PERCEPTUAL EVALUATION OF A CIRCULARLY SYMMETRIC MICROPHONE ARRAY FOR PANORAMIC RECORDING OF AUDIO

Enzo De Sena, Hüseyin Hacıhabiboğlu and Zoran Cvetković

Centre for Digital Signal Processing Research (CDSPR)
King's College London
Strand, WC2R 2LS, London, United Kingdom
{e.desena|huseyin.hacihabiboglu|zoran.cvetkovic}@kcl.ac.uk

ABSTRACT

A circularly symmetric multichannel audio recording and reproduction system has previously been proposed by Johnston and Lam [1]. This system aimed at perceptual reconstruction of the sound field and employed a sparse circular array of directional microphones driving a matching array of loudspeakers. An improvement to that system based on stereophonic time-intensity panning concepts has recently been proposed [2]. In this paper localisation performance of that system is compared with the original Johnston/Lam configuration, and 2nd-order Ambisonics using a formal subjective localisation test. It is found that, for a listener positioned at the centre of the multichannel reproduction system, the newly proposed improvement to Johnston/Lam system delivers a more accurate spatial rendition of the sound sources as compared to the 2nd-order Ambisonics decoder we used.

1. INTRODUCTION

In many conventional multichannel systems, the desired spatial features are obtained through manual mixing and artificial manipulation of audio material. This approach requires high-end equipment, intervention of a sound engineer, and long production processes making them infeasible for scenarios such as live broadcast. Moreover, spatial cues like localization and envelopment are obtained through artificial panning and reverberation, thus impairing the consistency of the actual with the reproduced sound field. First-order Ambisonics [3, 4] provides an elegant solution to the problem. It aims at the physical reconstruction of the sound field, however the reproduction accuracy can be maintained only at a narrow optimal listening area. Higher-order Ambisonics (HOA) overcomes this problem by providing increased flexibility and enlarged listening area. HOA is therefore an ideal solution for custom-made setups, however it requires careful calibration, and recording and reproduction of audio material for HOA is not straightforward. Similarly wave-field synthesis (WFS) [5] accurately reproduces the wave front and has a wide listening area, but the number of input and output channels is too high for consumer grade reproduction systems.

In order to provide a wide listening area with a commercially feasible equipment setup, a sparse multichannel recording and reproduction system was proposed by Johnston and Lam [1]. This system relies to some degree on psychoacoustic phenomena as it aims to capture and reproduce binaural cues that a listener would experience in the actual sound field. Its design includes a spherical array of seven microphones driving a circular array of five loudspeakers. Such a system proved to be effective in rendering the localisation and envelopment cues in several real-world scenarios, however its design was mainly based on empirical observations. Subsequent studies [6, 2, 7] showed that a more methodological design approach can substantially improve its performance.

In this paper a comparison between the state-of-the-art Johnston/Lam system [2] and 2nd-order Ambisonics is presented on the basis of a subjective localisation study. The test involves matching acoustic pointers with the simulated source directions under well-controlled experimental conditions. It is found that the Johnston/Lam system achieves more accurate rendition of the direction of the auditory images. The original version proposed by Johnston in [8] is also included and it is observed that its performance is comparable with 2nd-order Ambisonics.

In Section 2 a brief description of the recording/reproduction systems under investigation is given. Section 3 presents the listening test setup and its results. Section 4 concludes the paper.

2. PANORAMIC RECORDING OF AUDIO

2.1. Johnston/Lam array

The recording system proposed by Johnston and Lam in [1] consists of five directional microphones evenly distributed on a circular array on the horizontal plane. Two superdirectional microphones aimed vertically up and down were added to improve ambience capturing. The reproduction system is composed of five loudspeakers equally spaced on a circle. Each horizontal microphone drives the matching loudspeaker, whereas the two remaining channels are suitably mixed and played back by all the loudspeakers.

In the original design, the microphone array diameter was 31cm, the horizontal microphones were hypercardioid[1] and the vertical ones were superdirectional. Further studies analysed the horizontal part of the system and explored the impact of different array radii and microphone directivity functions on the system performance. In [6] it has been suggested that higher order

The work reported in this paper was funded by the Engineering and Physical Sciences Research Council (EPSRC) Research Grant EP/F001142/1 "Perceptual Sound Field Reconstruction and Coherent Emulation".

[1] In a more recent specification [8], the same author suggests a directivity pattern with the primary lobe down by 3dB at 72° and down to effectively zero at 144°. This specification is employed in the present paper.

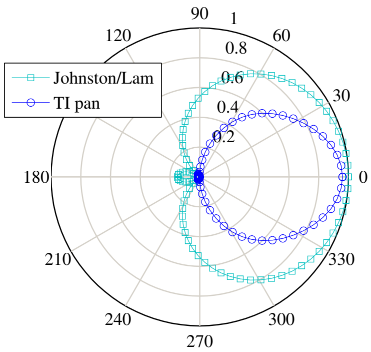

Figure 1: This figure shows the directivity pattern initially proposed in [8] (Johnston/Lam), and the improved directivity based on time-intensity panning (TI pan).

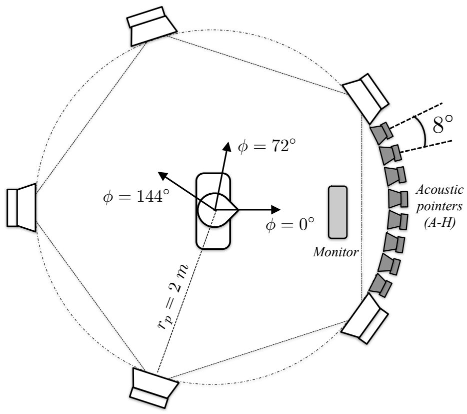

Figure 2: The test setup: the reproduction system is formed by the five white loudspeakers, whereas the grey loudspeakers are the acoustic pointers. Three listener seating directions, $\phi = 0°$, $\phi = 72°$ and $\phi = 144°$ are denoted as arrows.

microphones facilitate more accurate reproduction of monochromatic plane waves. More recently in [7] it was pointed out that natural sound sources trigger ILD and ITD cues that are highly correlated and that the phantom images reproduced by the surround system should ideally preserve this same property. On the basis of this observation, extensive simulations were run and their result showed that array diameters around 31 cm deliver more "natural" and mutually consistent ILD and ITD cues. Such a diameter is employed in [2] where a new approach to the microphone directivity design is established within the framework of time-intensity stereophony: if the inter-channel crosstalk is sufficiently small, the surround system behaves like a simple stereophonic system, and, under this assumption, stereophonic time-intensity panning curves as given by Franssen [9] can be used to design microphone directivity for a given array radius. In Figure 1 the resulting directivity is shown (TI pan) as well as the original directivity proposed in [8].

2.2. Second-order Ambisonics

The Johnston/Lam system is compared with a horizontal 2nd-order Ambisonics. The B-Format signals are encoded via the Furse-Malham 2nd-order coefficients (FMH-Format) [10] and decoded using the "in-phase" coefficients. The CDP Multi-Channel software toolkit available at [11] has been employed. More specifically the *abfpan2* routine has been used for encoding, while *fmdcode* for decoding. The employed loudspeaker layout is pentagon. This layout, which represents a sub-optimal setup for 2nd-order Ambisonics, has been chosen in order to compare three audio reproduction systems with the same number and topology of loudspeakers.

3. SUBJECTIVE EVALUATION

3.1. Subjects

Six subjects - five males and one female - with no reported hearing impairments attended the listening test. Three of them were the authors of the present paper and the other three were naive to the hypothesis under test. The presentation of the stimuli was fully randomised, thus ensuring the test was "blind" for the authors as well.

3.2. Test Setup

The test has been carried out in an audio booth with a reverberation time of $T_{60} \approx 200$ ms. The room dimensions were: $W = 4.5$ m, $L = 6$ m and $H = 2.2$ m. The walls and the ceiling were almost completely absorbent and the only major reflection was due to the floor.

The stimuli were played back by five MACKIE HR824 active monitor loudspeakers equally spaced on a circle of 4 meters of diameter. The acoustic pointers, eight GENELEC 6010, were placed between two consecutive channels of the five-channel system and separated by 8° as depicted in Figure 2. The 13 loudspeakers were driven by two sound-cards, a MOTU 828mk3 and a MOTU 896HD connected via Firewire to the audio workstation running Matlab 6.5. Loudspeaker levels were calibrated following the recommendation [12] to a nominal level of 78 ± 0.25 dBA .

All the loudspeakers were positioned at the ear level (1.2 meters). A computer monitor placed in front of the listener was connected to the audio workstation over which the test routine was executed.

3.3. Methodology

The listening test aimed to compare the localisation performance of 2nd-order Ambisonics with two versions of Johnston/Lam system: the original design [8] - denoted as "Johnston/Lam" - and the modified version recently proposed in [2] - denoted as "TI pan".

For each of the three systems, the microphone recordings were artificially synthesised for 8 different directions corresponding to the directions of the acoustic pointers - refer to Figure 2.

White Gaussian noise of 100 ms duration was used as a stimulus. A new random sample was generated at each trial in order to prevent bias due to fixed stimulus spectrum. A cosine tapered window with a 30% taper ratio was used to obtain a relatively

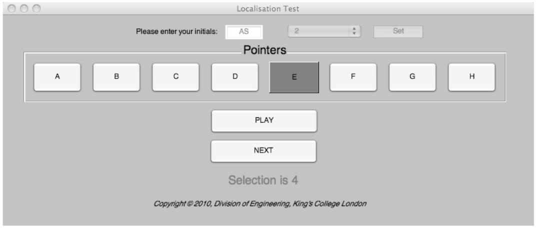

Figure 3: The GUI used to elicit subjects' responses.

smooth stimulus onset and offset in order to reduce transient response of the loudspeakers.

The subjects' task was to listen to the five-channel system stimuli - playing one of the three systems under test - and respond by listening to and selecting the acoustic pointer which is closest to the perceived direction of the auditory image. The subjects were given the possibility to play the stimuli and the acoustic pointers at will. The responses were collected using the MATLAB graphical user interface shown in Figure 3. The listeners were instructed to always watch the monitor - thus keeping the correct heading - but were allowed to make little head movements.

To test the localisation performances for a wider range of source directions, three different seating orientations, $\phi = 0^\circ, 72^\circ, 144^\circ$ were used - refer to Figure 2. For each seating orientation, each system-direction pair was presented 15 times. The presentation of the stimuli was fully randomised. In summary, each subject gave responses 15 times for 3 systems, 3 seatings, 8 directions, thus totalling $15 \cdot 3 \cdot 3 \cdot 8 = 1080$ trials. The test was divided in three blocks corresponding to the three seating orientations and a break was given between each block to prevent fatigue. Each block took between 30 and 40 minutes to complete for each subject.

3.4. Results

In Figures 4, 5 and 6 the average response angles are shown as a function of the stimulus angles. Ideally the response angles given by the listeners should be equal to the stimulus angles, therefore, the closer the points are to the bisecting line, the more accurate is the system.

Figure 4 shows the result for the first listening orientation, $\phi = 0^\circ$. The angles in abscissa and ordinate refer to the angles relative to listener heading. It can be observed that with 2nd-order Ambisonics, the average responses lie within an interval $\approx (-15^\circ, 15^\circ)$ suggesting that the auditory images are contracted in a sector between the two frontal loudspeakers. The *TI pan* system provides a more uniform panning as the stimulus directions that are closer to the edges are rendered properly. The original Johnston/Lam design performs slightly better than Ambisonics but worse than *TI pan*. It should also be noted that the results appear symmetrical with respect to front direction, which is due to the symmetry of the hearing system.

In Figure 5 is shown the result for the side listening direction, $\phi = 72^\circ$. It appears that in the interval between 44° and 68° the performance of all the systems are equally bad, being the listeners' choice the left-most acoustic pointer (44°) most of the time. This is possibly due to the sparsity of the surround system and the poor localisation accuracy of the auditory system

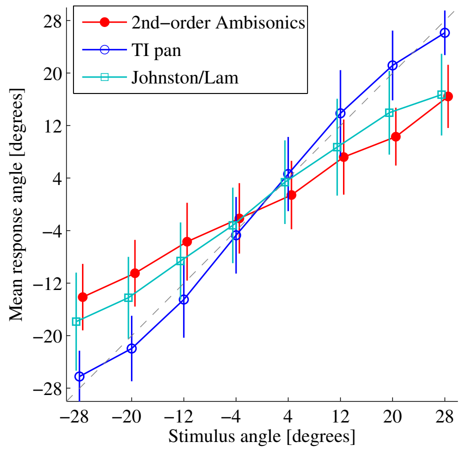

Figure 4: Mean response angle for the first listening position, $\phi = 0^\circ$ (front). The error bars show the $\pm\sigma$ intervals. Ideally the response angle should be equal to the stimulus angle (bisecting line).

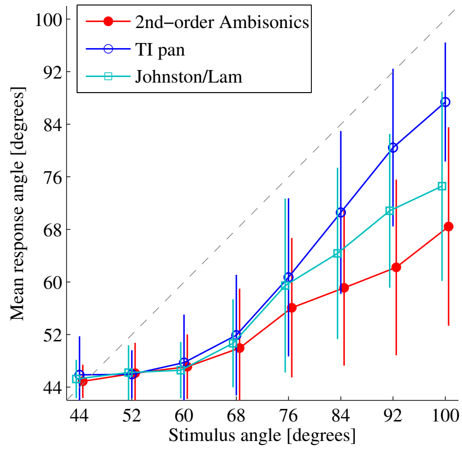

Figure 5: Mean response angle for the second listening position, $\phi = 72^\circ$ (side). The error bars show the $\pm\sigma$ intervals.

for side angles [13]. As the stimulus angle increases, the performance begin to separate for the different systems, with *TI pan* delivering again the best performance.

In Figure 6 is shown the result for the back listening direction, $\phi = 144^\circ$, where the same observations made for seating $\phi = 72^\circ$ can be repeated.

An overview of the results gathered among the three listening orientations is given in Table 1. The localisation error is defined as the difference between the stimulus angle and the response angle. In Table 1 its mean and standard deviation are given for the different systems. The mean errors are all positive and this is due to the listeners' tendency to choose left-most directions that has been commented for seating $\phi = 72^\circ$ and $\phi = 144^\circ$. Nevertheless, *TI pan* system shows the lowest bias,

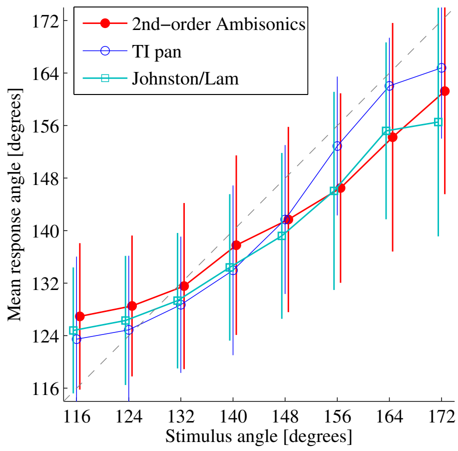

Figure 6: Mean response angle for the third listening position, $\phi = 144°$ (back). The error bars show the $\pm\sigma$ intervals.

followed by *Johnston/Lam* and 2nd-order Ambisonics. It should be emphasised that - assuming symmetry of the auditory system - the mean error would have been approximately zero if the remaining two listening directions $\phi = 216°$ and $\phi = 288°$ were considered in the test. However such directions are redundant in this analysis as they represent angles that are symmetric in relation to user heading, being $\phi = 216°$ equivalent to $\phi = 144°$ and $\phi = 288°$ equivalent to $\phi = 72°$.

System	Mean error	Std. deviation
TI pan	4.44°	10.80°
Johnston/Lam	6.64°	13.74°
2-order Ambisonics	6.78°	15.71°

Table 1: Means and standard deviations of the localisation error.

An important degradation in multichannel audio systems is due to the widening of the auditory image which reduces the locatedness of the auditory image [13]. We hypothesise that the spread of localisation error reflects the level of difficulty at which the subjects gave their responses. Therefore, pairwise comparisons of standard deviations of the subjective localisation errors were carried out via F-tests. In Table 1 it is shown that the SD of 2nd-order Ambisonics is higher than the SD of the original Johnston/Lam system, that, in turn, is higher than the SD of *TI pan*. Right-tailed tests were applied to check whether this result have statistical significance. The null hypothesis that the error variances are equal for 2nd-order Ambisonics and the original Johnston/Lam system can be rejected with an F ratio of $F(2159, 2159) = 1.3066, p < .0001$. The null hypothesis that the error variances are equal for *Johnston/Lam* and *TI pan* can also be rejected; the F ratio in this case is $F(2159, 2159) = 1.6191, p < .0001$. This points out that *TI pan* provides the better option in terms of locatedness of the auditory images.

4. CONCLUSIONS

In this paper a comparison between horizontal 2nd-order Ambisonics and two versions of Johnston/Lam surround system was presented. Their localisation capability has been studied by means of a subjective localisation experiment. The localisation test involved matching acoustic pointers with the simulated source directions under well-controlled experimental conditions. It was observed that the Ambisonics decoder we used renders phantom images that are contracted in between the two frontal loudspeakers. The newly proposed version of Johnston/Lam system [2] has much more uniform panning and delivers the best overall performance. The original Johnston/Lam version performance lie in between these two. An extension of the presented test paradigm including off-centre listening positions is planned.

5. REFERENCES

[1] J. Johnston and Y. Lam, “Perceptual soundfield reconstruction,” AES 109 Convention - Preprint # 5202, September 2000.

[2] H. Hacıhabiboğlu, E. De Sena, and Z. Cvetković, “Design of a circular microphone array for panoramic audio recording and reproduction: Microphone directivity,” Accepted for presentation at AES 128 Convention, London, UK, May 2010.

[3] M. Gerzon, “Ambisonics in multichannel broadcasting and video,” *J. Audio Eng. Soc.*, vol. 33, no. 11, pp. 859–871, November 1985.

[4] R. Furness, “Ambisonics–an overview,” in *Proc. of the 8th Int. Conference of the Audio Eng. Soc.*, 1990, pp. 181–189.

[5] M. Boone, E. Verheijen, and P. Van Tol, “Spatial soundfield reproduction by wave-field synthesis,” *J. Audio Eng. Soc.*, vol. 43, no. 12, pp. 1003–1012, December 1995.

[6] H. Hacıhabiboğlu and Z. Cvetković, “Panoramic recording and reproduction of multichannel audio using a circular microphone array,” in *Proc. 2009 IEEE Workshop on Appl. of Signal Process. to Audio and Acoust. (WASPAA'09)*, 2009, pp. 117–120.

[7] E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, “Design of a circular microphone array for panoramic audio recording and reproduction: Array radius,” Accepted for presentation at AES 128 Convention, London, UK, May 2010.

[8] J. Johnston and E. Wagner, “Microphone array for preserving sound field perceptual cues,” US Patent 6,845,163 B1, 2005.

[9] N. Franssen, *Stereophony*. Eindhoven, the Netherlands: Philips Research Laboratories, 1964.

[10] D. Malham, “Higher order ambisonic systems for the spatialisation of sound,” *Proceedings, ICMC99, Beijing*, pp. 484–487, 1999.

[11] [Online]. Available: http://people.bath.ac.uk/masrwd/mctools.html

[12] ITU-R, “Methods for the subjective assessment of small impairments in audio systems including multichannel audio systems,” Recom. BS.1116-1, 1997.

[13] J. Blauert, *Spatial hearing: The psychophysics of human sound localization*. The MIT Press, 1997.