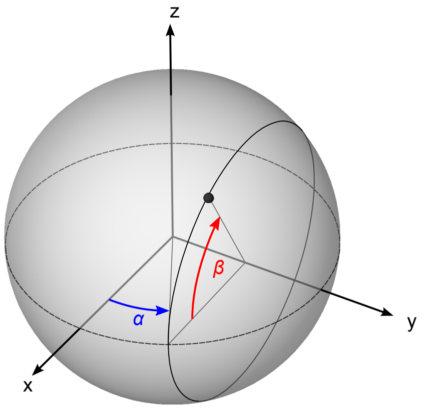

Figure 4: The sagittal coordinate system.

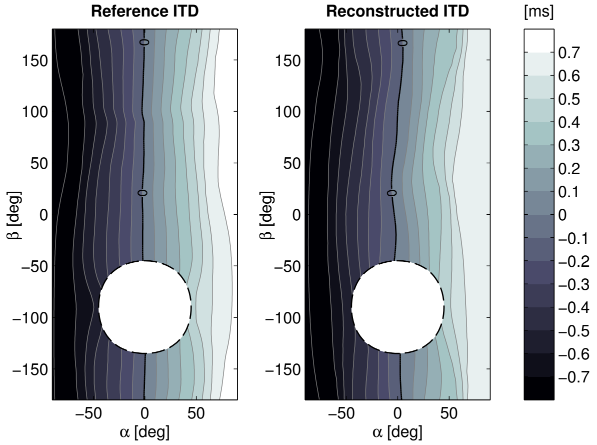

Figure 5: Comparison of the reference and reconstructed Interaural Time Delays (ITDs), as a function of the source direction in the sagittal coordinates α and β. The white ellipse represents the area where no anechoic measurements were available.

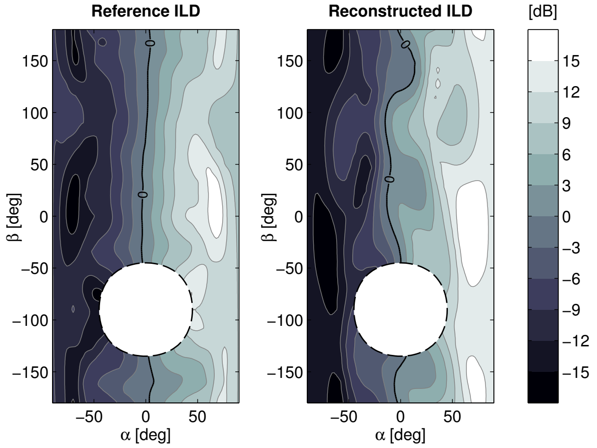

Figure 6: Comparison of the reference and reconstructed Interaural Level Differences (ILDs), as a function of the source direction in the sagittal coordinates α and β. The white ellipse represents the area where no anechoic measurements were available.

We now compare the values of the interaural cues calculated for the reference HRIRs and the HOA-reconstructed HRIRs obtained with the manikin located at the exact centre of the loudspeaker array. A particularly meaningful representation of the interaural cues is obtained when plotting the ITD and ILD in the sagittal coordinate system, as shown on Fig. 4. Each value of the angle α represents a different cone of confusion, along which the reference interaural cues are nearly constant.

Fig. 5 shows the reference and reconstructed ITDs as a function of the source direction, represented in the sagittal coordinate system. With the exception of the source directions with a β angle close to -90°, which correspond to sources located at low elevations, the reconstructed ITDs match the reference ones accurately. This result suggests that source localisation will be good in azimuth for broadband sounds.

Compared to the ITD reconstruction, the ILD reconstruction is not very accurate, as shown on Fig. 6. Although the reconstructed ILDs follow the general trend of the reference ILDs, they are generally shifted towards negative α angle values for β values comprised between -30 and 110 °, which correspond to sources located in the frontal hemisphere. Also, note that the reconstructed ILDs are exaggeratedly large (resp. small) for β angles greater than 50 ° (resp. less than -50 °). Nevertheless, these errors should not affect the localisation of sound sources as the ITD is accurately reconstructed and it is known to be a more dominant localisation cue for source azimuth. These ILD reconstruction errors are explained by the fact that the ILD is calculated using the HRTF magnitude values up to 3.5 kHz. This frequency is above the upper frequency limit for perfect sound field reconstruction using an order-4 HOA system, as given by Eq. 5.

The interaural cues can be used for source localisation in azimuth only. In order for the listener to localise sound sources in elevation, a good reconstruction of the monaural cues is required. These cues are provided by the magnitude of the HRTFs, which change according to the source direction and frequency. We already showed that the reconstruction of the HRTF magnitude was inaccurate above approximately 3 kHz for sources located in the horizontal plane. As our system is designed to reconstruct 3D sound fields, however, it is useful to look at how well the HRTF magnitudes are reconstructed for sources located outside the horizontal plane.

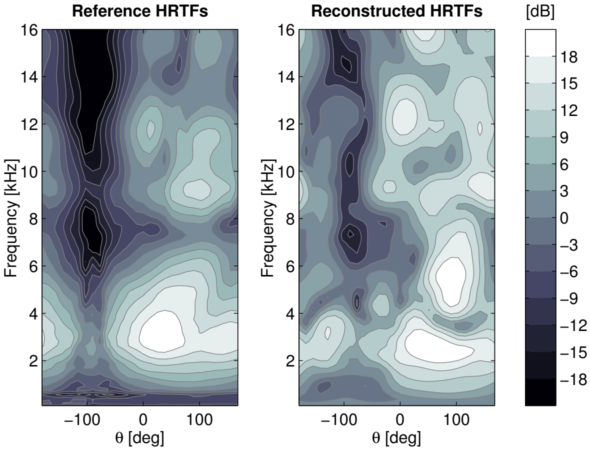

Figure 7: Comparison of the reference and reconstructed HRTF magnitudes for sources located in the plane of elevation -40 °, as a function of the frequency and the source azimuth. Left: magnitude of HATS left-ear HRTF measured in an anechoic room. Right: magnitude of HATS left-ear HRTF reconstructed by the HOA system.

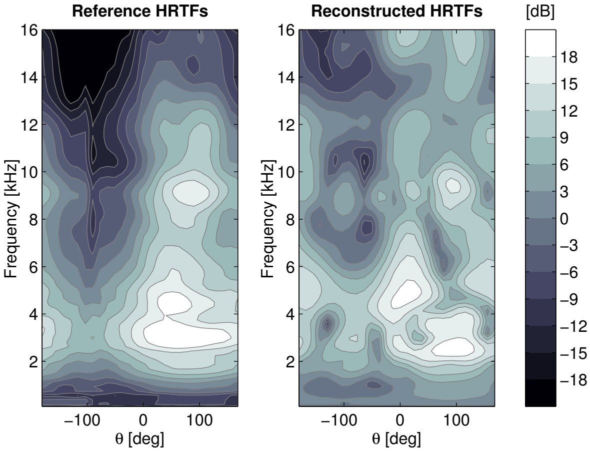

Figure 8: Comparison of the reference and reconstructed HRTF magnitudes for sources located in the plane of elevation +40 °, as a function of the frequency and the source azimuth. Left: magnitude of HATS left-ear HRTF measured in an anechoic room. Right: magnitude of HATS left-ear HRTF reconstructed by the HOA system.

Figs. 7 and 8 show the amplitude of the reconstructed HRTFs for sources located in the planes of elevation -40 ° and +40 °, respectively. As is the case for sources in the horizontal plane, the sound field reconstruction is reasonably accurate up to around 2 kHz. Above this frequency, however, the HRTFs are much less accurately reconstructed than in the case of sources in the horizontal plane. This result suggests that listeners will not be able to localise high and low-elevation sources well using our 3D sound field reproduction system.

Effect of moving the manikin away from the centre of the loudspeaker array

In addition to the centre of the loudspeaker array, HATS loudspeaker array HRIRs have been measured with the manikin being located at 30 other positions along the x, y and z axes. For each of these measurement positions, we then calculated the reconstructed HRIRs for every source position, as well as the corresponding ITDs and ILDs. Finally, we calculated the average ITD and ILD reconstruction error as given by Eqs. 23 and 24. These two global reconstruction errors provide a description of the effect of moving away from the sweet-spot.

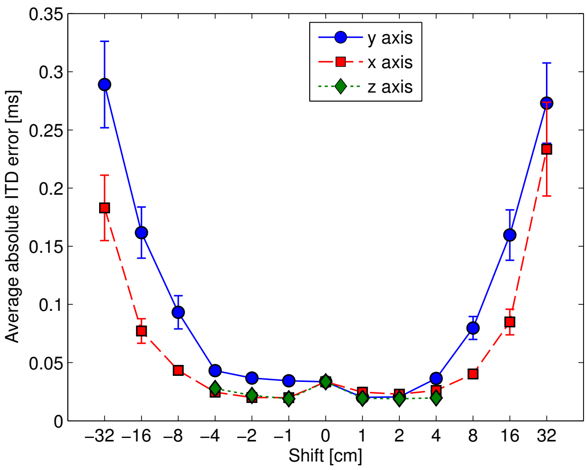

Figure 9: Average ITD reconstruction error as a function of the manikin displacement from the centre of the loudspeaker array along the x, y and z axes. The bars represent the 95 % confidence interval for each mean value.

Fig. 9 shows the value of the average ITD error as a function of the manikin shift along the x, y and z axes. Clearly, the ITD reconstruction error increases when moving away from the centre of the loudspeaker array. However, the error increases significantly only for shifts greater than 8 cm: up to 4 cm, the average error is less than 0.05 ms with a tight confidence interval which suggests excellent ITD reconstruction. From 8 cm onwards, the error increases dramatically to reach about 0.3 ms, which is very large considering maximum ITD values are around 0.8 ms. Also, the confidence intervals are then much wider, which suggests some extreme error values for particular source positions. Finally, note that the error increases faster when moving the manikin along the y-axis. This is not surprising since the y-axis is along the interaural axis.

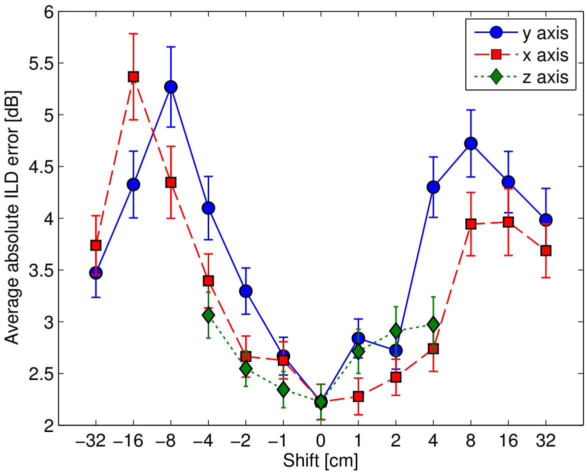

Figure 10: ILD reconstruction error as a function of the manikin displacement from the centre of the loudspeaker array along the x, y and z axes. The bars represent the 95 % confidence interval for each mean value.

Fig. 10 shows the value of the average ILD error as a function of the manikin shift along the x, y and z axes. As is the case with the ITD, the error clearly increases when moving away from the sweet-spot. However, this increase occurs much faster than in the case of the ITD: significantly larger error values are observed when moving the manikin 2 cm only to the left. The error also increase faster along the y-axis, which has already been observed for the ITD. Finally, note that the maximum error values are surprisingly not obtained for the largest shifts: the worst position in terms of ILD seems to be located 8 to 16 cm away from the centre.

These results show that the performance of our sound field reproduction system strongly depends on the position of the listener. The results suggest the existence of two spatial zones: (i) within 2 cm around the centre of the loudspeaker array, the ITD reconstruction is accurate and the ILD error varies moderately; (ii) from 8 cm onwards, on the other hand, the ITD error increases significantly, while the ILD reconstruction is clearly less accurate than in the exact centre of the loudspeaker array. This indicates that the listener can move his/her head slightly without substantial changes in the sound field reconstruction quality.

CONCLUSIONS

Three main conclusions can be derived from our results. First, our system seems to provide reasonably accurate sound localisation cues for source localisation in azimuth. In the case where the manikin is located at the centre of the loudspeaker array, the system achieves an almost perfect reconstruction of the ITD for every source direction. On the other hand, the ILD reconstruction is less accurate. However, the ITD cue is known to dominate localisation perception when the interaural cues contradict each other. Second, the localisation of sources in elevation will probably be imprecise using our system. To precisely localise sources in elevation requires that the monaural cues be reconstructed accurately, which our system does not achieve. Nevertheless, informal listening tests suggest that the system can recreate the impression of a source being at a low or high elevation. In other words, while the HOA loudspeaker panning is not accurate at the level of reproducing exact monaural spectral cues, it does provide sufficient acoustic cues regarding source location to give some impression of elevation. Third, the listener can move a few centimeters away from the centre of the loudspeaker array without any noticeable decrease in the sound field reconstruction quality. This is an important result as small head movements are known to improve the stability of the perceived sound scene image. In addition, the listener can then acquire dynamic cues which help in localising sources in elevation.

These results will be further investigated via a sound localisation test, which we intend to conduct in the near future. We also intend to use our loudspeaker array to compare HOA with the Vector Base Amplitude Panning (VBAP) method (Pulkki 1997).

REFERENCES

S. Bertet, J. Daniel, E. Parizet, and O. Warusfel. Influence of microphone and loudspeaker setup on perceived higher order ambisonics reproduced sound field. In *Proceedings of the Ambisonics Symposium 2009*, Graz, Austria, 25-27 June 2009.

J. Blauert. *Spatial Hearing*. MIT Press, Cambridge Mass., 1997.

J. Daniel. *Représentation de Champs Acoustiques, Application à la Transmission et à la Reproduction de Scènes Sonores Complexes dans un Contexte Multimédia*. PhD thesis, Université Paris 6, Paris, France, 2000.

R.O. Duda and W.L. Martens. Range dependence of the response of a spherical head model. *The Journal of the Acoustical Society of America*, 104(5):3048–3058, 1998.

A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Proceedings of the 108th AES Convention*, pages 350–5093, 2000.

M. Gerzon. Practical periphony: The reproduction of full-sphere sound. In *65th Convention of the Audio Engineering Society*, London, February 25-28 1980.

N.A. Gumerov and R. Duraiswami. *Fast multipole methods for the Helmholtz equation in three dimensions*. Elsevier, The Netherlands, 2005.

D.J. Kistler and F.L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *The Journal of the Acoustical Society of America*, 91(3):1637–1647, 1992.

V. Larchet. *Techniques de spatialisation du son pour la réalité virtuelle*. PhD thesis, Université Paris 6, Paris, France, 2001.

A. Parthy, C. Jin, and A. van Schaik. Optimisation of co-centred rigid and open spherical microphone arrays. In *Proceedings of the AES 120th Convention*, Paris, France, 20-23 May 2006.

V. Pulkki. Virtual sound source positioning using Vector Base Amplitude Panning. *Journal of the Audio Engineering Society*, 45(6):456–466, June 1997.

D. Sun, C. Jin, A. van Schaik, and D. Cabrera. The design and evaluation of an economically constructed anechoic chamber. *Architectural Science Review*, 52:312–319, 2009.

Bin Zhou, D.M. Green, and J.C. Middlebrooks. Characterization of external ear impulse responses using Golay codes. *The Journal of the Acoustical Society of America*, 92(2): 1169–1171, 1992.