AMBISONICS SYMPOSIUM 2009
June 25-27, Graz

INFLUENCE OF MICROPHONE AND LOUDSPEAKER SETUP ON PERCEIVED HIGHER ORDER AMBISONICS REPRODUCED SOUND FIELD

Stéphanie Bertet[1], Jérôme Daniel[2], Etienne Parizet[3], Olivier Warusfel[1]

[1] Dep. Acoustiques des salles, IRCAM, Paris, France (stephanie.bertet@free.fr)
[2] Orange Labs TECH/OPERA, Lannion, France
[3] Laboratoire Vibrations Acoustiques, INSA Lyon, France

Abstract:
Among the spatial audio reproduction techniques, the ambisonic approach is based on a spherical harmonics sound field decomposition. By truncating the decomposition to the Mth order, a finite number of ambisonic components that form the spatial ambisonic format remains and gives a partial recreation of the sound field. The higher the order M is, the more accurate the sound field is reproduced. Microphone arrays are used to encode natural sound field into spatial components. The encoded sound field is then decoded for a dedicated reproduction system. The goal of this study was to evaluate the influence of these devices. In a first experiment, four ambisonic microphones (from first to fourth order) were evaluated. Six sound scenes were reproduced over a fixed loudspeaker setup. In a second experiment, synthetic encoding processes from first to fourth order were reproduced on different loudspeaker configurations. Besides the ambisonic order, the encoding and reproduction systems also had a perceived influence on the reproduced sound field.

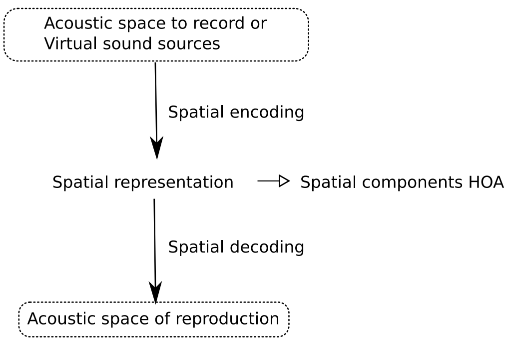
Key words: subjective test, HOA systems

1 INTRODUCTION

The aim of 3D sound reproduction systems is to reproduce the timbral and spatial information of sound sources as they are found in the original scene. It means that the direction of each sound source situated in the recording area should be respected when the scene is played back.

Among 3D sound reproduction techniques, the Ambisonics and Higher Order Ambisonics (HOA) technologies have some advantages. These technologies involve a spatial decomposition of the sound field. It is encoding in an intermediate format allowing a flexible selection of the reproduction system.

The ambisonics systems are decomposed in three steps : sound field recording or sound synthesis, data transmission and sound field reproduction.



This technology is based on spatial sound field decomposition using spherical harmonics. The more components are used, the more accurate the sound field is reproduced [7]. The sound field is decomposed on the spherical harmonics basis, forming the ambisonics components. The aim of the reproduction is to obtain at the center of a loudspeaker setup the same sound field than the original one. Then, the encoded sound field, represented by the ambisonic components vector is transformed by a spatial decoding process to be reproduced in the sweet area. The decoding process recreates the encoded sound field for a reproduction system. For a two-dimensional listening setup, the classic reproduction setup is an evenly distributed loudspeaker configuration. The reproduction setup should be composed by at least $2M + 1$ loudspeakers where M is the ambisonic order. Gerzon advises the use of more loudspeakers than this number to avoid the detent effect (where the sound is pulled toward the closest loudspeaker) [11].

In order to optimize the sound field reconstruction, different decoding options have been developed. The basic decoder rebuilds faithfully the ambisonic components. However, it is limited in its reconstruction of large areas and at high frequencies. The $\mathrm{maxr_E}$ decoder tries to optimize the energy vector in order to satisfy the energy preservation criteria for a central area. These two decoders combined (basic for low frequencies and $\mathrm{maxr_E}$ at high frequencies) try to rebuild as close as possible the localization cues [6, 5, 12]. For a large listening area, Malham suggests a

Figure 1: SoundField microphone, 12-sensor microphone (second order) and 32-sensor microphone (fourth order)

controlled-opposite or inphase decoder [16].

The B-format represents the first harmonics of an angular sound field decomposition. The higher order ambisonics system includes spherical harmonics of higher orders. To reproduce a "real" sound field, ambisonic and HOA microphones have been developed over the years [9, 10, 8, 17, 18, 19]. We concentrate our study on some of them.
The commercialized SoundField microphone is composed of four coincident sensors. The sensor signals are combined to obtain the first-order ambisonic components (figure 1).

The higher order components cannot be built by linear combination due to their complexity. In order to build a HOA microphone, a rigid sphere where sensors are evenly distributed presents advantages. The HOA components are built using the diffraction properties of the sphere [19]. Therefore, a compromise has to be made between the size of the sphere (for reproduction at low frequency) and the number of sensors. The later determines the aliasing frequency limit at high frequency. These limits mean a partial reconstruction of the spherical harmonics over the frequency range. To optimize the recreation of HOA components and tend to push these limits at low and high frequency, filters can be computed using sensors' responses [15, 19].

OrangeLabs has built two higher-order microphones of second and fourth order (figure 1). The second-order prototype is composed of twelve sensors placed in dodecahedral configuration on a semi rigid sphere 7 cm in diameter. The fourth-order prototype is composed of thirty-two sensors placed in a pentaki dodecaedron on a semi rigid sphere of 7 cm in diameter. Then, the three ambisonic microphones, the SoundField microphone and the two HOA microphone prototypes were measured at IRCAM in the anechoic room. Their characteristics have been studied and integrated into a complete reproduction system to subjectively evaluate the recreated soundfield (details about the measurements and objective studies can be found [1, 19]).

In the first experiment, we evaluate the influence of using recording systems to reproduce a sound field. The test focuses on spatial accuracy and sound source direction perceived at the center of the reproduction area. The reproduction setup consists of twelve loudspeakers evenly distributed in a circle.

The second experiment deals with the influence of the loudspeaker configuration on a reproduced sound scene of a given order. The encoding process is then done with synthetic components for a reproduction setup with varying numbers of loudspeakers. The decoding conventions remain the same across all studies. A combined basic and $\mathrm{maxr_E}$ decoder is used. Shelf filters control the transition between the low and high frequency filters.

2 FIRST EXPERIMENT

The goal of this experiment is to evaluate the performance of the SoundField microphone and the HOA microphone prototypes built at Orange Labs. In theory the accuracy of the reproduced sound field increases with the order. Localisation tests showed the benefit of higher order components on localization using synthetic encoding systems [20, 23]. Also, a previous evaluation of the studied microphones showed an improvement using higher order systems by adjusting an encoded broadband noise sound source to a physical sound target [2]. To get closer to more realistic sound contents, synthetic sound scenes are created with voices and everyday sounds taking into account the microphones characteristics. A MUSHRA-like test focusing on perceived spatial resolution and spatial quality is carried out.

2.1. The MUSHRA test ITU - R. BS 1534

A MUltiple Stimuli with Hidden Reference and Anchor test is typically used for testing the quality of audio codecs. It follows the ITU - R BS. 1534 report [13]. This report defines a test procedure to evaluate systems of intermediate audio quality. It compares systems to a reference and between each other. All systems are presented at the same time (MUltiple Stimuli) to the listener. A hidden reference and an anchor (a 3.5 kHz low pass filtered signal) are part of the presented systems. A continuous quality scale is used for the evaluation. It is divided in five intervals denominated by adjectives. Test instructions specify the kind of signal degradation.
In our case, the spatial quality of the recreated sound field is evaluated. Thus, the MUSHRA test principle has been modified and does not include anchor.

2.2. Systems under test

The measured microphones are tested as well as a synthetic fourth order encoding system :

- the SoundField, first order ambisonic microphone
- the second order microphone prototype, denoted in the following as the *12 sensors*
- the fourth order microphone prototype denoted as the *32 sensors*,
- a third order system constituted by the 8 sensors placed in the horizontal plane of the 32-sensor microphone

(*the 8 sensors*),

- a theoretical fourth order encoding system (*ideal* 4^{th} *order*).

The measured impulse responses of each system have been considered to generate the HOA components. The impulse responses of the SoundField microphone have been measured in B-format (W, X, Y, Z signals) directly.

At least $2M + 2$ loudspeakers should be used to reproduce an encoded sound field of order M. Twelve loudspeakers evenly reparted among a 48-loudspeaker circle compose the reproduction setup in the horizontal plane. In order to compensate for the influence of the tranducers and the imperfect concentricity of the structure, the loudspeakers are measured at the center of the listening area. Their responses are inversed, applying a frequency-dependent regularisation factor to limit the inversion effort in high frequencies.

2.3. Sound scenes

To differenciate the five systems, six scenes are built using monophonic sound sources placed around the listener in the horizontal plane. Sources directions are chosen among the 48-loudspeakers circle. In each scene, the sources directions are inspired by the localisation test results (localisation blur, front-back confusions). Four scenes simulate audio meetings with three or four persons, two scenes where voices are placed in front of the listener and two scenes where the sources are placed around the listener. Conversations are not coherent to help the listener focusing on the direction of the source and not on the meaning of the talk. Scenes last between 8 and 14 seconds. By limiting the number of sources and the length of the scene we suppose that the listener focuses on the information of the all scene to establish his/her judgment. Sources are played one after the other with overlaps (figure 2).
Two scenes are composed by environemental sound sources. Contrary to the meeting scenes these scenes simulate sounds in a kitchen and in a classroom. The sources are continuous or coherent between each other helping the listener to immerse himself in the scene atmosphere. However dry sources are used that could limit the realism of these scenes.

The reference scenes are built in associating each sound source to the corresponding loudspeaker. The ambisonic and HOA scenes are created by encoding each sound source (and direction) on the recreated ambisonic components of each system.

2.4. Procedure

The test took place in one of a listening room of Orange Labs. The listener was placed at the center of the loudspeakers circle. The loudspeakers were hidden by an acoustically transparent curtain and a mark indicated the frontal direction. In front of the listener, a graphical interface where six systems under test (the five ambisonics systems and the hidden reference) and the reference system are displayed

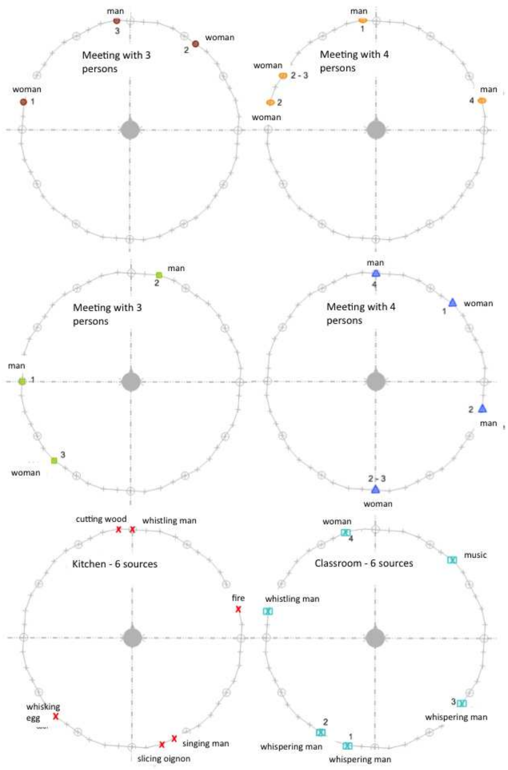


Figure 2: Placement, nature and playing order of the sound sources for the six synthetised scenes.

enabled him/her to manage the test sequences. It is composed of buttons and cursors to evaluate each system according to the defined scale. The evaluation pointed out the spatial quality of the systems. The scale was divided in five intervals qualified by adjectives (in french) and went from 0 to 100 (figure 3). Each system could be played as much as the listener wants and he/she could switch from one system to the other whenever he/she wanted in the scene. Before the test a learning phase included another scene to familiarize the listener with the task. Then, the six scenes were randomly presented for comparison to the listener. The test lasted around thirty minutes.

2.5. Subjects

Eighteen participants (seventeen men and one woman) passed the test. Twelve of whom were experienced. They reported no hearing problem but their hearing threshold had not been measured. The results of one listener have not be taken into account for the analysis since he did not find all the hidden references. The results of the seventeen listeners are analysed.

2.6. Analysis

Data of all seventeen listeners are retained and analysed even though five of them are not experts. Globally there is a bigger standard deviation for the group of naive listeners than for the one of experts. However a analysis of variance

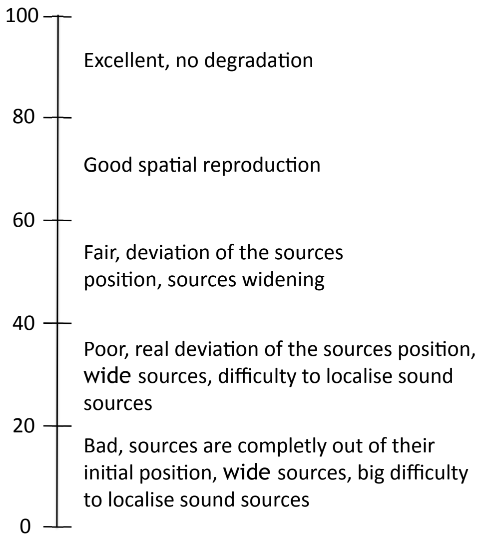


Figure 3: Scale helping participants to rate systems (the adjectives are in french)

at one factor (naive or expert) is carried out on scores. No significant differences are found between the two groups of listeners (F(1) = 2.17, p = 0.14).

An analysis of variance is carried out on scores considering the factors recording system (the SoundField, the 12 sensors, the 8 sensors, the 32 sensors, the ideal 4^{th} order and the reference sytem) and scene (kitchen, classroom, front meeting three persons, surround meeting three persons, front meeting four persons, surround meeting four persons). The analysis reveals a significant principal effect of factor system ($F(5) = 192.08$, $p < 0.01$). However, there is no significant principal effect of factor scene ($F(5) = 1.25$, $p = 0.292$). The figure 4 shows the mean scores of the six systems for each scene. There is a small effect of the interaction system scene on scores ($F(25) = 1.59$, $p = 0.0378$). Globally we can observe that each system has a score quite homogeneous among all scenes.

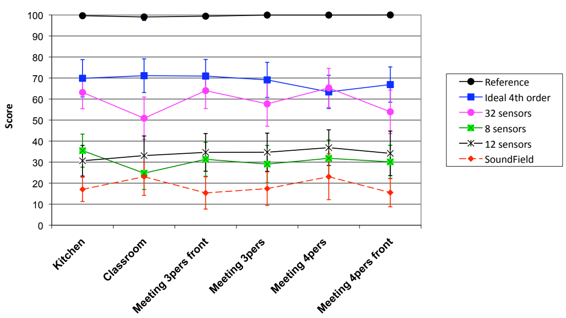


Figure 4: Mean scores and 95% confident interval for each system. The results of the seventeen participants are grouped by scene

The mean scores of all scenes are computed for each sys-

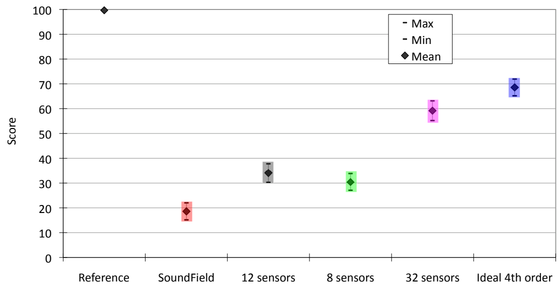


Figure 5: Mean scores and 95% confident interval for each system. The results of the seventeen participants and of the six scenes are grouped

tem (figure 5). A post-hoc test (Tukey HSD test) on scores reveals five groups of systems :

- the SoundField microphone (first-order system)
- the 12 sensors (second-order microphone) and the 8 sensors (third-order microphone)
- the 32 sensors (fourth-order microphone)
- the ideal 4^{th} order
- the reference system

Projecting scores on the rating scale (figure 6), the SoundField is judged like "bad" and does not reconstruct the source direction. The 12 and 8 sensors (second- and third-order microphones) are "poor" in term of spatial quality and recreate sound source direction with deviation. The scenes recreated by the 32-sensor system (fourth-order microphone) are judged like a "fair" spatial reproduction of the original sound scenes. Finally, eventhough the ideal fourth-order system is significantly different of the reference system, it is judged like making a "good" spatial reproduction.

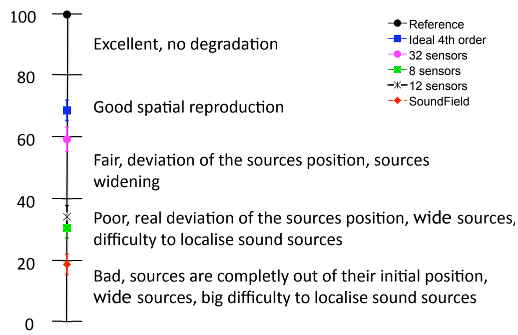


Figure 6: Scores projected on the rating scale

This test reveals a significant difference between the synthetic fourth order encoding system and the 32-sensor microphone. This result dissents from the localisation test carried out on these systems in which the two fourth order systems give equivalent results [2]. The use of real sound sources whom spectrum is not constant over frequency (broadband noise was used in the localisation test) highlights the encoding differences between a synthetic and a microphone system, Furthermore, the second and third order systems show no difference (in phase with the localisation test results).

Eventhough there is a significant difference between the fourth-order systems, the distance between the 8 sensors (third-order system) and the 32 sensors (fourth-order system) is obvious and more important.

3 SECOND EXPERIMENT

The first experiment is based on the evaluation of different ambisonics recording devices decoded for a 12-loudspeaker setup. The reproduction system is fixed. In the second experiment, the reproduction setup configuration is studied. A pairwise comparison test is used to evaluate the difference between synthetic ambisonic systems from first to fourth order decoded for two loudspeaker configurations.

3.1. Systems under test

This evaluation is performed on the reproduction systems using synthetic encoding process instead of microphone responses. Eight systems are under evaluation :

- a first-order encoding system decoded on four (minimum number of loudspeakers for a first order reproduction system) and twelve loudspeakers (number of loudspeakers used in the first experiment), named *o1spk4* and *o1spk12* respectively
- a second-order encoding system decoded on six (minimum number of loudspeakers for a second order reproduction system) named *o2spk6* and twelve loudspeakers (*o2spk12*)
- a third-order encoding system decoded on eight (minimum number of loudspeakers for a first order reproduction system), named *o3spk8* and twelve loudspeakers (*o3spk12*)
- a fourth-order encoding system decoded on twelve loudspeakers named *o4spk12* (the configuration with the minimum number of loudspeaker - ten - cannot be reproduced on a 48-loudspeakers array).
- a reference (*ref*) where the sound source are played through a physical loudspeaker placed in the right direction.

As in the first experiment, the loudspeakers are measured at the center of the listening area. Their responses are inversed, applying a frequency-dependent regularisation factor to compensate for their influence and the imperfect concentricity of the structure.

3.2. Sound scenes

Two sound scenes are chosen among the scenes created for the first experiment : a meeting scene with three talkers in front of the listener and a surround scene with everyday sound sources simulating a sound environment of a kitchen.

3.3. Procedure

A pairwise comparison was performed on the eight systems under test. Twenty-eight pairs ($\frac{8\times7}{2}$) were built and presented randomly to the listener for each scene. The dissimilarity between systems was judged on a horizontal continuous scale. The evaluation is not attribute oriented (e.g. spatial impression, timbre, natural sensations..), the "global" perception difference between systems has been evaluated.

The test took place in one of a listening room of Orange Labs. The listener was placed at the center of the loudspeakers circle. The loudspeakers were hidden by an acoustically transparent curtain and a mark indicated the frontal direction. In front of the listener, a graphical interface displayed two buttons corresponding to the two systems to compare. The listener had to rate the difference on a horizontal scale ranging from "identical" to "very different" (from 0 to 100). Each system could be played as much as the listener wanted and he/she could switch from one system to the other whenever he/she wanted in the scene.
A preliminary listening phase presented the eight systems to the listener to show the difference he/she could find in the test between systems. Then, a learning phase composed of seven comparisons has been done on a different scene to familiarise the listener to the task. At last, 31 pairs (28 pairs + 3 pairs of control, where the same system are presented twice) were compared for each scene. The test was divided into two sessions, one per scene. Half of the listeners started by the surround scene "kitchen" (twelve persons), the other half by the frontal scene "meeting".

3.4. Subjects

Twenty-five listeners passed the test (five women and twenty men), Fourteen of whom were experienced. Twelve of the listeners have done the first experiment. All listeners reported no hearing problem but their hearing threshold had not been measured.

3.5. Raw data analysis

The dissimilarity between systems are collected for each subject and both scenes. Besides the pairs of control which have very low scores, mean scores are spread from 1.68 over 100 (dissimilarity found for the pair of systems o3spk12 / o4spk12) to 83.6 where the dissimilarity between system o1spk12 and the reference system is the largest. A global mean reachs 48.78 for the scene "kitchen" (figure 8).

For the scene "meeting", mean scores go from 6.16 (between systems o3spk12 and o4spk12) to 83.1 (between systems o1spk12 and the reference). The global mean is 49.39 (figure 7).

A paired t-test with 1% error (α = 0.0001, t = 3.75) has been carried out on the dissimilarity scores of each couple of systems . Only the pairs of control and the pairs o3spk12 / o4spk12 and o2spk12 / o3spk12 contain systems with no significant dissimilarities. Scores for the other 26 pairs have been analysed significantly different from zero. This means that the listeners were able to distinguish one system from the other.

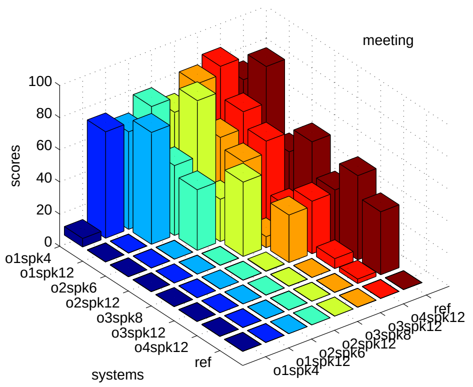


Figure 7: Mean scores of dissimilarity between systems for the frontal scene "meeting"

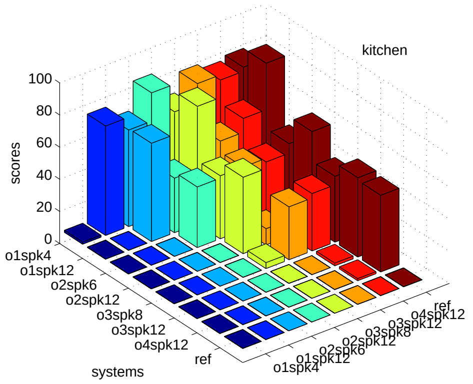


Figure 8: Mean scores of dissimilarity between systems for the surround scene "kitchen"

Globally, the couples in which the first-order systems are involved show the highest dissimilarity judgement. In the other hand, the couples of systems decoded on twelve loudspeakers o2spk12 / o3spk12, o3spk12 / o4spk12, and o2spk12 / o4spk12 obtain scores below 20 for the two sound scenes.

A correlation between the mean scores for each pair for the two scenes obtain 95 %.

3.6. Multidimensional Scaling analysis

A multidimensional scaling (MDS) analysis allows a representation of the dissimilarity in terms of perceptual distance in a space X with Q dimensions. The classical MDS estimates the euclidian distance d_{ij} between the object i and the object j [3]. This distance is expressed as a function of the perceived dissimilarity, $f(\delta_{ij}) = d_{ij}(X)$, where

$$d_{ij} = \sqrt{\sum_{q=1}^{Q-1} (x_{iq} - x_{jq})^2} \qquad (1)$$

the indice q is the dimension, Q the number of dimensions in the space X, and x represents the stimuli coordinates in this space. This transformation is based on triangle inequality that expresses the most direct path is the shortest ($d(i,j) \leqslant d(i,k) + d(k,j)$). If this inequality is not verified between objects i and j in space X, their projection implies an error $e(_{ij}) = f(\delta_{ij}) - d_{ij}(X)$.

The goal of this analysis is to define the space that best fits the given data (minimal error) with least number of dimensions possible (model complexity).

The first step of the analysis is to find the suitable number of dimensions required to define the perceived space. The stress calculates the mean square error between the dissimilarity matrices and the model. It is a quantitative measure of the adjustment between the measured data and the found configuration of Q dimensions [14]. Thus, the smaller the stress value, the better is the fit of the reproduced distance matrix to the observed distance matrix. In our case the stress reduces of 60 and 70 % between a configuration at one and two dimensions (figure 9).

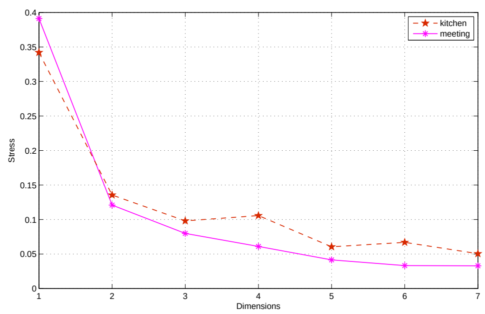


Figure 9: Stress - mean square error between the dissimilarity matrices and the model

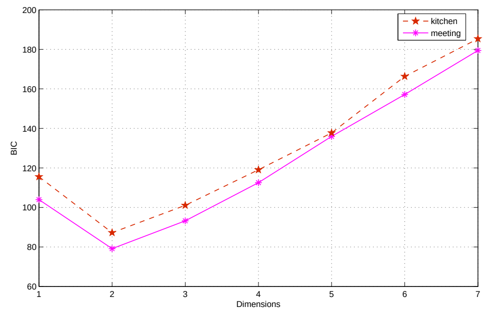


Figure 10: Bayesian Information Criterion computed Lee formula

Furthermore, model quality can be estimated using the bayesian information criterion (BIC) [21]. Based on

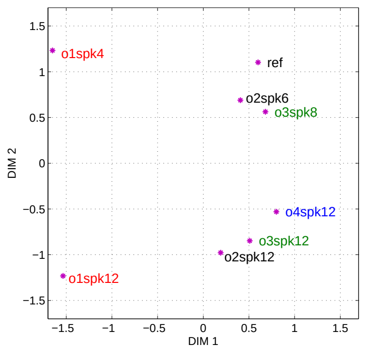


Figure 11: 2-dimensionnal space for the frontal scene meeting

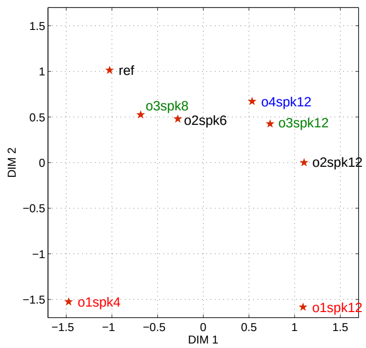


Figure 12: 2-dimensionnal space for the surround scene kitchen

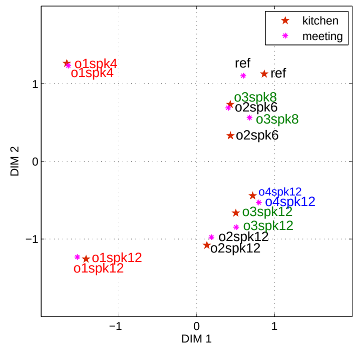


Figure 13: Procrustean transformation of the space of the kitchen scene on the space of the meeting scene

maximum likelyhood, the number of parameters and the number of sample, it is a compromise between the adjustment quality of the model and the complexity of the data representation. A minimum value of BIC points out the best compromise, two dimensions (figure 10). Also, the correlation of the measured distances and the rebuilt distances is 90 % for a 2-dimensional space for the scene "kitchen" and 92 % for the scene "meeting". Consequently, the space is defined with two dimensions for both scene.

In order to take into account the inter-individual differences, an individual scaling (INDSCAL [4]) analysis is carried out on the raw data for each scene. Two 2-dimensional spaces are generated (figures 11 and 12).

In the 2-dimensional space of the scene "meeting", the systems appear ordered according to the ambisonic order on dimension 1, regardless of loudspeaker setup. A big difference appears between the first-order systems and the higher-order systems. Two groups are distinguished considering the loudspeaker setup along dimension 2. The systems decoded for the minimum number of loudspeaker are grouped as well as the ones using twelve loudspeakers. The large distance between the 1st order systems tends to shrink when the system order increases because a bigger number of lousspeaker is needed at higher order.
For the space of the scene "kitchen", axes are inverted. The first dimension shows the difference between loudspeaker configuration, and the second dimension is linked to the ambisonic order.

A procrustean transformation of the space of the "kitchen" on the space of the "meeting" is done (figure 13). A correlation of 98% between the two spaces is found.

4 DISCUSSION

Ambisonics and HOA systems are based on spherical harmonics. The more components are used, the more accurate the sound field is reproduced in a given area. In order to reproduce a sound field, microphones have been built, from first- to fourth-order.
In the first experiment, the performances of these devices to reproduce a sound field have been evaluated in terms of perceived accuracy of sound sources and spatial quality. Systems are ordered depending on the order but in three groups : the first-order microphone SoundField, the 12 sensors and the 8 sensors (second- and third-order microphones, respectively) and the fourth-order 32- sensor microphone.
The second experiment is carried out with the same system order but using synthetic encoding contrary to the first experiment. It is shown that there is an obvious difference between first order systems and HOA systems. On the other hand, the dissimilarities between HOA systems are smaller particularly for the systems decoded on 12-loudspeaker setup where differences are barely perceived.
Even though the question to the listener was different between the two tests, there are differences in the results concerning the HOA systems. Considering the same reproduction system (scene encoded from first to fourth order decoded over 12-loudspeaker setup), the "real" encoding systems of second and third order are not differenciated in both tests. However the third and fourth order systems are clearly differenciated in the first test but no dissimilarities are perceived between the two systems in the second experiment.
The first test focuses on spatial quality while the second experiment rates the global differences. If the systems used in the second experiment (synthetic encoding) have brought spatial degradation, this would have been seen on results. It is not the case. Then, the degradation that has been noticed

between the 12 sensors, 8 sensors (second and third order) and the 32 sensors microphone (fourth-order system) in the first experiment shows an influence of the recording device on sound field reproduction.

Furthermore, the first experiment reveals a significant difference between the synthetic fourth-order encoding system and the 32-sensors microphone, highlighting the encoding difference between a synthetic and a microphone system.

Focusing on play-back system configuration, from Gerzon criteria (energy and velocity vectors), the phase propagation and the energy concentration reproduced at the center of the loudspeaker setup are equivalent when using $2M+2$ loudspeakers or more (M is the ambisonic order). These objective factors do not take into account the physical limits of the ambisonic reproduction, in particular the sweet area diameter which is function of frequency. Solvang made a study on the frequency-dependent spectral degradation linked to the order and the reproduction area [22]. He distinguishes two situations : when the wave number kr (depending on the area diameter and the frequency) is smaller than the ambisonics order M and when the wave number is higher than the ambisonics order of the reproduction system. If $kr < M$, the additionnal loudspeakers do not influence the reproduced sound field. However when $kr > M$ the intensity error increases by using more than loudspeakers needed.
The second experiment corroborates the influence of the number of loudspeakers used for a given order. Two groups of systems are clearly defined in the perceptual space. Despite, there is less distance between the systems decoded over the minimum number of loudspeaker setup and the reference system than between the ones decoded on 12-loudspeaker setup and the reference system. Yet, the influence is noteworthy between the first-order system and the higher-order systems. Therefore, the increased number of loudspeakers for a given order seems to bring perceived sound field impairment. However, we cannot conclude on the perceptual attributes linked to this difference.

In these experiments, a decoding option has been chosen for all systems. We considered that since it is the same for all systems in both tests, the obtained results would not change if another option is used. However, the influence of these options should be investigated for futur developments.

5 CONCLUSION

A test is performed with "real" sound sources placed around the listener, virtually recorded by the microphones. Still concentrating on spatial quality, a comparison test shows the contribution of the higher orders and highlights the differences between synthetic and microphone encoding systems. The second test focuses on the reproduction setup, showing the influence of the number of loudspeakers used.

Consequently, the ambisonic restitution depends not only on the ambisonic order, but also on the recording and playback systems.

REFERENCES

[1] S. Bertet. *Formats audio 3D hiérarchiques: caractérisation objective et perceptive des systèmes Ambisoniques d'ordres supérieurs*. PhD thesis, Institut National des Sciences Appliquées de Lyon, March 2009.

[2] S. Bertet, J. Daniel, E. Parizet, L. Gros, and O. Warusfel. Investigation of the perceived spatial resolution of higher order ambisonic sound fields : a subjective evaluation involving virtual and real 3d microphones. In *AES 30th International Conference*, Saariselkä, Finland, 2007.

[3] I. Borg and P. J. F. Groenen. *Modern Multidimensional Scaling - Theory and Applications*. Springer Series in Statistics, New York, second edition, 2007.

[4] J. D. Carroll and J. Chang. Analysis of individual differences in multidimensional scaling via an n-way generalization of "eckart-young" decomposition. *Psychometrika*, 35(3):283–319, September 1970.

[5] 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.

[6] J. Daniel, J. B. Rault, and J. D. Polack. Ambisonics encoding of other audio formats for multiple listening conditions. In *Audio Eng. Soc. 105th Convention*, San Fransisco, California, September 1998.

[7] J. Daniel, J. B. Rault, and J. D. Polack. Acoustic properties and perceptive implications of stereophonic phenomena. In *Audio Eng. Soc. 16th Int. Conf*, 1999.

[8] G. Elko and J. Meyer. A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield. In *IEEE ICASSP - 02*, May 2002.

[9] K. Farrar. Soundfield microphone. *Wireless World*, 85(1526):48–50, October 1979.

[10] K. Farrar. Soundfield microphone - 2. *Wireless World*, 85:99–103, November 1979.

[11] M. A. Gerzon. Criteria for evaluating surround-sound systems. *Journal Audio Eng. Soc.*, 25(6):400–408, June 1977.

[12] M. A. Gerzon. General metatheory of auditory localisation. In *Audio Eng. Soc. 92nd Convention*, 1992.

[13] ITU-R. Rec. bs. 1534-1 method for the subjective assessment of intermediate quality level of coding systems. Technical report, 2003.

[14] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, March 1964.

[15] Ewan A. Macpherson. *A Review of Auditory Perceptual Theories and the Prospects for an Ecological Account*. PhD thesis, University of Winsconsin - Madison, USA, 1995.

[16] D. Malham. Experience with large area 3d ambisonic sound systems. *Proc. of the Institute of Acoustics*, 14, 1992.

[17] J. Meyer and T. Agnello. Spherical microphone array for spatial sound recording. In *Audio Eng. Soc. 115th Convention*, October 2003.

[18] 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.

[19] 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 *Audio Eng. Soc. 120th Convention*, Paris, France, May 2006.

[20] V. Pulkki and T. Hirvonen. Localization of virtual sources in multichannel audio reproduction. *IEEE Transactions on Speech and Audio Proc.*, 13(1):105–119, January 2005.

[21] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6(2):461–464, March 1978.

[22] A. Solvang. Spectral impairment for two-dimensional higher order ambisonics. *Journal Audio Eng. Soc.*, 56(4):267–279, April 2008.

[23] 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.