



**HAL**  
open science

# Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources

Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, O. Warusfel

► **To cite this version:**

Stéphanie Bertet, Jérôme Daniel, Etienne Parizet, O. Warusfel. Investigation on Localisation Accuracy for First and Higher Order Ambisonics Reproduced Sound Sources. *Acta Acustica united with Acustica*, 2013, 99, pp.642-657. hal-00848764

**HAL Id: hal-00848764**

**<https://hal.science/hal-00848764v1>**

Submitted on 29 Jul 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources

Stéphanie Bertet<sup>1</sup>, Jérôme Daniel<sup>3</sup>, Etienne Parizet<sup>4</sup>, Olivier Warusfel<sup>2</sup>

<sup>1</sup> Sonic Arts Research Centre, Queen's University Belfast, UK (s.bertet@qub.ac.uk)

<sup>2</sup> UMR 9912 - Sciences et Techniques de la Musique et du Son - IRCAM-CNRS-UPMC, Paris, France

<sup>3</sup> Orange Labs TECH/OPERA, Lannion, France

<sup>4</sup> Laboratoire Vibrations Acoustiques, INSA Lyon, France

## Abstract

Ambisonics and higher order ambisonics (HOA) technologies aim at reproducing sound field either synthesised or previously recorded with dedicated microphones. Based on a spherical harmonic decomposition, the sound field is more precisely described when higher-order components are used. The presented study evaluated the sound field reproduced with four microphones of order one to four, perceptively and objectively in terms of localisation accuracy. A localisation test was performed on four microphones and a synthetic fourth order system. Results showed an improvement of the localisation with higher order ambisonic microphones. Reproduced localisation indices were calculated for the four microphones and the respective synthetic systems of order one to four. The perceptive and objective analysis revealed the same conclusions. The localisation accuracy depends on the ambisonic order as well as the source incidence. Furthermore, impairments linked to the microphones were highlighted.

## 1 Introduction

In every points of a source-free sphere centred on the origin of a chosen referential, the acoustic pressure can be expressed by a Fourier-Bessel decomposition. The theoretical decomposition induces an infinite number of harmonics for a exact sound field representation. In

practice, the series are truncated to a finite order  $M$ . The sound field description lies on a limited number of spherical Fourier  $B_{mn}^\sigma$  coefficients, inducing a partial sound field reproduction.

In the particular case of a plane wave reproduction of amplitude  $S$ , coming from a direction  $(\theta_S, \delta_S)$ , spherical Fourier coefficients are defined by the spherical

harmonics values for the direction of the plane wave weighted by the wave amplitude  $S$ :

$$B_{mn}^\sigma = S \cdot Y_{mn}^\sigma(\theta_S, \delta_S). \quad (1)$$

The ambisonic theory is based on this principle [41]. The encoding process describes as faithfully as possible the original sound field on a limited number of spherical Fourier coefficient named ambisonic components. It consists of decomposing the sound field on the spherical harmonics basis  $Y_{mn}^\sigma$ . The decoding process aims to reproduce the original encoded sound field in an area as large as possible to feed a regular reproduction system, composed of at least  $2M + 1$  loudspeakers for 2D or  $(M + 1)^2$  loudspeakers for a 3D rendering,  $M$  being the ambisonic order. Different decoders exist either to reproduce the original sound field, the *basic* decoder [12], to optimise the rendering for a large reproduction area - the *controlled-opposite* or *inphase* decoder [36], or to concentrate the energy of the loudspeaker configuration toward the direction of the source, called *maxr<sub>E</sub>*. A mixed decoding option that combines *basic* and *maxr<sub>E</sub>* was proposed to optimise, for a given order the resolution of the reproduced sound field [24, 14, 12]. Poletti [46] demonstrated mathematically that a suitable component weighting improved the decoding process. Furthermore, decoders for non-regular loudspeaker array were recently developed [54].

In 1977, the National Research Development Corporation in UK and Calrec Audio under the direction of Farrar [20, 21] published a patent that translates Gerzon's theoretical principles [11] to build the first order ambisonics SoundField microphone. Since then, higher order ambisonics (HOA) microphones have been investigated using spherical microphone array, either

open such as in [44], using a rigid sphere such as in [48, 43, 40, 18, 61, 2], or investigating hybrid arrays [1]. When evaluating the efficiency of ambisonic microphones, both the ambisonic technology and the microphone are sources of impairments.

As the ambisonics reproduction is a partial representation of the sound field, limits linked to the truncation of the finite order  $M$  of the Fourier-Bessel series appear. Sound field impairment has been defined in terms of quadratic error between the ideal and the reproduced waves [45, 13, 38, 52], expressing a reduction of the reproduction area proportional to the frequency. The reproduction error was linked by Daniel to a limit frequency for a reproduction area that is slightly bigger than a centred listener's head, for increasing truncation orders (table 1) [14]. One would need order 16 to reproduce the full audible bandwidth for the area of an average head. To express synthetically the reproduction of the sound source position, Gerzon defined the velocity vector, based on Makita's theory [34], which makes a connection between the apparent speed of the wave front reproduced in stereophony and the localisation phenomenon; and the energy vector, which is defined as the barycentre of the loudspeaker directions weighted by their energy contribution and points in the direction of the highest concentration [24].

Regarding a partial set of localisation cues at high frequency, the energy vector was interpreted by Daniel as providing a rough estimate of the blur width of the reproduced sound source. This blur width is objectively characterised by the angle  $\alpha_E = \arccos(r_E)$  [12], where  $r_E$  is the norm of energy vector. The use of HOA components  $M$  leads  $r_E$  close to 1. Therefore the blur width

of the sound image gets smaller with increasing order (table 1).

Order M	1	2	3	4
$f_{lim}$ (Hz)	700	1300	1900	2500
$\alpha_E$	45°	30°	22,5°	18°

Table 1: Limit frequency for an accurate reproduction of a plane wave in an area of the size of an average head, for an order M. Source width estimated from the energy vector with a  $\max_{\mathbb{R}}E$  decoding

Ambisonics microphones have been simulated and evaluated objectively in terms of component reproduction and physical sound field reproduction depending on sensor configuration, sampling scheme, sensor quality, sensor mismatch and placement [40, 48, 17, 60, 49]. Most perceptive studies are based on evaluating the reproduced sound field using a first order microphone [19, 28, 59, 29, 5]. Studies evaluating the perceived spatial resolution for HOA systems are mainly based on synthesised ambisonic sound field and show that higher order systems perform better than first order ones. Sontacchi et al. [53], Pulkki [47] and Bates et al. [4] evaluated HOA systems in terms of localisation performance. Results have shown an improvement of localisation with an increase of the ambisonic order (a third order approach is recommended by Sontacchi et al. [53]).

This study investigates the performance of four ambisonics microphones: the SoundField microphone and three HOA microphone prototypes developed in Orange Labs in the horizontal plane. First a localisation test performed with the four microphones and a synthetic

fourth order system is presented. Then an objective evaluation was carried out on the four microphones and four synthetic first to fourth order systems, using the same decoding option than the one used for the perceptive study. A distortion index was derived from theory to describe the blur width of a reproduced source as a function of its incidence angle  $\alpha_E$ . In order to compare the sound field the listener’s ears, interaural time differences were calculated for the eight systems. Finally, a localisation model estimating the direction of a sound source was used to compare the performance of the systems.

## 2 Evaluated microphones

### 2.1 SoundField microphone

The SoundField microphone reconstructs the directivity of ambisonics signals (W, X, Y and Z) combining signals from 4 hypercardioid sensors placed at the surface of a regular tetrahedron, thus allowing a quasi-coincident recording (figure 1). The error is negligible at low frequency. However, beyond a certain frequency the component directivities cannot be strictly reconstructed due to phase conflict. An equalisation to improve the ratio between the omni and bi-directional component was proposed by Gerzon and Craven [11] and implemented in the SoundField encoder.

### 2.2 The Higher Order Ambisonics prototypes

If the first order harmonics (omnidirectional and bidirectional components) can be directly created by com-

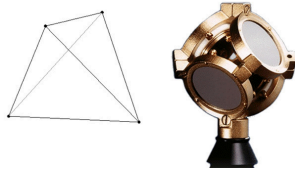


Figure 1: SoundField microphone. Left, sketch of a regular tetrahedron. The capsules are placed on each side. Right, SoundField microphone from Farrar’s work [20, 21].

binning elementary microphone capsules (cardioid or bi-directional) as if they were coincident, it is not the case for higher order harmonics. For higher order directivities that are not intrinsically contained in elementary combinations to appear, one has to involve more numerous and non-coincident capsules, exploiting both phase and amplitude differences between captured signals. Two HOA microphone prototypes of order 2 and 4 have been developed in Orange Labs. Moreau detailed the description of the fourth order prototype [39, 38, 40].

### 2.2.1 Second order microphone

The second order microphone (figure 2) is composed of 12 sensors positioned in a dodecahedron configuration on a semi rigid sphere (plastic ball) of 7 cm diameter. The sensors are evenly distributed on the sphere and spaced from 3.875 cm (arc length). The phase relationships between captured signals are unambiguous spatial information only for wavelengths larger than twice the inter-sensor spacing, e.g. for frequencies below the so-called spatial aliasing frequency. If a classic encoder is considered the aliasing frequency is 4400 Hz, above

which this limit spatial informations become inconsistent. This microphone is named 12-sensor in the following sections.



Figure 2: Second order ambisonics microphone. Left, sketch of 12 sensors placed in dodecahedron. Right, the prototype developed at Orange Labs.

### 2.2.2 Third order microphone

A third order microphone can be derived from the 8 sensors positioned on the horizontal plane of the 32 sensor microphone (figure 3). This encoder is aimed to reproduce a sound field in the horizontal plane only as it does not have elevated sensors. This microphone is named 8-sensor in the following sections. The eight sensors are not equally distributed on the horizontal ring (as the positions depend on the 32-sensor configuration). Therefore two theoretical aliasing frequencies are calculated : 6700 Hz and 7500 Hz.

### 2.2.3 Fourth order microphone

The fourth order microphone (figure 3) is composed of 32 sensors positioned in a pentaki dodecahedron configuration on a semi rigid sphere (plastic ball) of 7 cm

diameter (identical at the second order microphone's one). This microphone is named 32-sensor in the following sections. This configuration creates two different distances between adjacent sensors, 2.5 cm and 2.283 cm, which results in theoretical spatial aliasing frequencies that would be between 6700 Hz and 7500 Hz, with a classic encoder.



Figure 3: Fourth order ambisonics microphone. Left, sketch of 32 sensors placed in a pentaki dodecahedron configuration. Right, the prototype developed at Orange Labs.

### 2.3 Optimised encoding

The four microphones have been measured in the anechoic chamber of IRCAM. For each microphone, the measurements were sampled from  $-40^\circ$  to  $90^\circ$  in elevation and from  $0^\circ$  to  $360^\circ$  in azimuth with a stepsize of  $5^\circ$ . The procedure is described in details in [40].

In order to exploit the best potential of the four microphones, a specific encoding optimising the performance was used. The ambisonic components of the Sound-Field microphone were recorded directly, therefore taking into account the equalisation proposed by Gerzon

and Craven [11]. Furthermore for each HOA prototype, sensor responses were measured and combined to form an optimised encoding that minimises the influence of their non-flatness and their discrepancies [40]. By measuring the sensors responses for a representative set of sound source incidences in space, we obtain a transfer matrix that links the sensors responses to the sound source positions. In parallel, a target HOA encoding matrix can be obtained by computing the spherical harmonic coefficients for the same set of sound source directions. The matrix of the optimised encoding is then defined as the pseudo-inverse of the transfer matrix multiplied by the matrix of the HOA encoding coefficient. A regularisation factor is applied during the inversion of the transfer matrix in order to limit the low frequency amplification and control the signal-to-noise ratio. The optimised encoding aims at reproducing the sound field restricted to the horizontal plane. It pushes up the limit of the aliasing frequency and reduces the equalisation effort at low frequency. The reproduction of non-horizontal sound sources could induce spatial and spectral distortion. This choice has been made to compare the performances of all the microphones for an horizontal rendering. This encoding has been implemented for the three HOA microphones for an optimised rendering in the horizontal plane.

For 12-sensor and 32-sensor microphones (order 2 and 4 respectively) the response of the non-horizontal sensors was projected to the horizontal plane, which reduced the distance between sensors, therefore pushes the limit of the aliasing frequency. This operation did not occurred for the 8-sensor (third order) that used already the 8 sensors of the horizontal plane of the 32-

sensor, therefore its frequency limit did not improve.

Table 0 summarize

## 2.4 Encoding limits

From Shannon criteria, an ambiguity in the encoding process occurs from the aliasing frequency defined as:

$$f_{aliasing} = \frac{c}{2R\gamma} \quad (2)$$

where  $c$  is the speed of sound,  $R$  the radius of the sphere and  $\gamma$  the maximum angle between two sensors.

To characterise the encoding of the four microphones a spatial correlation and a level difference has been calculated between the directivity of the measured components and theoretical spherical harmonics. This allowed to synthetically estimate the degree of similarity of the directivity shape of the ambisonics components. The contribution of the encoded components could then be deduced from these indices. A frequency range for which components were well reproduced for a threshold of attenuation lower than  $-3$  dB is exposed in table 3.

The SoundField microphone shows a good reproduction of the ambisonics components on almost all the frequency range. Similarly to the SoundField microphone, the HOA microphones have the components 0 and 1 well synthesised up to a frequency much higher than the theoretical limit frequency  $f_{lim}$  thanks to the optimized encoding. The frequency limit of reconstruction of these components is lower for the 12- and 8-sensor microphones than for the 32-sensor microphone. The analysis shows a good reproduction of the higher order components associated to each HOA microphone but for a limited frequency range only. Although an

optimised encoding is used for each HOA microphone pushing the lower and upper limit frequencies of the reproduction of each component, the frequency range where the components are well reproduced can be considered as non-optimal if the theoretical limit frequencies  $f_{lim}$  (table 1) above which the area of reproduction is smaller than an average head, are considered. Increasing the gain of the high order components to compensate for these non-optimal limits would raise the noise floor of the recording at low frequency. Therefore the HOA components were kept as such.

## 2.5 Decoder

For the perceptive and objective evaluation, all the systems were decoded using a mix decoding option, optimising the resolution of the reproduced sound field. The *basic* decoder [12] reproduces a plane wave in a zone slightly bigger than an average head until a limit frequency  $f_{lim}$  depending on the order (cf. table 1). Beyond this frequency, the *maxr<sub>E</sub>* decoding option concentrates the energy contributions of the loudspeakers in the sound source direction [14, 12]. Shelf filters [23] placed before the decoding matrix were implemented. A biquadratic infinite impulse response (IIR) was created for each order with  $f_{lim}$  (cf. table 1) as a transition frequency. The system was composed of more loudspeakers than the minimum required number,  $2M + 1$  ( $M$  being the ambisonics order), to avoid the detent effect that pulls the sound toward the closest loudspeaker of the reproduced sound source direction [25, 23, 22]. All systems were decoded in order to feed a 12-loudspeaker reproduction system.

	SoundField microphone	12-sensor	8-sensor	32-sensor
Diameter	3.14 cm?	7 cm	7 cm	7 cm
Sensors spacing	?	3.875 cm	2.5 cm and 2.283 cm	2.5 cm and 2.283 cm
Aliasing frequency	7500 Hz ?	4400 Hz	6700 Hz - 7500 Hz	6700 Hz - 7500 Hz
Optimized frequency	7500 Hz ?	?	6700 Hz - 7500 Hz	?

Table 2: Specification of the four microphones

	order 1	order 2	order 3	order 4
SoundField microphone	>16500	-	-	-
12-sensor microphone	100 - 8300	900 - 4700	-	-
8-sensor microphone	300 - 7000	1400 - 5300	2800 - 6000	-
32-sensor microphone	100 - 18000	1000 - 17400	2000 - 18000	3400 - 17800
$f_{lim}$	700	1300	1900	2500

Table 3: Frequency range for which ambisonics components of the measured microphones are well reproduced for a threshold of attenuation lower than  $-3$  dB. The frequency  $f_{lim}$  represents the upper frequency limit of accurate reproduction over an area about the size of a head.



### 3 Subjective evaluation

To investigate perceptively the performance of the four microphones on reproducing accurately the direction of a sound source, a localisation test was performed<sup>1</sup>.

#### 3.1 Methods of report

In localisation tests, listeners are asked to report their judgment of a perceived sound source direction. Depending on the application, a great number of different reporting techniques have been used either in free field evaluation with physical sound sources or in sound reproduction (virtual sound source position) evaluation [37, 42, 57, 35, 26, 10, 31, 27, 33]. However, reporting methods may introduce bias or imprecision, depending on the equipment, the test (and training) duration, pointing precision, ability to report sound source position in all directions [51, 16].

The presented test aimed to measure the localisation accuracy of the listener for reproduced sources of different position on the horizontal plane, therefore independent of visual pointing. In previous studies using an acoustics pointer, the target was either a physical source or a synthesised static virtual sound source. The acoustic pointer was a moveable loudspeaker mounted on a rotating device aiming at matching the position of the static source [31, 47]. The proposed method reversed the respective role of the pointer and the target sound. The spatialisation method was used to build the virtual rotating pointer, whereas the reference sound was static and reproduced by a single loudspeaker. This inversion

---

<sup>1</sup>preliminary version of the results was presented at the Audio Eng. Soc. 30th Int. Conference, Saariselkä, Finland, March 2007.

was made possible because the ambisonics technique (as well as all spatialisation techniques) allows a sound source to be continuously reproduced on a circle using a limited number of loudspeakers.

The task of the subject consisted in matching a virtual sound source (created by one of the ambisonic encoding systems) with a real sound source (a loudspeaker). The virtual source was moved with a digital knob (without notch and stop) with one degree precision. It should be noted that the relation between the knob position and the pointer direction was not absolute. The only established information was the knob rotation direction. This prevented subjects from relying mainly on gesture, and thus put the emphasis on the auditory feedback. However, this constraint may have increased the difficulty of the task since the knob and pointer relation was not deterministic.

#### 3.2 Systems under test

The four microphones as well as a synthetic 4th order encoder were tested:

- SoundField microphone, first order,
- 12-sensor microphone, second order,
- 8-sensor microphone, third order,
- 32-sensor microphone, fourth order,
- a synthetic 4th order encoder.

The synthetic 4th order encoder was calculated from the equations of a theoretical encoding for a reproduction in the horizontal plane.

The sound source (the acoustic pointer) was projected on the spherical harmonics reproduced with the

measured microphones for every directions. The spherical harmonics were then decoded using a mix decoding option, as explained in section 2.5.

### 3.3 Reproduction system

Twelve loudspeakers evenly distributed among a ring of 48 loudspeakers composed the HOA reproduction system (displayed in figure 4).

The 48 loudspeakers were regularly spaced every  $7.5^\circ$  on a dodecagonal structure with a radius of 1.5 m. They were mid-range monitor prototypes developed by Studer (100 Hz - 20000 Hz). To compensate for the loudspeaker influence and the unperfect concentricity of the structure, the loudspeakers were measured by a microphone placed at the centre of the system (listener's position). Filters were calculated and applied for all loudspeakers' equalization. A regularisation factor was applied to limit the inversion effort of the filters. Consequently above 16 kHz, the sound was filtered by the loudspeaker responses.

The reproduction system was located in a listening room composed of absorbent wall panels and ceiling at Orange Labs. The room reverberation time was 0.3 s for frequencies below 500 Hz and 0.2 s above. An acoustically transparent curtain was hiding the system to the listener.

### 3.4 Stimuli

The target stimulus was played through a single loudspeaker while the pointer was an ambisonic reproduced sound source. The target and pointer were spectrally different to avoid tonal match.

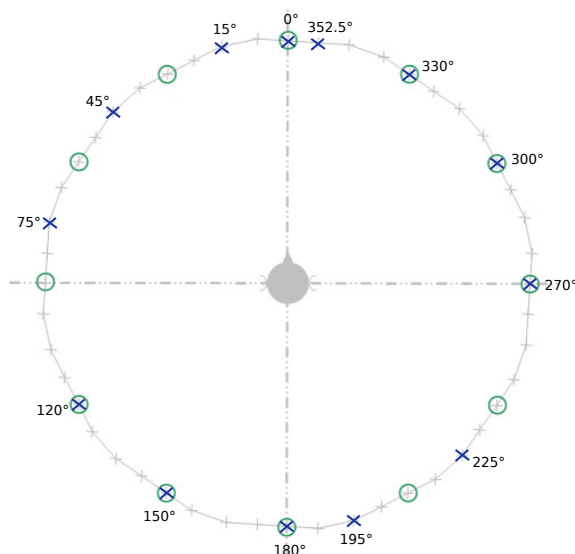


Figure 4: Listening system and target positions. The 48 loudspeakers of the ring are depicted by small crosses (+). The loudspeakers used for the various HOA sound pointer decoding are displayed with a superimposed circle (o). The loudspeakers used for the different target directions are depicted by superimposed diagonal crosses (x).

### 3.4.1 Target stimulus

The target stimulus was a 206 ms train of nine 22 ms white noise bursts modulated in amplitude. Two bursts were separated by 1 ms. The average sound level was 70 dB SPL at the ears of the listener.

### 3.4.2 Acoustic pointer stimulus

A broadband uniformly masking noise [62] (20 Hz to 20 kHz) was used to build the pointer stimulus. Its large frequency range ensured that all localisation cues were used. The noise had to be filtered to account for both the decoding process and compensation of the loudspeaker reproduction system. However, the pointer being dynamic (following the knob rotation) a real-time implementation would have resulted in possible delays due to the size of the filters. Therefore artefacts may have occurred during the listening phase. It was decided to pre-process the pointer stimuli. Therefore the noise was convolved with the encoded impulse responses of each system, corresponding to the 360 possible directions of the pointer. Adjusting the pointer simply consisted in playing and switching between pre-processed multi-channel sound files. Each noise burst duration was limited to 150 ms to avoid any sensation of static sound and large jump in between different sound directions.

## 3.5 Test procedure

The pointer and target sounds were alternately presented and separated by a 150 ms silence (as shown in figure 5). The listener had twenty-five target - pointer presentations to adjust the pointer to the target direction. The overall presentation duration (17.4 s) was a com-

promise between the answer's time and the repetition number that the listener needed to adjust the pointer. The initial direction of the pointer was randomly chosen between  $\pm 20^\circ$  and  $\pm 60^\circ$  around the target direction. The listener was not aware of this first direction.

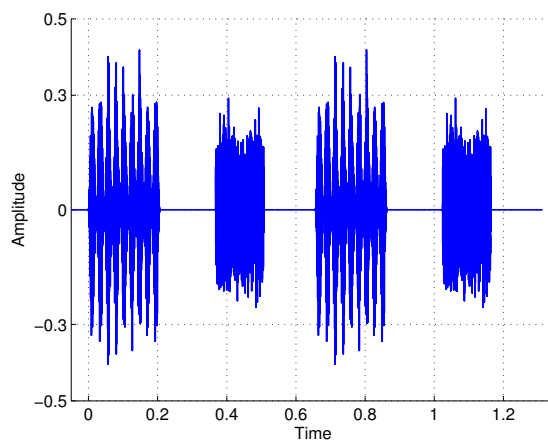


Figure 5: Sequence of 2 target - acoustic pointer stimuli, presentation in time domain.

As the pointer stimulus was pre-processed for each pointer direction it could not move simultaneously with the knob rotation. However, all knob movements were recorded and taken into account for pointer play back. Thirteen target directions were chosen in the horizontal plane, favouring the evaluation of the frontal listening area (figure 4). The spatial distribution of the sound source positions was not symmetrical (left/right) but were evenly distributed in order to span the horizontal plane. Seven target positions corresponded to one of the loudspeakers used for ambisonic reproduction. The remaining six were chosen among the other 36 loudspeakers. The listener was standing in the centre of the ring. Although the head was not fixed she/he was

instructed not to move it and to fix a visual reference mark that was placed in the  $0^\circ$  direction.

The listener was first familiarised with all the virtual ambisonic pointers (without naming them) and the relation between the rotation of the knob and the moving sound source. Then, a short training session (10 sequences) was proposed, where all systems were presented for two random target directions. Afterwards, the test was composed of 195 sequences (13 directions  $\times$  5 systems  $\times$  3 repetitions) randomly presented. Each target direction was started by pressing a trigger. This allowed the listener to anticipate the end of the sequence by switching to the next one or to take a break during the test whenever she/he wanted. The test lasted approximately one hour and breaks were advised after each 20 min.

### 3.6 Listeners

Fourteen listeners including four women and ten men took part in the experiment. Their age varied from 22 to 45. All listeners reported no hearing problem but their hearing threshold had not been measured before the experiment. All listeners had listening test experience in the past but only three of them had experience in spatial sound listening, and were therefore considered as experienced listeners.

### 3.7 Analysis

#### 3.7.1 Choice of the pointing value

The listener had a restricted time to adjust the pointer to the target direction. The number of 25 repetitions was determined by trials testing the time of adjusting

the pointer to the target with all systems (from first to fourth order). Nevertheless the pointer value to consider for analysis was investigated. In 76.4% of cases the listener anticipated the answer stopping the sequence before the end of the 25 repetitions, as shown in figure 7 a), b) and d). This means that 25 iterations were sufficient for listeners to adjust their answers in the majority of cases. The final pointer value was considered for the analysis.

For the other 23.6% it is interesting to check whether the listener was actually still manipulating the knob or if she/he had converged to a stable answer. Firstly, an analysis of variance on the number of repetitions showed that the evolution of the diverging answers did not decrease along the test ( $F(194) = 1$ ,  $p = 0.49$ ). Therefore the non convergence was not due to a training effect. Secondly, the variability of the pointer was analysed. The standard deviation of the pointer direction for the last five repetitions (i.e. the last 3 seconds of the sequence) was considered with a threshold set at  $10^\circ$ . If the standard deviation was lower than  $10^\circ$  then the pointer was considered as converging. The standard deviation was greater than  $10^\circ$  for 5.27% of cases (144 over 2730). This means the listener had not succeeded to find a stable pointer position before the end of the sequence. An example is visualised figure 7 c). Interestingly, the majority of these cases occurred with the microphones of order 1 to 3 and very few occurred with the fourth order systems. Most divergences occurred with the SoundField microphone (48.6% of cases or 2.56% of total occurrence). The 8-sensor microphone accounted for 27.8% of cases (or 1.46% of total occurrence) and the 12-sensor microphone for 18%

of cases (or 0.95% of total occurrence). Very few divergences occurred with the 4th order systems (3.5% of cases or 0.18% of total occurrence with the 32-sensor microphone and 2.1% of cases or 0.10% of total occurrence for the synthetic 4th order encoder) as displayed in figure 6. The 8-sensor microphone (third order) accounted higher number of pointer divergences than the 12-sensor (second order). This can be due to the non-optimised frequency range of harmonics recreation of first and second order for the 8-sensor (3), even though a large frequency band noise was used. Furthermore, a general trend shows divergences were mainly occurring for the side target directions (namely  $120^\circ$ ,  $225^\circ$ ,  $270^\circ$  and  $300^\circ$ ).

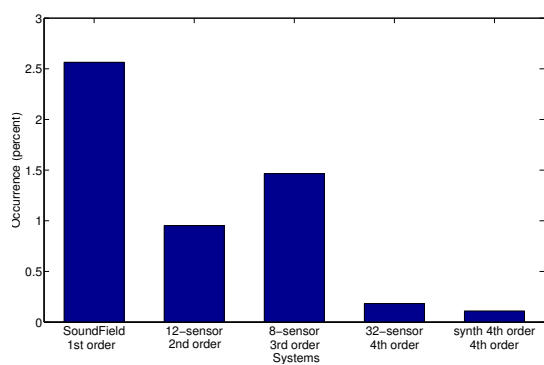


Figure 6: **Percentage of the number of cases where the pointer was diverging during the last five target-pointer presentations for the five systems under test.**

This observation reveals the difficulty of the listener to find an accurate adjustment and a conclusive direction in some cases. All the pointer values were included in the subsequent analysis.

### 3.7.2 Listeners' performance

Prior to comparing the different systems performances, an analysis was conducted on individual performances to verify that the error was brought by the system and not by the ability of the listener to localise a source.

Previous localisation tests demonstrated inter-listener variability in free field and with greater magnitude with reproduction systems [27, 55, 58]. In localisation of natural sources situated in front of the listener, the error is  $1^\circ$  with a localisation blur of  $\pm 3.6^\circ$  for a broad band signal [7].

The last pointer positions were interpreted as the perceived direction of the auditory event reproduced by one of the ambisonic systems. The interquartile range designates 25 % and 75 % of the pointer positions for each target direction and represents the dispersion of the pointer positions for a given target direction. It is interpreted as an uncertainty of matching the pointer to the target direction. Therefore it is an estimate of the system accuracy in reproducing the sound source direction.

The median of the unsigned angular error for all systems and all target directions of each participant varied from  $6^\circ$  to  $12^\circ$  with a maximum interquartile range of  $20^\circ$  for some listeners, highlighting the difficulty of the task. Although the listener's ability to localise real sound sources has not been assessed, the synthetic 4th order system showed results close to natural localisation resolution. Indeed a signed angular error of  $-0.5^\circ$  (with an interquartile range of  $6^\circ$ ) has been calculated for the frontal direction  $0^\circ$  and a signed angular error of  $2^\circ$  (with an interquartile range of  $4^\circ$ ) was calculated for  $352^\circ$ . The pointing error using the synthetic 4th or-

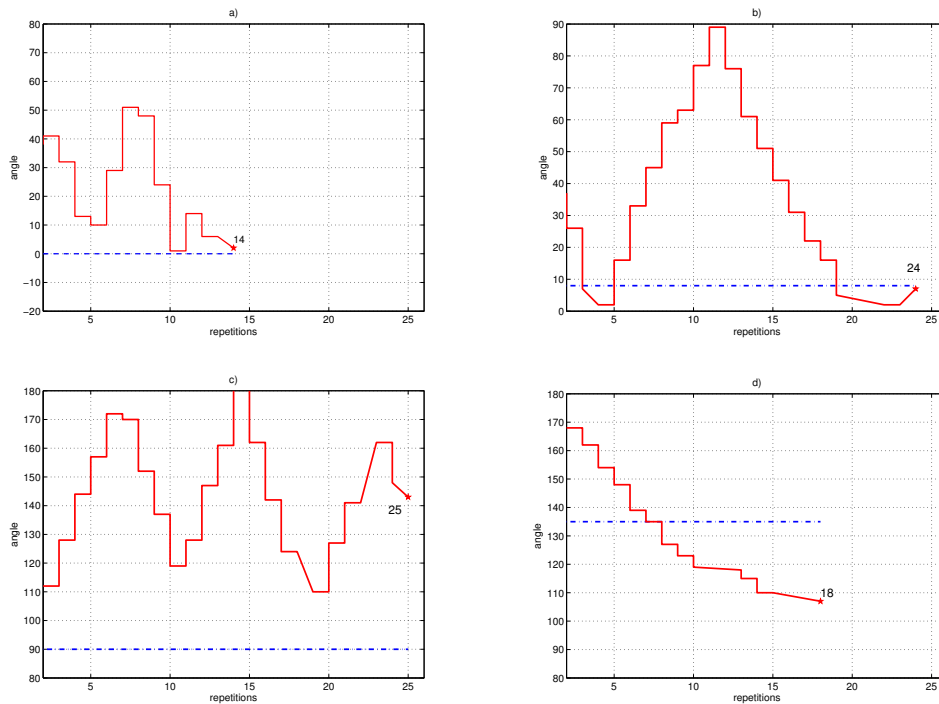


Figure 7: Pointer trajectory examples. The dotted line represents the target direction. The star indicates the last pointer position. The associated number corresponds to the repetition number before validation of the pointer direction by the listener.

der system is within the range of the natural localisation blur for frontal sound source directions. Therefore, we considered that the system could reproduce frontal sound sources and that the listeners were able to report the sound source direction when well reproduced. An analysis of variance (repeated measures) on the unsigned error, based on the following factors: 1) system (SoundField, 12-sensor, 8-sensor, 32-sensor and synthetic 4th order), 2) direction ( $0^\circ$ ,  $15^\circ$ ,  $45^\circ$ ,  $75^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$ ,  $195^\circ$ ,  $225^\circ$ ,  $270^\circ$ ,  $300^\circ$ ,  $330^\circ$  and  $352^\circ$ ), 3) listeners (14) and 4) repetition (1 to 3), revealed that the listeners' performance was significantly different ( $F(13) = 5.62$ ,  $p < 0.01$ ) and that the interaction be-

tween listeners and systems was also significantly different ( $F(52) = 2.6$ ,  $p < 0.01$ ). However the interaction between listeners and source directions was not significantly different ( $F(156) = 1.09$ ,  $p = 0.22$ ), which means that there is a coherency across listener's judgment.

A hierarchical ascendant method established that one listener had different score than the 13 others. From individual results, apart from the fact that he/she obtained an average unsigned error of  $23^\circ$  (total average unsigned error across all listeners:  $12^\circ$ ), it was observed that the listener did not rate differently the SoundField microphone and the 8-sensor microphone, but had "average" discrepancies for the three other systems. Fur-

thermore, some listeners had more discrepancies between the different systems than others, which had more uniform errors across the different systems (results of three listeners were found to be non significantly different for the five systems ( $p = 0.3$ ,  $p = 0.13$  and  $p = 0.03$ )). Surprisingly, this difference was not conducted by the expertise of the listener. Despite these discrepancies, the trend of the results was similar for all the listeners. Therefore all the results were kept for further analysis.

### 3.7.3 Perceived pointer direction

The analysis carried out on the unsigned error (presented in section 3.7.2) revealed a strong influence of the system ( $F(4) = 21.28$ ,  $p < 0.01$ ) and of the direction ( $F(12) = 9.14$ ,  $p < 0.01$ ).

As stated in 3.5, six target positions ( $0^\circ$ ,  $120^\circ$ ,  $150^\circ$ ,  $180^\circ$ ,  $270^\circ$  and  $330^\circ$ ) were located on one of the loudspeakers used to render the ambisonic pointer. The seven other targets were played by loudspeakers located outside the HOA reproduction system. An analysis of variance (repeated measures) was carried out on the unsigned errors considering the two target playback systems. It showed no influence ( $F(1) = 0.28$ ,  $p = 0.6$ ) of the rendering system on the target source. Hence, the matching between the pointer and target direction was not influenced by any spectral similarity related to the reproduction system.

The unsigned angular errors as well as the associated interquartile ranges of the different systems for each direction are displayed in figure 8. Figure 9 displays the perceived directions as a function of target directions for each system. For both figures, results are folded in a semi plane  $0^\circ - 180^\circ$ . The effect of the source direction

from one side of the listener or the other could not be statistically evaluated as no directions were duplicated in both sides, but the results do not show prevalence to one side or the other.

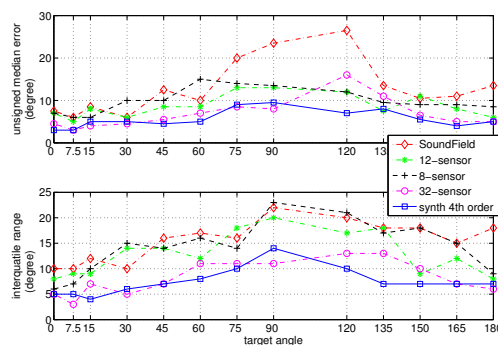


Figure 8: **Top: median of the absolute (unsigned) error for the five systems under test as a function of the target directions. Bottom : associated interquartile range.**

For all systems the frontal target directions ( $0^\circ$ ,  $7.5^\circ$ ,  $15^\circ$  and  $30^\circ$ ) were all well matched. A total signed angular error of  $2^\circ$  (with an interquartile range of  $11^\circ$ ) was measured. The smallest signed angular error  $0^\circ$  (with an interquartile range of  $8.5^\circ$ ) was measured with the synthetic 4th order system. The SoundField microphone had a signed angular error of  $0.5^\circ$  but an interquartile range of  $14.5^\circ$ . The highest signed angular error was measured with the 8-sensor:  $4^\circ$  (with  $13^\circ$  of interquartile range).

For lateral directions ( $45^\circ$ ,  $60^\circ$ ,  $75^\circ$ ,  $90^\circ$ ,  $120^\circ$  and  $135^\circ$ ) larger errors appeared, especially for the SoundField microphone (signed angular error of  $25^\circ$  at  $120^\circ$ ) and for the 8-sensor (signed angular error of  $10^\circ$  and  $12^\circ$  at  $60^\circ$  and  $75^\circ$  respectively). The error was even

more pronounced when considering the interquartile range that varied from  $10^\circ$  (for the 4th order systems at  $45^\circ$ ) to  $50^\circ$  for the SoundField microphone at  $90^\circ$ . Lastly, as illustrated in figures 8 and 9, the error was decreasing for target directions behind the listener ( $150^\circ$ ,  $165^\circ$ ,  $180^\circ$ ).

### 3.7.4 Confusions

Front to back and back to front confusions were observed in 3.85 % of cases (105 over 2730). All pointer positions recorded in the opposite hemisphere to the target direction was considered as a confusion. With this calculation, angular mismatch could have been identified as a confusion. Therefore the confusion rate is displayed in figure 10 as a function of target position. The majority of the confusions was observed for the lateral target directions  $75^\circ$  and  $120^\circ$  (as shown figure 10). The SoundField microphone alone accounted for 39 % of all confusions (or 1.50 % of the total results), the 12-sensor microphone for 17.1 % of confusions (or 0.66 % of the total results), the 8-sensor microphone for 26.7 % (or 1.03 % of the total results) and the fourth order systems (the 32-sensor microphone and the synthetic encoder) accounted for 7.6 % and 9.5 % of confusions (or 0.29 % and 0.37 % of the total results) respectively. In free field Makous and Middlebrooks observed that confusions occurred in average in 6 % of cases when using a pointing method [35]. In sound reproduction, Wenzel et al. measured more than 30 % (mean value) of front to back confusion in binaural reproduction with non-individual HRTF [55]. Guastavino et al. found a confusion rate of 38 % when using a first order ambisonics

system for a localisation task [29]. In natural listening environment, confusions are often resolved by small head movements that slightly change the interaural indices. In our test, the listeners did not have their head fixed and, even though they were asked to not move the head, small movements might have occurred removing localisation ambiguities therefore helping them to resolve some confusions. Wightman and Kistler showed that confusions can be reduced by head movement but also by displacement of the source, only if the listener is responsible for the sound source movement [58]. Even though the target was fixed, the participant was dynamically controlling the movement of the reproduced pointer with a knob. Therefore, the method of report might have helped reducing the front to back confusions.

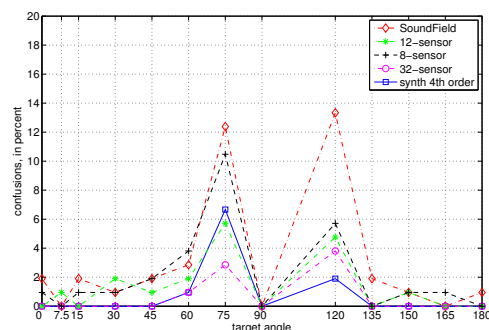


Figure 10: Confusions as a function of azimuth angle

## 4 Objective evaluation of the reproduced sound field

The subjective evaluation concluded that localisation error was increasing for lateral directions and that the



sound source was perceived more accurately when reproduced with higher order systems. To investigate the impairment brought by the system, an objective evaluation was performed on first and higher order ambisonics systems.

## 4.1 Theoretical limits

The blur width characterised by Daniel [12] (introduced in section 1) provides a generic indication of the theoretical performance of the reproduction system. This index can be expanded as a function of the incidence angle of the reproduced source. In localisation of natural sound sources, the reduction of interaural differences leads to a reduction of sound source lateralisation, which means that the sound source is perceived closer to the median plane (of the listener). By projecting the energy vector on the interaural axis and assuming that the listener does not move (or rotate) his/her head, the angle  $\alpha_E$  predicts a deviation angle of the reproduced sound source located at  $\pm 90^\circ$ . More generally according to the energy vector prediction the unity vector that would describe the possible perceived direction has the same projection on the interaural axis  $\vec{y}$  as the energy vector  $\vec{E}$ . It lays on a cone at an angular distance from the interaural axis given by  $\theta_{cone} = \arccos(\vec{E} \cdot \vec{y})$ , one possible position on the horizontal plane being

$$\theta_{deviation} = \arcsin(r_E \cdot \sin(\alpha_E)). \quad (3)$$

The projection of the angular difference between the energy vector and the unity vector varies as a function of the sound source direction and the norm of the energy vector (linked to the ambisonics order). The angular distortion between the ideal and the reproduced source

direction is displayed as a function of the sound source incidence in figure 11. The blur width of the reproduced sound source gets smaller at the side of the listener when the ambisonic order increases. However it is less affected by the system order when the source gets closer to the median plane (in front and at the back of the listener).

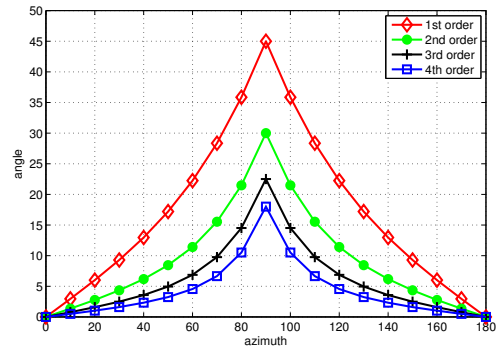


Figure 11: Angular distortion for systems of order 1 to 4 (using a  $\max r_E$  decoding option). Virtual sound source directions are displayed along the x-axis. Azimuth  $0^\circ$  represents the frontal direction.

## 4.2 Simulated reproduction at the listener's ears

In order to characterise the rendering at the listener's ears, the reproduced interaural time difference (ITD) was estimated for all different systems. ITD is used to predict the sound source direction and it is predominant at low frequency. From the duplex theory, above 1500 Hz, interaural level difference (ILD) prevails [50]. Since 1907, this simple theory has been expanded and it has been demonstrated that the fluctuation of the envelope of the signal, visualised with the interaural group

delay (IGD), reinforce the prediction at high frequency [8, 32]. Therefore, a prediction model that integrates the different localisation indices ITD, IGD and ILD was then used to display the behaviour of each system for the entire frequency range.

The SoundField microphone, 12-, 8- and 32-sensor microphones as well as four synthetic first to fourth order ambisonic systems were evaluated. The ambisonic components of each system were decoded using the mix decoder on twelve virtual loudspeakers evenly distributed around the listener. The binaural impulse responses corresponding to the twelve loudspeaker positions were imported from MIT measurements of a Kemar dummy head and weighted by the respective ambisonics filter applied to each loudspeaker.

#### 4.2.1 ITD

Busson et al. demonstrated that the most relevant ITD estimation from a psychoacoustics point of view was obtained with maxIACC methods [9]. The ITD calculation was performed using Gaussian MaxIACCe. This method uses the signal envelope of right and left head related impulse responses (HRIR) and calculate the maximum of intercorrelation between them, to take into account high frequency variations [7]. A gaussian envelope is then applied to the resulting signal of the cross correlation to extract the maximal correlation index. The ITDs for the microphones (SoundField microphone, 12-, 8-, 32-sensor microphones) and for the synthetic first, second, third and fourth order systems are displayed in figure 12. The ITDs of all systems are well reproduced for the front and back positions whereas discrepancies increase with a decreasing order for lateral

positions.

This showed that sound source lateralisation increased along with the system order, giving an indication of the HRIR reproduction quality for a centred listener. Previous studies that investigated first [30, 56, 47] and second [56, 47] order ambisonic systems showed a deterioration of the simulated localisation cues and large localisation errors.

Furthermore, apart from the synthetic 4th order system and the 32-sensor microphone that had a good correlation (98 %), the microphones performed less well than the synthetic system of same order (correlation of 82 % for the SoundField microphone and the synthetic first order system, of 69 % for the 12-sensor microphone and the synthetic second order system and 89 % for the 8-sensor microphone and the synthetic third order system). This suggests that the non-ideal ambisonic components from the microphones influenced the ITD reproduction.

#### 4.2.2 Baskind's model

The ITD calculation synthetically showed the influence of the ambisonic order and of the microphone reproduction. In order to get closer to the mechanism of localisation, a model predicting the direction of a sound source and developed by Baskind [3] was adapted to compare the eight systems. The original model uses temporal signals in input to calculate a global estimation of the direction of the source in azimuth and elevation. In our study, binaural responses in the horizontal plane only were used to calculate a frequency-dependent estimation of the source direction. The error was evaluated as a function of frequency.

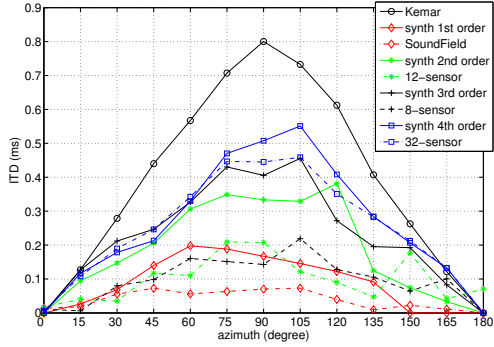


Figure 12: ITD calculated with GaussianMaxIACCe method for the four microphones and the four synthetic systems of respective order one to four and target direction. Angle of the sound source is displayed along the abscissa.  $0^\circ$  designates the frontal direction.

In order to analyse the reproduction on the entire frequency range, the HRIR were divided in 24 equivalent rectangular bandwidth (ERB) frequency bands from 100 Hz to 20 kHz. For each frequency band the interaural indices were calculated. MaxIACC method was used the signal to calculate the ITD and on its envelope to derive the IGD. The ILD was calculated dividing the left and right spectrum for each frequency band.

The interaural indices calculated from a pair of HRIR of one target source direction were compared to the interaural indices calculated from a database composed of HRIR from all space directions (limited by the spatial sampling of the base). Figure 13 illustrates the different steps.

The root mean square error  $d$  equation (4) estimated the distance between the interaural indices of the target and the ones of the database for each frequency band and for all the directions of the database. Therefore, a minimal

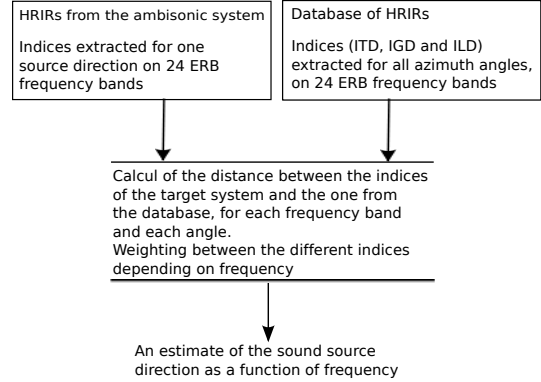


Figure 13: Flow graph illustrating the different steps of the adapted Baskind's estimation model.

error corresponds to an estimation of the most likely target direction. For each frequency band, the contribution of the interaural indices were then weighted, limiting the prevalence of the ITD in high frequency and the one of the IGD and ILD in low frequency. The transition frequency between low and high frequencies was set at 1500 Hz.

$$d(\theta, k) = \sqrt{\left( \alpha(k) \cdot \frac{(ITD(k) - itd^0(\theta, k))^2}{\sigma_{ITD(k)}^2} + \beta(k) \cdot \frac{(IGD(k) - igd^0(\theta, k))^2}{\sigma_{IGD(k)}^2} + \mu(k) \cdot \frac{(ILD(k) - ild^0(\theta, k))^2}{\sigma_{ILD(k)}^2} \right)}, \quad (4)$$

$\theta$  is the azimuth angle,  $k$  the frequency band.  $ITD$ ,  $IGD$  and  $ILD$  are the interaural indices calculated for each frequency band and  $itd^0$ ,  $igd^0$  and  $ild^0$  are the nominal indices.  $\sigma_{ITD(k)}^2$ ,  $\sigma_{IGD(k)}^2$  et  $\sigma_{ILD(k)}^2$  are the standard deviation corresponding to the estimation noise of each index.  $\alpha$ ,  $\beta$  and  $\mu$  are the weights to apply to each corresponding index in the  $k$ th frequency band. The Kemar dataset that has a  $5^\circ$  step between directions, was used. Intrinsicly, the model has a lack of

accuracy at low frequencies and an "estimation blur" in between 1500 Hz and 3000 Hz. However the estimation of the reproduced direction using Kemar's HRIRs as target system as well as database shows a perfect match for the reproduced direction. An example for a sound source direction of  $120^\circ$  is displayed in figure 15.

Figure 14 displays the estimation error of the reproduced direction as a function of frequency for the directions  $0^\circ$  and  $120^\circ$  for the eight systems (microphones and their respective synthetic encoders of order one to four). An error of zero means that the distance between the indices derived from the HRIRs of the ambisonic system and the ones derived from the HRIRs database was null; so that the system was able to reproduce a correct localisation index for that target direction. For the direction  $0^\circ$ , only small errors appeared ( $0.05 < d < 0.15$ ) for the SoundField microphone at 900 Hz and between 6 kHz and 8 kHz. The error is increasing for lateral sound sources directions. At  $120^\circ$  large errors appeared for the first order systems ( $d = 0.89$  for the synthetic first order system and  $d = 0.98$  for the Sound Field microphone at 1400 Hz). For the HOA systems, the error appears at higher frequencies. Errors above 0.2 appeared at 3600 Hz for the second order systems, at 4300 Hz for the third order systems and above 11 kHz for the fourth order systems. Discrepancies were observed between synthetic systems and microphones. The estimation showed bigger errors for the microphones than for the synthetic systems of corresponding order.

Figure 16 displays the estimation error for the eight sys-

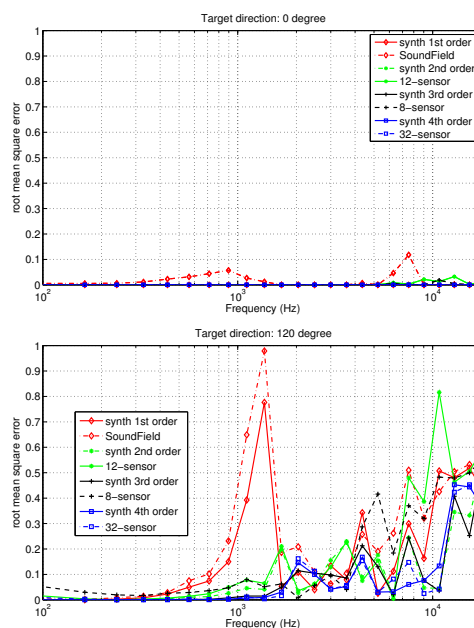


Figure 14: Estimation of the reproduced direction for  $0^\circ$  and  $120^\circ$  for the frequency range 100 Hz - 18000 Hz for the eight ambisonic systems.

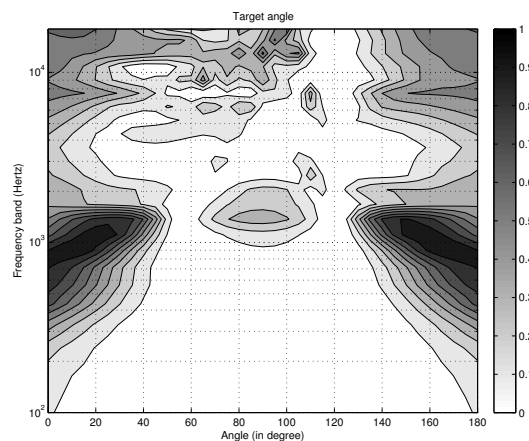


Figure 15: Estimation of the reproduced direction  $120^\circ$  using Baskind's model using Kemar's HRIR. The estimation is displayed for the frequency range 100 Hz - 18000 Hz.

tems as a function of the azimuth. The under lateralisation can clearly be seen for the synthetic first order system and the SoundField microphone from 700 Hz. The estimation error decreases at high frequency with systems of increasing order. Furthermore, the error is minimal for the opposite direction lying on the cone of confusion ( $60^\circ$ ) for all the systems and for the same frequency range as the target direction, unlike the reference estimation calculated with Kemar HRIRs (figure 15). This suggests that the reproduced direction could either be  $120^\circ$  or  $60^\circ$ .

To estimate the impairment introduced by each system on the estimation of the target direction, a correlation between the reference estimation and the one of each ambisonic system was calculated. Table 4 displays the correlation for seven target directions. As expected, a very high correlation is shown for  $0^\circ$  and  $180^\circ$ , decreasing for lateral directions. A discrepancy can be observed between the synthetic system and the microphones (of first, second and third order), revealing a higher correlation with the synthetic systems. The two fourth order systems show similar performances.

## 5 Discussion

In the objective and perceptive analysis the performance of the system was dependent on the ambisonic order and the source incidence. The reproduction of lateral sources brought bigger errors, either in simulation or during the localisation task, inducing a under lateralisation of the source. In a localisation test on first and second ambisonics order systems, Pulkki et al. showed a better localisation for a frontal source that decreased

for a lateral one [47]. Guastavino et al. using a first order ambisonic system with an *inphase* decoder showed a pronounce under lateralisation of the perceived source directions from  $60^\circ$  and a lot of confusions [29]. De-Sena et al. used a second order ambisonic system with an *inphase* decoder and found also an obvious under lateralisation [15]. Even though we used a decoder developed to give the most accurate reproduced sound field [25] and tested as such [6], under lateralisation of source directions was also observed in our localisation test for all the ambisonic systems. It is particularly obvious for the SoundField microphone but this error is visible on the results of all the systems, decreasing with an increasing ambisonic order.

The theoretical deviation angle  $\theta_{distortion}$  (equation 3) expresses, for each ambisonic order, the possible perceived direction of the reproduced sound source as a function of its incidence angle. Figure 12 illustrates the reproduced ITD calculated for each system as a function of the source incidence. These two indices were correlated for each system and their corresponding ambisonic order system, results of the correlation ( $\rho_{ITD}$ ) are displayed in table 5. The computed deviation angle and the ITD were well correlated individually for each system. A global correlation taking into account all systems was performed and resulted in 90% correlation score.

Furthermore, the angular distortion displayed in figure 11 expresses the theoretical angular difference between a reproduced and an ideal sound source direction as a function of its incidence for each ambisonic order. This theoretical angular distortion was correlated with the interquartile range measured during the localisation test

Azimuth	0°	30°	60°	90°	120°	150°	180°
synthetic 1st order	1.00	0.71	0.77	0.66	0.75	0.77	1.00
SoundField microphone	0.98	0.77	0.64	0.53	0.60	0.81	0.99
synthetic 2nd order	1.00	0.89	0.88	0.82	0.90	0.92	1.00
12-sensor microphone	1.00	0.84	0.77	0.70	0.77	0.90	0.98
synthetic 3rd order	1.00	0.87	0.90	0.82	0.91	0.93	1.00
8-sensor microphone	1.00	0.87	0.73	0.65	0.78	0.90	1.00
synthetic 4th order	1.00	0.89	0.92	0.85	0.91	0.96	1.00
32-sensor microphone	1.00	0.89	0.92	0.85	0.92	0.96	1.00

Table 4: Correlation between the estimation of the source direction computed with Kemar and the ones computed with the eight ambisonic systems for the target directions 0°, 30°, 60°, 90°, 120°, 150°, 180°

Systems	Order	$\rho_{ITD}$	$\rho_{iqr}$
SoundField	1	86%	83%
12-sensor	2	94%	85%
8-sensor	3	97%	76%
32-sensor	4	98%	42%
synthetic 4th order	4	99%	79%

Table 5: Correlation between the theoretical indices, objective criteria and subjective results. The third column displays the correlation  $\rho_{ITD}$  between the theoretical deviation angle and the calculated ITD for each system. Fourth column displays the correlation  $\rho_{iqr}$  between the theoretical angular distortion and the interquartile range for each system.

for each system ( $\rho_{iqr}$  in table 5). A correlation of more than 80% was found for the first and second order systems and a correlation of 76% and 79% were obtained for the SoundField microphone and the synthetic 4th order respectively. Only the 32-sensor microphone obtained a correlation of 46%. Part of this low correlation was due to the error around the direction 120°. However, even if this direction was removed from the analysis, a 69% correlation was obtained. The angular distortion assumed a 0° error in front and in the back of the listener, which systems did not achieve in practice. The interquartile range had smaller values across the different angles so the correlation was not as high as the correlation with the ITD indices. However, the overall shape of the results matched the angular distortion.

From the objective evaluation of the various systems, the ITD and the Baskind’s model showed that the performance of the 8- and 12-sensor microphones did not reach the performance of their respective synthetic systems, whereas the 32-sensor microphone showed sim-

ilar results as the synthetic 4th order system. The localisation test shows no significant differences between the results for the synthetic fourth order system and the ones for the 32-sensor microphone. This suggest that the ambisonic components of the first, second and third order microphones impaired the sound field reproduction, even if the SoundField microphone components seemed to be well reproduced on the entire frequency range (table 3).

Nevertheless, using more components increased the accuracy even if the components were not perfectly reproduced on the entire frequency range. Indeed the higher order ambisonic systems showed a more accurate reproduction than the first order one.

However, the 12- and 8-sensor microphones (second and third order) did not show significant differences in the localisation test. Either the third order components did not bring improvement on the precision of the reproduced sound field compared to a second order system or the microphone impairment had an influence on the reproduced sound field. From the spatial correlation analysis (table 3), the components of the 8-sensor microphone were well reproduced on a non optimal limited frequency range. Consequently the performance of such system became equivalent to a second order system. This suggests that the influence of the non-optimal encoding is prominent if the HOA components are reproduced on a limited frequency range. On another hand the higher order ambisonics components of the 32-sensor microphone were reproduced on a limited frequency range as well but the results were much less affected. This range was admittedly larger than the one of the 8-sensor microphone especially at high fre-

quency.

## 6 Conclusion

This study aimed at evaluating objectively and perceptively the spatial resolution of a sound field reproduced by first and higher order ambisonics systems of different complexities. First, a localisation test on a series of microphones of first order (SoundField microphone), second order (12-sensor microphone), third order (8-sensor microphone) and fourth order (32-sensor microphone) and a synthesis 4th order encoder was presented. An objective criterion based on the energy vector's norm was defined, and predicted the distortion angle of a reproduced sound source as a function of its incidence angle. The sound field was then simulated at the listener's ears and localisation indices were calculated. The perceptive and objective evaluation revealed that the ambisonic order as well as the microphones had an influence on the reproduced sound sources. Furthermore, the accuracy depending on the source incidence showing a good reproduction for frontal, and to some extent back sources, even with low order. The accuracy decreased for lateral sources where under lateralisation appeared, especially for low order systems.

## References

- [1] T.D. Abhayapala and A. Gupta. Spherical harmonic analysis of wavefields using multiple circular sensor arrays. *IEEE Trans. Audio, Speech and Language Processing*, 18(6):1655–1666, August 2010.

- [2] Mh Acoustics. <http://www.mhacoustics.com>, 12 2011.
- [3] A. Baskind. *Modèles et méthodes de description spatiale de scènes sonores*. Thèse de doctorat, Université Paris 6, France, 2003.
- [4] E. Bates, K. Gavin, F. Boland, and D. Furlong. Localization accuracy of advanced spatialization techniques in small concert halls. In *153rd meeting of the Acoust. Soc. Am.*, June 2007.
- [5] C. Baume and A. Churnside. Upping the auntie: A broadcaster’s take on ambisonics. In *128th Audio Eng. Soc. Convention*, May 2010.
- [6] E. Benjamin, R. Lee, and Heller A. Localization in horizontal-only ambisonic systems. In *Audio Eng. Soc. 121st Convention*, San Fransisco, October 2006.
- [7] J. Blauert. *Spatial hearing: the psychophysics of human sound localization*. The MIT Press, Cambridge, MA., 1997.
- [8] J. Braasch. *Communication Acoustics*, chapter 4. Modelling of Binaural Hearing, pages 75 – 108. Jens Blauert, Springer, Berlin, 2005.
- [9] S. Busson, R. Nicol, B. Katz, and O. Warusfel. Subjective investigations of the interaural time difference in the horizontal plane. In *Audio Eng. Soc. 118th Convention*, Barcelona, Spain, May 2005.
- [10] S. Carlile, P. Leong, and S. Hyams. The nature and distribution of errors in sound localization by human listeners. *Hearing Research*, 114(1/2):179–196, December 1997.
- [11] P. Craven and M.A. Gerzon. Coincident microphone simulation covering three dimensional space and yielding various directional outputs. U.S. Patent 4,042,779, August 1977.
- [12] 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 Pierre et Marie Curie (Paris VI), France, 2000.
- [13] J. Daniel. Spatial sound encoding including near field effect: Introducing distance coding filters and a viable new ambisonic format. In *Audio Eng. Soc. 23rd Int. Conference*, Copenhagen, Denmark, 2003.
- [14] J. Daniel, J. B. Rault, and J. D. Polack. Ambisonics encoding of other audio formats for multiple listening conditions. In *105th Audio Eng. Soc. Convention*, San Fransisco, California, September 1998.
- [15] Enzo De Sena, Hüseyin Hacıhabiboğlu, and Zoran Cvetković. Perceptual evaluation of a circularly symmetric microphone array for panoramic recording of audio. In *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, May 2010.
- [16] T. Djelani, C. Pörschmann, J. Sahrhage, and J. Blauert. An interactive virtual-environment generator for psychoacoustic research ii: Collection of head-related impulse responses and evaluation of auditory localization. *Acta Acustica united with Acustica*, 86(6):1046–1053, November 2000.



- [17] R. Duraiswami, Z. Li, D. N. Zotkin, E. Grassi, and N. A. Gumerov. Plane-wave decomposition analysis for spherical microphone arrays. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pages 150–153, October 2005.
- [18] N. Epain and J. Daniel. Improving spherical microphone arrays. In *124th Audio Eng. Soc. Convention*, Amsterdam, The Netherlands, May 2008.
- [19] A. Farina and E. Ugolotti. Subjective comparison between stereo dipole and 3d ambisonic surround systems for automotive applications. In *Audio Eng. Soc. 16th Int. Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, March 1999.
- [20] K. Farrar. Soundfield microphone. *Wireless World*, 85(1526):48–50, October 1979.
- [21] K. Farrar. Soundfield microphone - 2. *Wireless World*, 85:99–103, November 1979.
- [22] M. A. Gerzon. Criteria for evaluating surround-sound systems. *Journal Audio Eng. Soc.*, 25(6):400–408, June 1977.
- [23] M. A. Gerzon. Ambisonics in multichannel broadcasting and video. *Journal Audio Eng. Soc.*, 33(11):859–871, November 1985.
- [24] M. A. Gerzon. General metatheory of auditory localisation. In *92nd Audio Eng. Soc. Convention*, 1992.
- [25] M. A. Gerzon. Optimum reproduction matrices for multispeaker stereo. *Journal Audio Eng. Soc.*, 40(7/8):571–589, July/August 1992.
- [26] R. H. Gilkey, M. D. Good, M. A. Ericson, J. Brinkman, and J. M. Stewart. A pointing technique for rapidly collecting localization responses in auditory research. *Behavior Research Methods, Instruments and Computers*, 27(1):1–11, 1995.
- [27] M. Gröhn. Localization of a moving virtual sound source in a virtual room, the effect of a distracting auditory stimulus. In *International Conference on Auditory Display*, Kyoto, Japan, July 2002.
- [28] C. Guastavino and B. Katz. Perceptual evaluation of multi-dimensional spatial audio reproduction. *Journal Acoust. Soc. Am.*, 116(2):1105–1115, August 2004.
- [29] C. Guastavino, V. Larcher, G. Catusseau, and P. Boussard. Spatial audio quality evaluation : comparing transaural, ambisonics and stereo. In *13th International Conference on Auditory Display*, pages 26–29, Montréal, Canada, June 2007.
- [30] J. M. Jot, V. Larcher, and Pernaux J. M. A comparative study of 3-d audio encoding and rendering techniques. In *Audio Eng. Soc. 16th Int. Conference on Spatial Sound Reproduction*, Rovaniemi, Finland, March 1999.
- [31] E. H. Langendijk and A. W. Bronkhorst. Collecting localization response with a virtual acoustic pointer. *Journal Acoust. Soc. Am.*, 101(5):3106, May 1997.
- [32] E. A. Macpherson and J. C. Middlebrooks. Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. *Jour-*

- nal Acoust. Soc. Am.*, 111(5):2219–2236, May 2002.
- [33] Piotr Majdak, Matthew J. Goupell, and Bernhard Laback. 3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training. *Attention, Perception, and Psychophysics*, 72(2):454–469, February 2010.
- [34] Y. Makita. On the directional localisation of sound in the stereophonic sound field. *EBU Review*, A(73):102–108, 1962.
- [35] J. C. Makous and J. C. Middlebrooks. Two-dimensional sound localization by human listeners. *Journal Acoust. Soc. Am.*, 87(5):2168–2180, May 1990.
- [36] D. Malham. Experience with large area 3d ambisonic sound systems. *Proc. of the Institute of Acoustics*, 14, 1992.
- [37] A. W. Mills. Minimum audible angle. *Journal Acoust. Soc. Am.*, 30(4):237–246, April 1958.
- [38] 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*. PhD thesis, Université du Maine, France, 2006.
- [39] S. Moreau and J. Daniel. Study of higher order ambisonic microphone. In *7e Congrès Français d’Acoustique - DAGA’06*, Strasbourg, France, 2004.
- [40] S. Moreau, J. Daniel, and S. Bertet. 3d sound field recording with higher order ambisonics - objective measurements and validation of a 4th order spherical microphone. In *120th Audio Eng. Soc. Convention*, Paris, France, May 2006.
- [41] Rozenn Nicol. Sound spatialization by higher order ambisonics: Encoding and decoding a sound scene in practice from a theoretical point of view. In *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France, May 2010.
- [42] S. R. Oldfield and S. P. A. Parker. Acuity of sound localisation : a topography of auditory space. i. normal hearing conditions. *Perception*, 13(5):581–600, 1984.
- [43] M. Park and B. Rafaely. Sound-field analysis by plane-wave decomposition using spherical microphone array. *Journal Acoust. Soc. Am.*, 118(5):3094–3103, 2005.
- [44] A. Parthy, C. Jin, and A. van Schaik. Evaluation of a concentric rigid and open spherical microphone array for sound reproduction. In *Ambisonics Symposium*, Graz, June 2009.
- [45] M. A. Poletti. The design of encoding functions for stereophonic and polyphonic sound systems. *Journal Audio Eng. Soc.*, 44(11), 1996.
- [46] M. A. Poletti. A unified theory of horizontal holographic sound systems. *Journal Audio Eng. Soc.*, 48(12):1155–1182, December 2000.
- [47] V. Pulkki and T. Hirvonen. Localization of virtual sources in multichannel audio reproduction. *IEEE Transactions on Speech and Audio Processing*, 13(1):105–119, January 2005.

- [48] B. Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on Speech and Audio Processing*, 13(1):135–143, January 2005.
- [49] B. Rafaely. The spherical-shell microphone array. *IEEE Trans. Audio, Speech and Language Processing*, 16(4):740–747, May 2008.
- [50] Lord Rayleigh. On our perception of sound direction. *Phylosophy Magazine*, 13:214–232, 1907.
- [51] B. Seeber. A new method for localization studies. *Acta Acustica united with Acustica*, 88(3):446–450, 2002.
- [52] A. Solvang. Spectral impairment for two-dimensional higher order ambisonics. *Journal Audio Eng. Soc.*, 56(4):267–279, April 2008.
- [53] A. Sontacchi, M. Noisternig, P. Majdak, and R. Höldrich. Subjective validation of perception properties in binaural sound reproduction systems. In *Audio Eng. Soc. 21st Int. Conference*, St Petersburg, Russia, June 2002.
- [54] J. Trevino, T. Okamoto, Y. Iwaya, and Y. Suzuki. High order ambisonic decoding method for irregular loudspeaker arrays. In *Proceedings of 20th International Congress on Acoustics*, August 2010.
- [55] E. Wenzel, M. Arruda, D. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *Journal Acoust. Soc. Am.*, 94(1):111–123, July 1993.
- [56] B. Wiggins, I. P. Stephens, and P. Schillebeeckx. The analysis of multi-channel sound reproduction algorithms using hrtf data. In *Audio Eng. Soc. 19th Int. Conference : Surround Sound - Techniques, Technology, and Perception*, Schloss Elmau, Germany, June 2001.
- [57] F. Wightman and D. Kistler. Headphone simulation of free field listening II: Psychophysical validation. *Journal Acoust. Soc. Am.*, 85(2):868–878, February 1989.
- [58] F. L. Wightman and D. J. Kistler. Resolution of front – back ambiguity in spatial hearing by listener and source movement. *Journal Acoust. Soc. Am.*, 105(5):2841–2853, May 1999.
- [59] P. Woloszyn. An acoustic ambience study by immersive sound recognition. In *European conference "Building with sounds"*, Paris, France, March 2005.
- [60] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov. Sound field decomposition using spherical microphone arrays. In *IEEE ICASSP*, pages 277–280, Las Vegas, NV, April 2008.
- [61] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov. Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays. *IEEE Signal Processing Society*, 18(1), January 2010.
- [62] E. Zwicker and E. Terhardt. *Facts and Models in Hearing*. Springer - Verlag, Berlin, 1974.

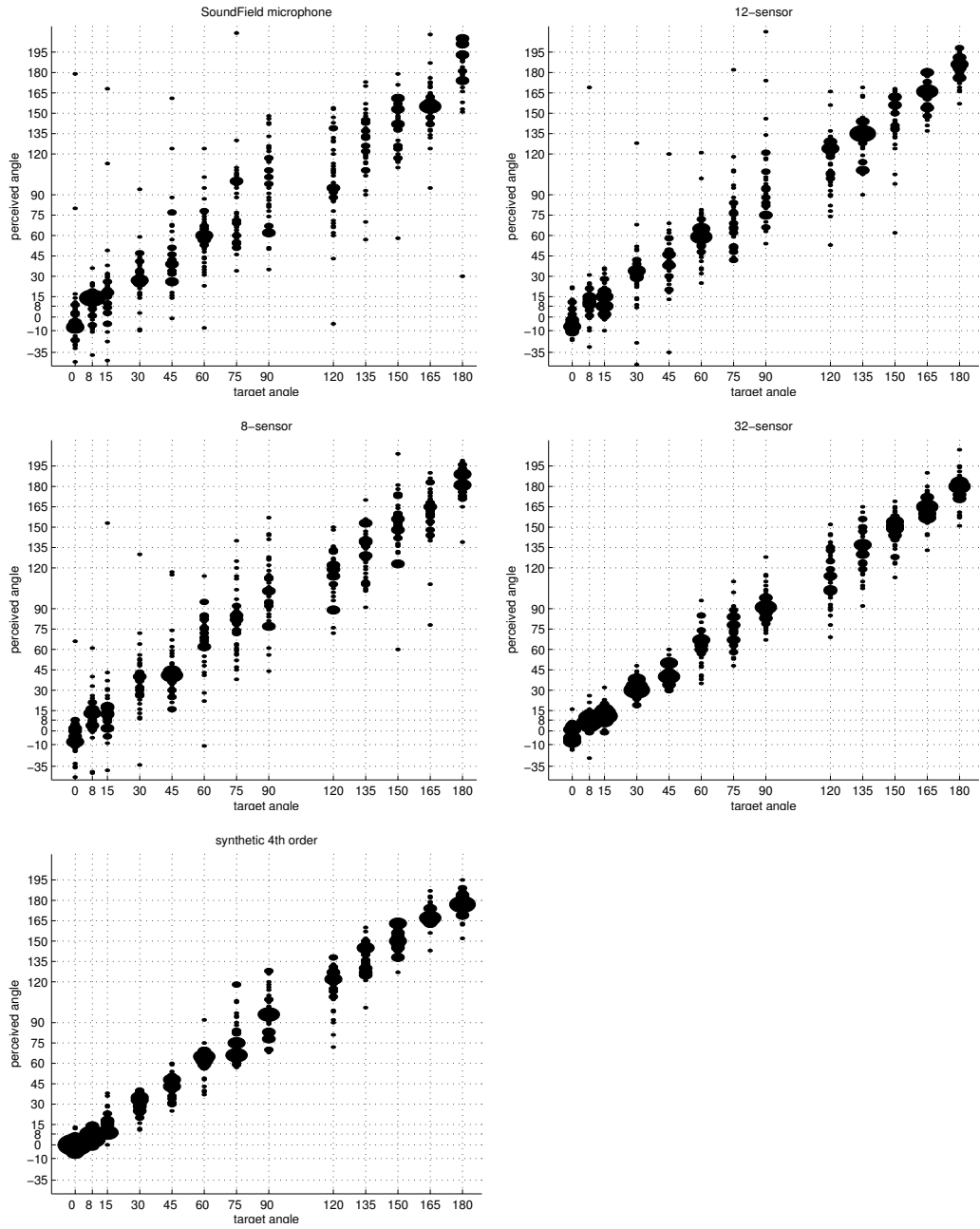


Figure 9: Perceived angle as a function of target angle for the five systems

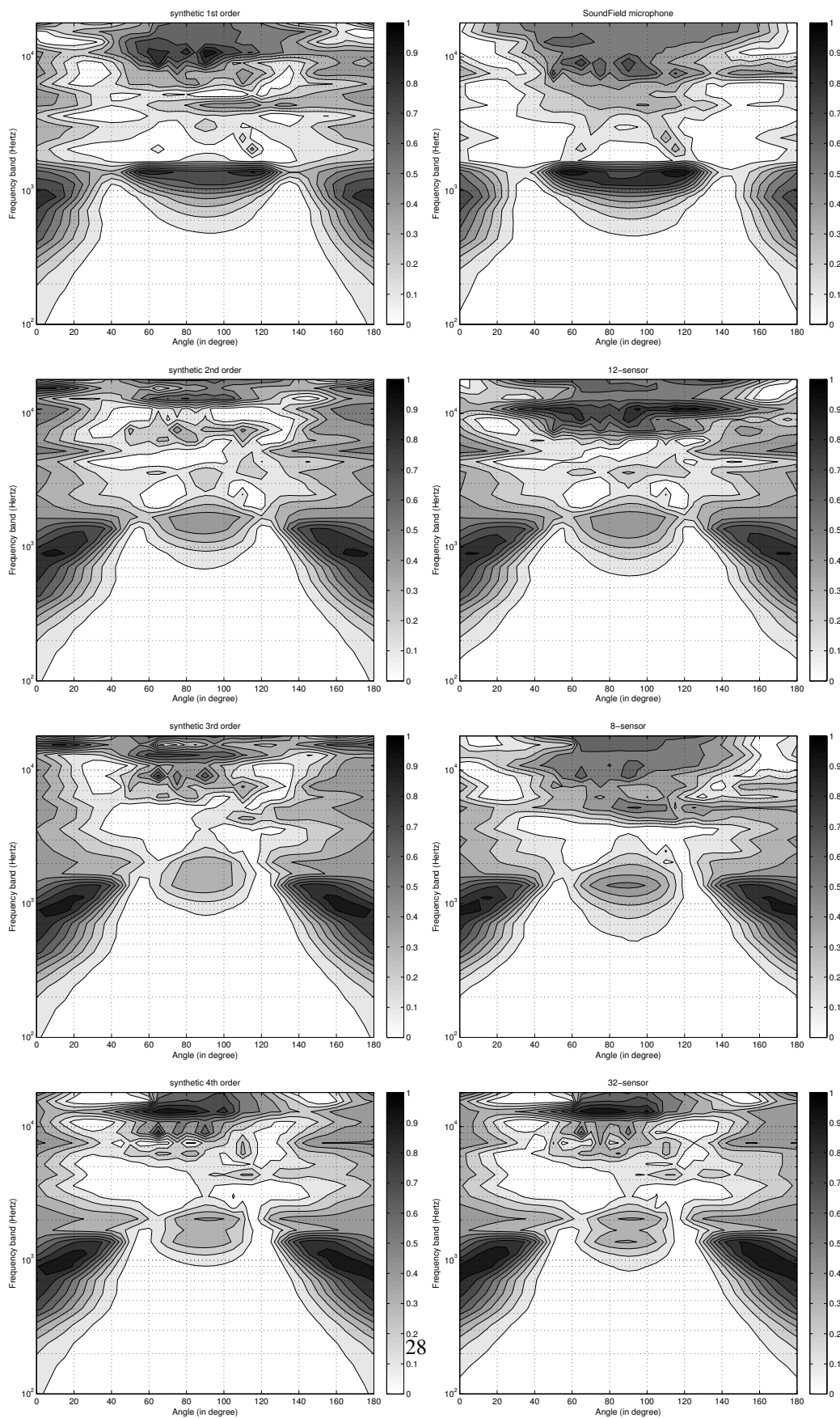


Figure 16: Estimation of the reproduced direction  $120^\circ$  using Baskind's model for the synthetic systems of order 1 to 4 and their corresponding microphones. The estimation is displayed for the frequency range 100 Hz - 18000 Hz.