EVALUATION OF AMBISONICS DECODING METHODS WITH EXPERIMENTAL MEASUREMENTS

Diego M. Murillo[1], Filippo M. Fazi[2], Mincheol Shin[3]

[1]Institute of Sound and Vibration,
University of Southampton
Southampton, United Kingdom
dmmg1c12@soton.ac.uk

[2] Institute of Sound and Vibration,
University of Southampton
Southampton, United Kingdom
ff1@isvr.soton.ac.uk

[3] Institute of Sound and Vibration,
University of Southampton
Southampton, United Kingdom
M.Shin@soton.ac.uk

ABSTRACT

Ambisonics is a sound reproduction technique based on the decomposition of the sound field using spherical harmonics. The truncation in the number of coefficients used to recreate the sound field leads to reproduction artifacts which depend on the frequency and the listener spatial location. In this work, the performance of three different decoding methods (Basic, Max-rE and In-Phase) has been studied and evaluated in the light of the results of experimental measurements. The latter were performed using a spherical array composed of 40 uniformly distributed loudspeakers and a translating 29-channel linear microphone array. An error analysis is presented based on the difference between the desired and synthesized sound pressure and acoustic intensity field. The results indicate that, as expected, the size of the region of accurate sound field reconstruction reduces as frequency increases, but with different trends depending on the type of decoder implemented.

1. INTRODUCTION

3-D audio reproduction allows the generation of virtual spaces where the user perceives the sound according to the acoustic characteristics of the environment. This immersive experience has wide applications in areas such an entertainment, education, and research, among others. One methodology commonly used to reconstruct 3-D sound is the use of multichannel systems that reproduce the desired sound field over a specific area. Some advantages of implementing these techniques are a better immersive experience due to the use of multiple loudspeakers and the fact that the listening cues as Interaural Time, Level and Phase Differences are created in a natural way by the listener [1].

Ambisonics is a multichannel technique which has been extensively applied from the seventies [2-4]. It is based on the decomposition of the sound field using spherical harmonics which are part of the solution of the wave equation when it is expressed in spherical coordinates [5]. In theory, an exact reconstruction of the sound field is given when an infinite number of coefficients are computed. However, if this number is finite, the truncation will decrease the accuracy of the reproduction, depending on the frequency and the spatial location. The selection of the order of spherical harmonics is determined by the number of loudspeakers available for the reproduction of the sound field. This relation is commonly expressed by the following rule of thumb [6]:

$$L \geq (kr+1)^2 \qquad (1)$$

where L is the number of loudspeakers, k is the wavenumber and r is the radius of the area where the reconstruction is accurate (radius of validity or reference radius). Equation 1 implies that a high number of loudspeakers when the reproduction of high frequency sound is attempted over a wide area[1], generating a trade-off between these two variables.

Due to the artifacts created by the truncation in the number of spherical harmonics, different methods have been proposed to increase the physical or perceptual accuracy of the sound field reconstruction. For example, Max-rE decoder aims to maximize the energy vector optimizing the high frequency sound reproduction. The energy vector is defined as [1]:

$$rE \cdot \hat{u}_E = \frac{\sum_{n=0}^{N} G_n^2 \hat{u}_n}{\sum_{n=0}^{N} G_n^2} \qquad (2)$$

Where G_n is the gain of the nth loudspeaker and $\hat{u}_n$ is a unitary vector which represents the direction of an incoming wave radiated by the nth loudspeaker. A different approach is used in the In-Phase decoder, which recreates the condition that the loudspeakers feed the signals in phase decreasing the localization artifacts [1]. A detailed description of these decoding methods is beyong to the scope of this paper, but the reader can find a comprehensive discussion in [7].

The implementation of these types of decoder is made by applying a monotonically decreasing weighting function (like a "fade out") to the spherical harmonics coefficients. Consequently, each decoder yields a different sound field reconstruction performance. The concept of this weighting function can be explained using an analogy to the Fourier transform of a Dirac Delta function with different window types. Figure 1 shows delta signals created by applying several different frequency-domain windows. According to the window type selected, the energy of the coefficients is weighted in different proportion leading to an altered signal when the inverse Fourier transform is applied.

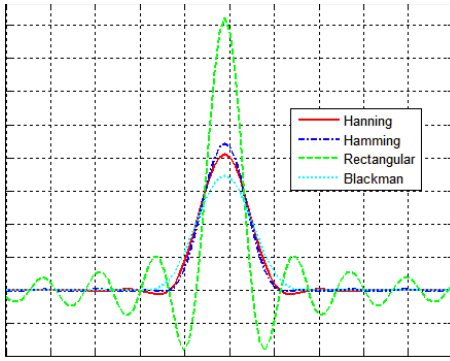[1] Radius of 0.1 m and frequency of 2 kHz require at least 25 loudspeakers

Figure 1: *A Delta Dirac signal after the application of the Fourier transform, truncation using different windows and subsequently inverse transform.*

Extensive work has been made to evaluate the performance of Ambisonics by means of perceptual or physical approaches. These assessments are commonly based on numerical simulations [4, 6, 8] or by listening test [9-11]. However, results obtained from experimental measurements of the acoustic pressure or the acoustic intensity field generated by HOA systems are less frequent in the scientific literature.

This paper evaluates the performance of three different Ambisonics decoding methods (Basic, Max-rE and In-Phase) by analyzing experimental measurements of objective parameters. To that end, a 5th order Ambisonics system was deployed using a spherical array of 40 loudspeakers [12]. The measurements were conducted in the anechoic chamber of the ISVR using a translating microphone array composed by 29 transducers (Ref. Brüel & Kjær 4189-L001) across 40 positions (see figure 2). The total number of measured points corresponded to 1160 with a spatial resolution of 0.05 m leading to an approximate spatial alias frequency of 3.4 kHz.

From the collected data, the sound pressure field and acoustic intensity field were computed and compared with the target field by means of pressure and intensity errors. The rest of this paper is organized as follows: Section 2 presents the methods used for the experiment. Section 3 shows the results of the measurements and the error analysis between the measured and target field. Finally, the conclusions of the current work are presented in Section 4.

Figure 2: *Measurement of the decoding methods.*

2. METHODS

The audio reproduction using an Ambisonics system can be mainly divided into two stages. Firstly, the audio signals are encoded in a finite number of spherical harmonic coefficients. This codification depends on the number of loudspeakers available but not on the size or shape of the array. A 5th order system involves the use of 36 spherical harmonics $(N + 1)^2$ to encode the signal. In the second stage, according to the number of the loudspeakers and the shape of the array, the signal is decoded and reproduced. One well established technique to decode the signal is called the mode-matching approach [13]. The reconstruction of a plane wave in direction (θ_i, ϕ_i) using a set of L plane waves each of them with different complex amplitude q_a and direction (θ_a, ϕ_a) can be expressed using the Jacobi-Anger expansion [14] as:

$$4\pi \sum_{n=0}^{\infty} (j^n) j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_i, \phi_i)^* =$$
$$4\pi \sum_{a=1}^{L} q_a(\omega) \sum_{n=0}^{\infty} (j^n) j_n(kr) \sum_{m=-n}^{n} Y_n^m(\theta, \phi) Y_n^m(\theta_a, \phi_a)^* \quad (3)$$

where k is the wavenumber, ω is the angular frequency, $j_n(kr)$ is the spherical Bessel function of first kind, $j = \sqrt{-1}$ and $Y_n^m(\theta, \phi)$ are the spherical harmonics defined as:

$$Y_n^m(\theta, \phi) = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} P_n^m \cos(\theta) e^{jm\phi} \quad (4)$$

in which P_n^m is the associated Legendre function. Simplifying equation 3 yields the following matching equation for each n and m:

$$Y_n^m(\theta_i, \phi_i)^* = \sum_{a=1}^{L} q_a(\omega) Y_n^m(\theta_a, \phi_a)^* \quad (5)$$

for $n = 0 ... N$ and $|m| \leq n$. This is a finite set of linear equations that can be written in a matrix form as $\boldsymbol{b}_{(m \times L)} = \boldsymbol{Y}_{(m \times L)} \boldsymbol{q}_{(m \times L)}$. In order to have at least one solution, the number of spherical harmonics $(N + 1)^2$ is required to be lower than, or equal to, the number of speakers, namely $L \geq (N + 1)^2$. Finally, the gains are calculated with the inverse matrix of $\boldsymbol{Y}_{((N+1)^2 \times L)}$ if $L = (N + 1)^2$ or pseudo-inverse matrix if $L > (N + 1)^2$. The stability of the inversion of the matrix Y depends on the loudspeaker array and can be checked by the condition number [15].

The algorithm to test the performance of the Ambisonics decoding methods was developed using the software package Max. Figure 3 shows a diagram of the decoder with its respective modules. The first part corresponds to the encoding stage using up to 5th order of spherical harmonics. Then, the resulting signals were weighted by a G_n function according to the chosen type of decoder (Basic, Max-rE or In-Phase). The values for the G_n functions were calculated using the methodology suggested by Jerome Daniel [7]. Table 1 reports the values of the gains G_n for each type of decoder, according to the order of the spherical harmonics. Finally, at the last stage, the signals are decoded

using the decoding matrix obtained by the mode matching approach and reproduced by the loudspeaker array.
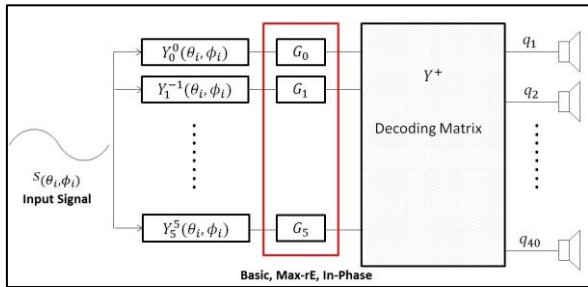

Figure 3: *Sketch of the decoder*

Table 1: *Decoder Gains.*

Order of SH	Basic	Max-rE	In-Phase
1	1	1	1
2	1	0.932	0.75
3	1	0.8029	0.4167
4	1	0.6259	0.1167
5	1	0.5186	0.0455

2.1. Sound pressure and acoustic intensity field

The sound pressure field was directly computed from the measurements. In the case of the acoustic intensity, the values were determined by taking the real part of the product between the sound pressure $p(\mathbf{x})$ and the conjugate of the particle velocity $\mathbf{u}(\mathbf{x})^*$ (see equation 6). The particle velocity was calculated based on the Euler equation (equation 7) by approximating the gradient of the pressure as the difference between neighbouring sound pressure measurement positions (equation 8).

$$\mathbf{I}(\mathbf{x}) = \frac{1}{2}\text{Re}\{p(\mathbf{x})\mathbf{u}(\mathbf{x})^*\} \qquad (6)$$

$$\mathbf{u}(\mathbf{x}) = -\frac{\nabla p(\mathbf{x})}{j\omega\rho_0} \qquad (7)$$

$$\mathbf{u}(\mathbf{x}) \approx -\frac{1}{j\omega\rho_0}\frac{[p(\mathbf{x}+d\mathbf{x})-p(\mathbf{x})]}{d\mathbf{x}} \qquad (8)$$

where $\nabla p(\mathbf{x})$ is the gradient of the pressure and ρ_o is the static density of the air.

3. RESULTS

The reconstruction of the acoustic pressure and acoustic intensity flow field for 250 Hz and 2 kHz are presented in Figures 4 and 5, respectively. Red color corresponds to zones of maximum acoustic pressure and blue to the minimum. The black circle represents the region of validity calculated from equation 1. In case of 250 Hz, the radius of validity is bigger than the dimension of the array so it is expected to have an accurate reconstruction over the whole measured area. Figures of 1 kHz are also presented in Appendix 1.

The measurement procedure involved the recording of the sound field generated by each type of decoder using the microphone array. The excitation signal corresponded to a virtual point source (white noise) located at 45° in azimuth [0°, 360°], 0° in elevation [90°, -90°] and 1.8 m far away. A comparison between figures 4 and 5 clearly identifies the limitation of Ambisonics to reproduce high frequencies. The radius of validity 'r' provides an insight on the area where the sound field reconstruction is accurate. However, it was found that this assumption is not always valid and depends strongly on the decoder. A more robust analysis of the data is performed in the next subsection.
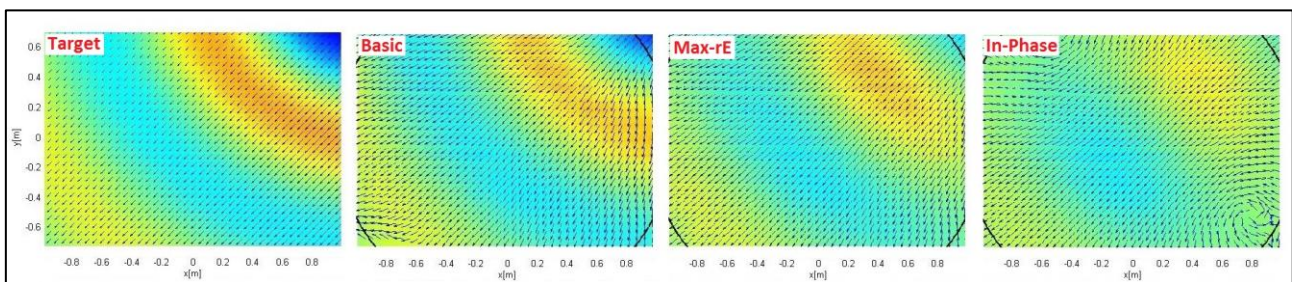

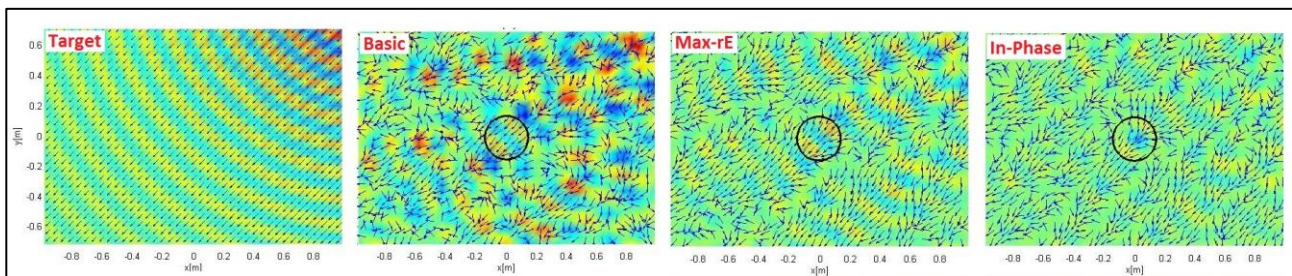Figure 4: *Sound pressure and acoustic intensity field flow reconstruction for 250 Hz.*

Figure 5: *Sound pressure and acoustic intensity flow field reconstruction for 2 kHz*

3.1. Error analysis

An error analysis was conducted on the sound pressure and the acoustic intensity data. The following error metrics have been adopted to assess the performance of the decoders:

Sound pressure errors:
Amplitude error:

$$E_{pa}(\mathbf{x}) = 10\log_{10}\left(\frac{|p_m(\mathbf{x})|}{|p_t(\mathbf{x})|}\right) \qquad (9)$$

Phase error:

$$E_{pp}(\mathbf{x}) = \text{angle}\left(\frac{p_t(\mathbf{x})\,p_m(\mathbf{x})^*}{|p_t(\mathbf{x})\|p_m(\mathbf{x})|}\right) \qquad (10)$$

where $p_m(\mathbf{x})$ is the measured pressure, $p_t(\mathbf{x})$ is the target pressure and $p_m(\mathbf{x})^*$ indicates the conjugate of the measured pressure.

Acoustic intensity error:
Angular error [16]:

$$E_{Ian} = \frac{\left(I_{x\_tar}I_{x\_mea}\right) + \left(I_{y\_tar}I_{y\_mea}\right)}{\left(\sqrt{\left(I_{x\_tar}\right)^2 + \left(I_{y\_tar}\right)^2}\right)\left(\sqrt{\left(I_{x\_mea}\right)^2 + \left(I_{y\_mea}\right)^2}\right)} \qquad (11)$$

in which I_{x_mea} and I_{y_mea} are the components of the measured acoustic intensity in $\hat{x}$ and $\hat{y}$ directions respectively. I_{x_tar} and I_{y_tar} are the components of the target acoustic intensity.

Figures 6 and 7 show the amplitude error of the sound pressure in dB. At 250 Hz, excellent agreement between the target field and the synthesized field is found for the Basic decoder. For the Max-rE and In-Phase decoders, the reconstructions are accurate at the center of the listener area, but over a region with a smaller radius than the predicted by the equation 1. At 2 kHz, the Basic decoder does not reconstruct the sound field as is expected, even within the radius of validity. The In-phase decoder presents a better performance compared to the Basic decoder, but the Max-rE decoder offers the best performance at this frequency.

The sound pressure phase error is illustrated in figures 8 and 9. The unit of the color bar corresponds to radians (from -π to π). At 250 Hz, the Basic decoder yields to the most accurate phase reconstruction. The Max-rE decoder also provides a good performance in terms of phase error except for the top left corner of the measured area, where a small mismatch can be observed. As in the case of the sound pressure amplitude, the In-Phase decoder leads to the largest errors at this frequency. At 2 kHz, the synthesized phase for all decoding methods tends to be more consistent with the measured data within the radius of validity. However, the Max-rE decoder achieves the best match for this case.

Regarding acoustic intensity, figures 10 and 11 show the angular error at 250 Hz and 2 kHz, respectively. The color bar represents the difference in degrees between 0° and 180°. At 250 Hz, the reference radius matches with the intensity flow created by the Basic and Max-rE decoders. This is not the case for the In-Phase decoder. Max-rE yields the best results for the intensity flow, but not in terms of the amplitude of the intensity where the Basic decoder is better. At 2 kHz, the angular error is almost zero inside of the reference radius for the Basic and Max-rE decoders. Nevertheless, using Max-rE, in some zones outside of this radius the intensity flow and the amplitude errors are comparatively small.
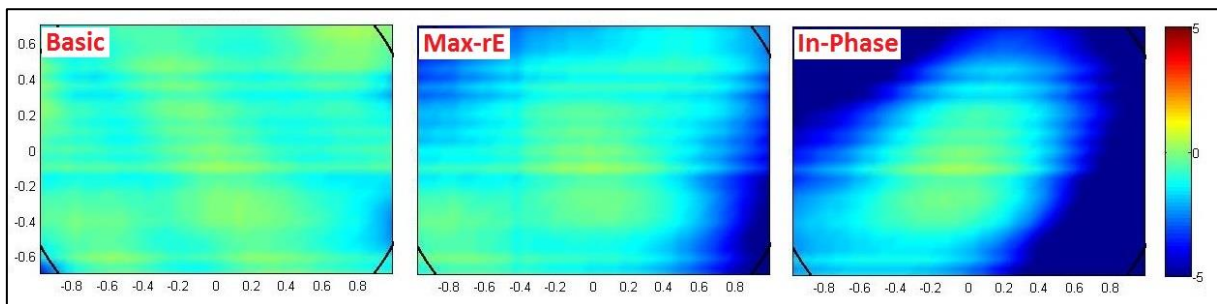

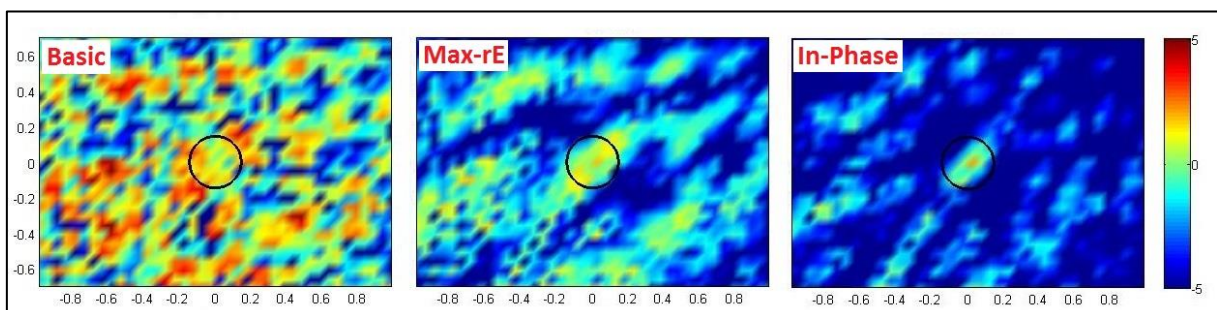Figure 6: *Sound pressure-amplitude error for 250 Hz*

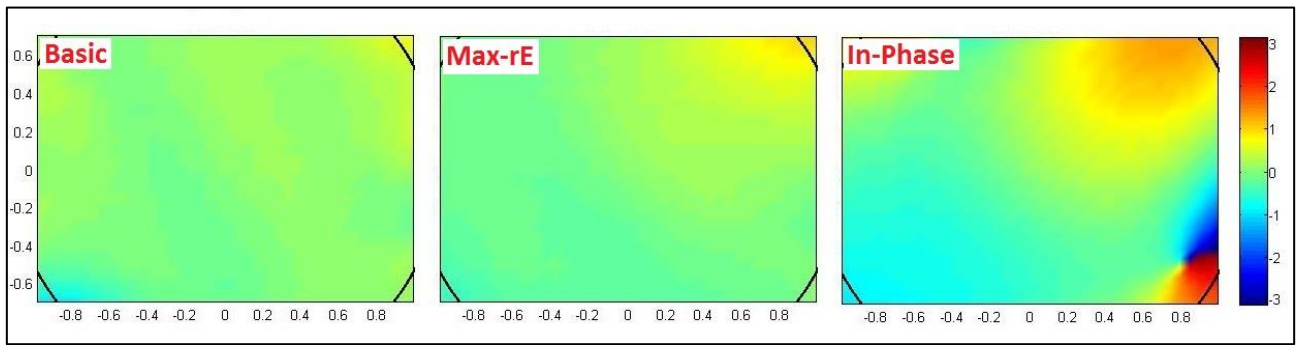Figure 7: *Sound pressure-amplitude error for 2 kHz*

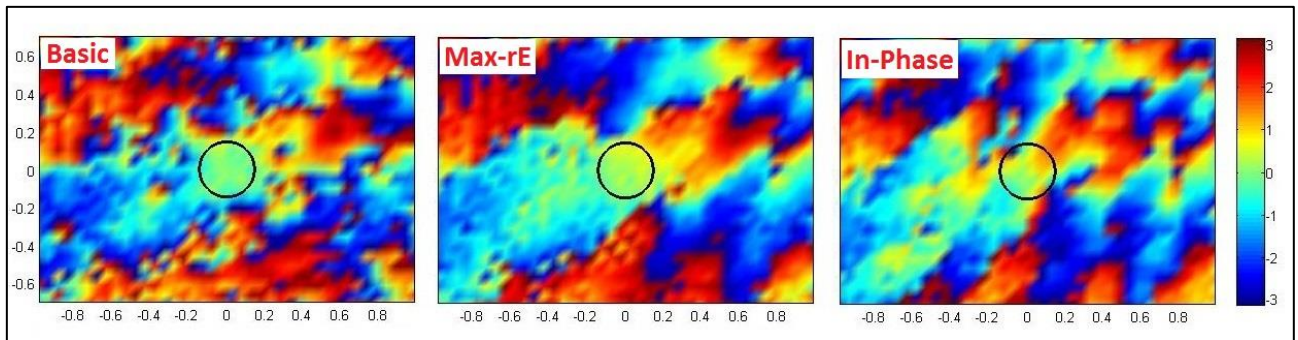Figure 8: *Sound pressure-phase error for 250 Hz*
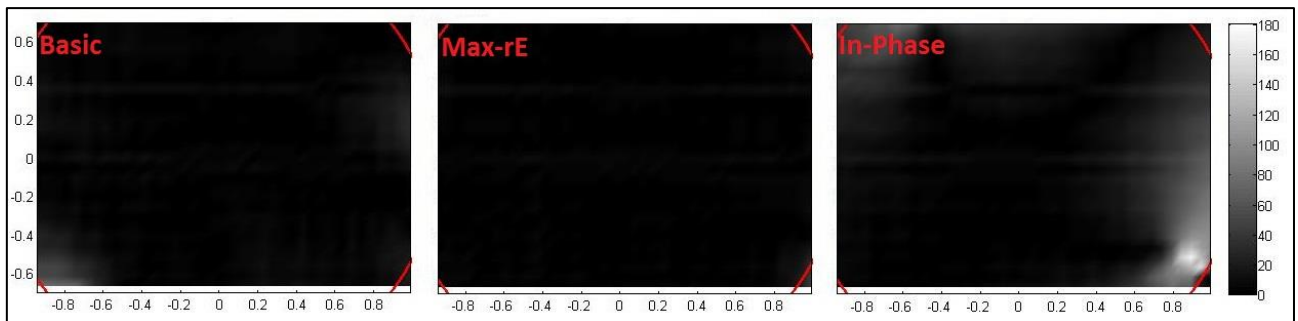

Figure 9: *Sound pressure-phase error for 2 kHz*

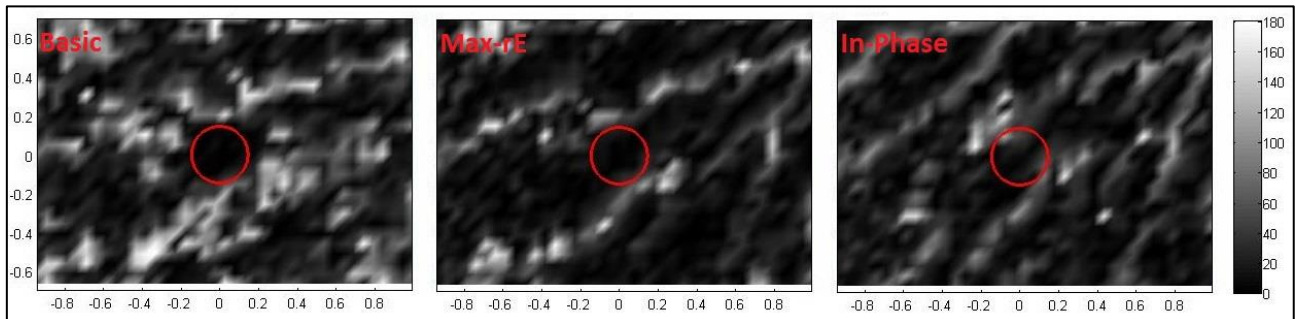Figure 10: *Acoustic intensity-angular error for 250 Hz*

Figure 11: *Acoustic intensity-angular error for 2 kHz*

3.2. Discussion of the results

From the error analysis, it is possible to identify that the performance of the decoders is highly dependant on frequency. At low frequencies, the basic decoder provides the best performance taking into consideration the sound pressure errors. Also, good agreement between the radius of validity and the area where the reconstruction is accurate has been described. However, the results of Max-rE and In-Phase decoders do not follow the rule of thumb $L \geq (kr+1)^2$ generating a smaller area where the pressure and intensity error are low. For this array, the In-Phase decoder is the worst method for reconstructing the low frequency sound field.

At high frequencies, the performance of the Basic decoder decreases noticeably compared to the other decoding methods. It shows the largest errors on both pressure and intensity amplitude[2] compared with the target fields. Nevertheless, the phase pressure and the angular intensity errors are low within the radius of validity. A comparison with the other decoding methods indicates that Max-rE offers the best results. This can be explained by the optimization of the energy vector which is the goal of this decoder. The errors on both pressure and intensity are the lowest when compared to other decoders.

Based on these results, if the aim is to reproduce an audio signal composed by a wide range of frequencies, the use of multiple decoding methods according to the frequencies may be advisable. To that end, the signal can be filtered and processed by different decoders based on the best performance in this specific frequency range. Examples of frequency dependent decoders can be found in [1,17].

4. CONCLUSIONS

The performance of three different Ambisonics decoding methods was evaluated in the light of experimental results. The findings confirm that the accuracy of sound field reproduction by a specific decoders depends on the frequency components. For this array, at low frequencies, the Basic decoder provides the best performance in terms of sound field reconstruction. In contrast, Max-rE presents the best performance at high frequencies. The implementation of combined Ambisonics decoding methods to reproduce a wide-frequency audio signal seems to be the most suitable option.

The concept of region of validity gives an indication of the area where the reconstruction is accurate. However, this assumption is not always valid in practice and significantly depends on the frequency and the type of decoder. The best match between the rule of thumb $L \geq (kr+1)^2$ and the reconstructed sound field was achieved, as expected, with the Basic decoder.

As the sound pressure, the acoustic intensity is another useful parameter that can be used to evaluate the performance of Ambisonics systems. Especially important is the angular error of the intensity, which cannot be evaluated for the acoustic pressure field as this does not contain directional information. An analysis in terms of pressure and intensity allows a more robust examination of reconstructed sound fields.

Finally, it is relevant to emphasize that the measurements were carried out in an anechoic environment using a spherical loudspeaker array which is far from the usual reproduction conditions. The performance of the decoders in regular rooms with comparatively low reverberation using a non-regular array is a topic for future research. Also, a near field compensation may be implemented in order to optimize the sound field for sources close to the listener.

5. REFERENCES

[1] D. Jerome, J. Rault and J. Polack "Ambisonics Encoding of Other Audio Formats for Multiple Listening Conditions*," in 105th Convention of Audio Engineering Society*, San Francisco, September 1998.

[2] M. Gerzon, "Periphony: With-Height Sound Reproduction*," J. Audio Engineering Society*, vol. 21, no. 1, pp.2-10, February 1973.

[3] S. Favrot and j. Buchholtz, "LoRA: A Loudspeaker-Based Room Auralization System," *Acta Acustica United with Acustica*, vol. 96, pp.364-375, 2010.

[4] D. Jerome, R. Nicol and S. Moreau, "Further Investigations of High Order Ambisonics and Wavefield Synthesis for Holophonic Sound Imaging*," in 114th Convention of Audio Engineering Society*, Amsterdam, March 2003.

[5] E. Williams. *Fourier Acoustics*. London: Academic Press, 1999.

[6] D. Ward and T. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transactions on Speech and Audio Processing*, vol. 9, pp. 697-707, Sep 2001.

[7] D. Jerome, Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context. Ph.D. dissertation, University of Paris, Paris, France, 2011.

[8] A. Solvang, "Spectral Impairment for Two-Dimensional Higher Order Ambisonics," *J. Audio Engineering Society*, vol. 56, no. 4, pp.267-279, April 2008.

[9] A. Horsburgh, R. Davis, M. Moffat and D. Fraser, "Subjective Assessments of Higher Order Ambisonic Sound Systems in Varying Acoustical Conditions*," in 133rd Convention of Audio Engineering Society*, San Francisco, October, 2012.

[10] G. Kearney, E. Bates, F. Boland and D. Furlong, "Comparative Study of the Performance of Spatialization Techniques for a Distributed Audience in a Concert Hall Environment, "

---

[2] No information about the error in the amplitude of the acoustic intensity was reported in this paper. However, it was calculated to analyse the performance of the decoding methods.

in 31ˢᵗ International Conference of Audio Engineering Society, London, June, 2007.

[11] S.Bernet, 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 30ˢᵗ International Conference of Audio Engineering Society*, Saariselka, March, 2007.

[12] F Fazi, Sound Field Reproduction. Ph.D. dissertation, University of Southampton, Southampton, United Kingdom, 2010.

[13] M. Poletti, "Three-Dimensional Surround Sound systems Based on Spherical Harmonics ," *J. Audio Engineering Society*, vol. 53, no. 11, pp.1004-1025, November 2005.

[14] B. Rafaely, "Plane Wave Decomposition of the Sound Field on a Sphere by Spherical Convolution", Institute of Sound and Vibration Research, ISVR, United Kingdom, Tech. Mem. 910, 2003.

[15] B. Stofringsdal and P.Svensson, "Conversion of Discretely Sampled Sound Field Data to Auralization Formats," *J. Audio Engineering Society*, vol. 54, no. 5, pp.380-400, May 2008.

[16] M. Shin, F. Fazi, P. Nelson and J. Seo "Control of Velocity for Sound Field Reproduction *," in 52nd InterCational conference of Audio Engineering Society*, Guildford, September, 2013.

[17] M. Gerzon, "Psychoacoustic Decoders for Multispeaker Stereo and Surround Sound," ," in 93rd Convention of Audio Engineering Society, San Francisco, October, 1992.
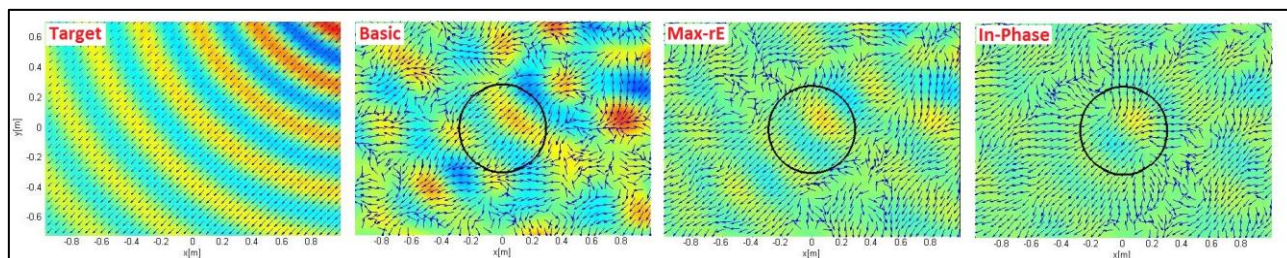
APPENDIX 1

Figure 12: *Sound pressure and acoustic intensity flow field reconstruction for 1 kHz.*
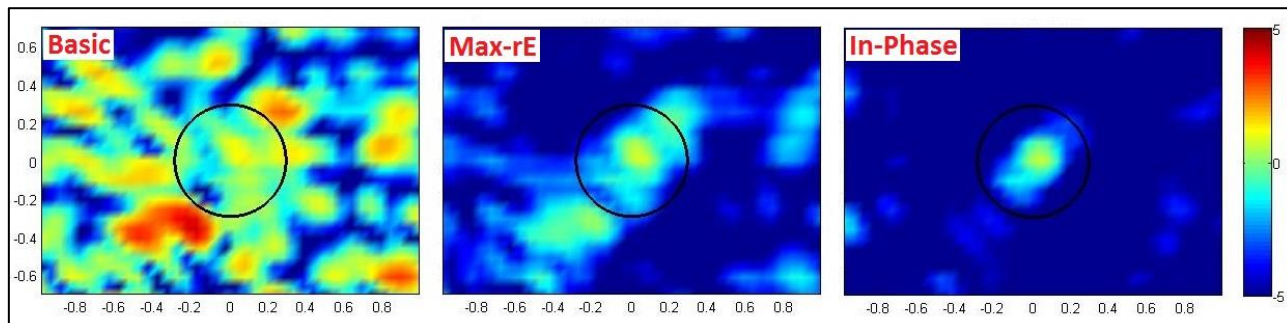

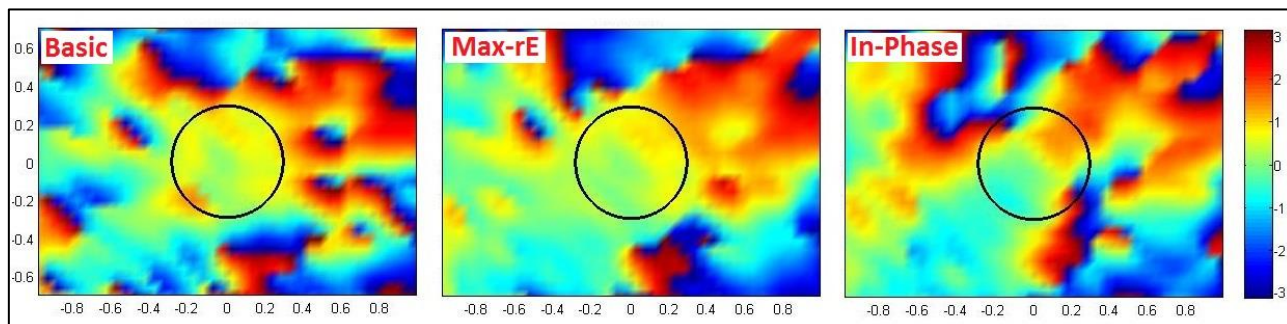Figure 13: *Sound pressure-amplitude error for 1 kHz*
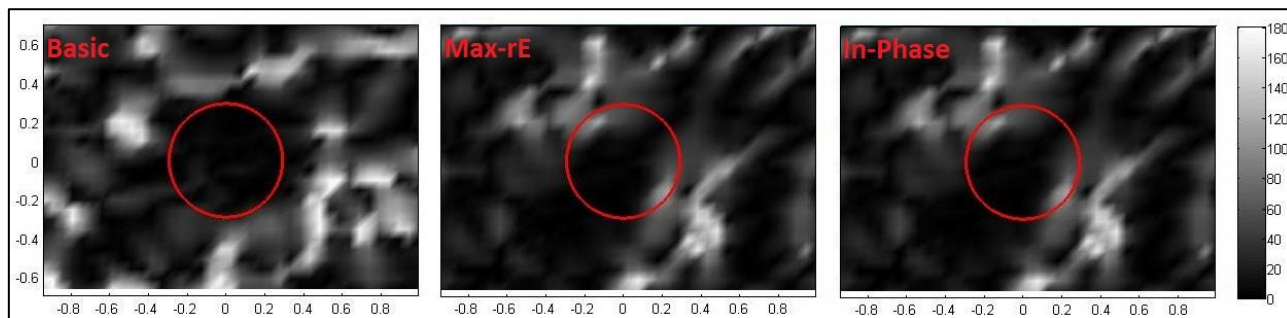

Figure 14: *Sound pressure-phase error for 1 kHz*

Figure 15: *Acoustic intensity-angular error for 1 kHz*