PANORAMIC RECORDING AND REPRODUCTION OF MULTICHANNEL AUDIO USING A CIRCULAR MICROPHONE ARRAY

Hüseyin Hacıhabiboğlu and Zoran Cvetković

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

ABSTRACT

Multichannel audio reproduction generally suffers from one or both of the following problems: i) the recorded audio has to be artificially manipulated to provide the necessary spatial cues, which reduces the consistency of the reproduced sound field with the actual one, and ii) reproduction is not panoramic, which degrades realism when the listener is not seated in a desired ideal position facing the center channel. A recording method using a circularly symmetric array of differential microphones, and a reproduction method using a corresponding array of loudspeakers is presented in this paper. Design of microphone directivity patterns to achieve a panoramic auditory scene is discussed. Objective results in the form of active intensity diagrams are presented.

Index Terms— Multichannel audio reproduction, differential microphones, microphone arrays

1. INTRODUCTION

Multichannel audio systems have become ubiquitous with the advent of new and effective audio compression, multimedia storage, and delivery methods. Horizontal multichannel systems using 5.1 channels are now the standard multichannel audio reproduction method accompanying DVD and other forms of digital video delivery. Significant effort has gone into standardizing the reproduction setup and desirable listening conditions for multichannel audio material. However, the recording, mixing and production of the audio material is still carried out heuristically and many spatial effects such as localization and envelopment are achieved by artificial manipulation of audio material such as by panning [1] or by *enhancing* the recording with artificial reverberation [2]. This impairs the consistency of the reproduced sound with that of the actual, recorded sound field.

Ambisonics which aims to record and reproduce the sound field exactly [3], provides an elegant solution to these problems. However, the reproduction accuracy can be maintained only at a narrow optimal listening area. In addition, it requires careful calibration and is thus, not yet fully adopted for consumer grade audio reproduction. Similarly, wave-field synthesis (WFS) [4] provides physically accurate reproduction of a sound field. However, the number of input and output channels are prohibitively high for adoption of the technology in a domestic audio reproduction.

A recording method using a sparse spherical array of cardioid microphones was proposed to capture acoustical cues necessary for multichannel reproduction [5]. This method was based on the idea of conjoining several near-coincident stereo pairs to capture localization cues necessary for a panoramic reproduction in the horizontal plane. The proposed array consisted of a circularly symmetric array of cardioid microphones and two additional "shotgun" microphones at the apexes of a sphere to capture *ambience*.

Recently, the impact of the diameter of the microphone array was investigated and it was found that increasing the diameter improves the localization performance [6], but the subjective localization performance was still lower than in real-life. An improvement over this panoramic recording and reproduction method is proposed in this paper. More specifically, each microphone in the recording array is replaced with a higher-order differential microphone array that allows using different directivity patterns. Design of a microphone directivity function that emulates stereophonic panning laws and minimizes inter-channel crosstalk is presented.

Sec. 2 presents a theoretical analysis of the proposed recording and reproduction method based on active intensity. The interchannel crosstalk terms are identified. Sec. 3 discusses the design of microphone directivity functions to capture the sound field. An objective evaluation of the system is given in Sec. 4. Sec. 5 concludes the paper.

2. MULTICHANNEL RECORDING AND REPRODUCTION

Any sound field can be represented as a superposition of monochromatic plane waves with different amplitudes, frequencies, and directions. It is therefore possible to assess directional accuracy of multichannel audio systems by analyzing their behavior for single monochromatic plane waves. The analysis given in this section follows this approach.

The multichannel recording system investigated in this paper consists of an array of N directional microphones with the same directivity function, $\Gamma(\theta)$, positioned on a circle of radius r_a at equal intervals and their acoustical axes facing out (see Fig. 1). Let us consider a complex monochromatic plane wave of frequency f_0, incident from the horizontal direction θ on the k^{th} microphone. The recorded signal can be expressed as:

$$m_k(t) = A\Gamma_k\left(\theta\right) e^{jk_0\left[ct - r_a \cos\left(\theta - \frac{2\pi k}{N}\right)\right]}, \tag{1}$$

where A is the amplitude of the wave, $\Gamma_k(\theta) = \Gamma\left(\theta - 2\pi k/N\right)$ is the directive response of the microphone, c is the speed of sound, and $k_0 = 2\pi f_0/c$ is the wave number. It should be noted that complex monochromatic waves are used in this expression to simplify the exposition.

Reproduction setup consists of N equispaced loudspeakers on a circle at ear level (See Fig. 2). Each loudspeaker plays back the

This work is funded by EPSRC Research Grant, EP/F001142/1, "Perceptual Sound Field Reconstruction and Coherent Emulation".

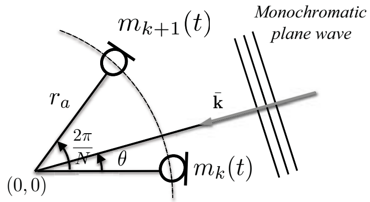


Figure 1: Two elements of the N-channel microphone array.

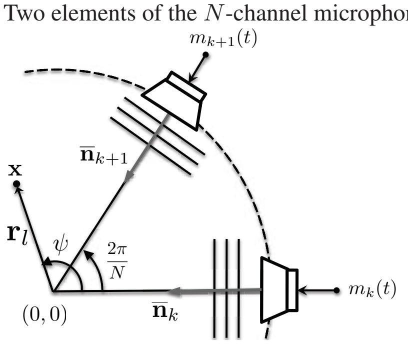


Figure 2: Two elements of the N-channel loudspeaker array.

audio signal recorded by the corresponding microphone without any additional processing. Let us assume that the loudspeakers are positioned in the acoustical far-field and can be approximated as plane-wave sources. We can express the pressure component of the sound field at an arbitrary position, $\mathbf{r}_l = r_l\,[\cos\psi \;\; \sin\psi]$, within the listening area due to the loudspeaker k as:

$$p_k(\mathbf{r}_l) = A\Gamma_k\left(\theta\right) e^{jk_0\left[ct - r_a\cos\left(\theta - \frac{2\pi k}{N}\right) - r_l\cos\left(\psi - \frac{2\pi k}{N}\right)\right]}. \quad (2)$$

The pressure and velocity components of the acoustical field in the listening area is a superposition of individual waves generated by these N loudspeakers:

$$p(\mathbf{r}_l) = \sum_{k=1}^{N} p_k(\mathbf{r}_l), \quad \overline{\mathbf{v}}(\mathbf{r}_l) = \frac{1}{\rho c}\sum_{k=1}^{N} p_k(\mathbf{r}_l)\overline{\mathbf{n}}_k. \quad (3)$$

where $\overline{\mathbf{n}}_k$ is the look direction of the loudspeaker k.

The directional properties of a sound field can be investigated using the real part of the complex intensity $\mathbf{I}_c(\mathbf{r}_l)$ also called *active intensity* [7]. Active intensity is co-directional with the propagation direction of a plane wave. As opposed to instantaneous intensity, complex intensity is not time-dependent for a monochromatic plane wave. The complex intensity can be expressed using the pressure and velocity components as:

$$\mathbf{I}_c(\mathbf{r}_l) = \frac{1}{2}p(\mathbf{r}_l)\overline{\mathbf{v}}^*(\mathbf{r}_l) = \frac{1}{2\rho c}\sum_{k=1}^{N}\sum_{m=1}^{N} p_k(\mathbf{r}_l)p_m^*(\mathbf{r}_l)\overline{\mathbf{n}}_k \quad (4)$$

The summand can be expressed as

$$\mathbf{I}_{c,mk}(\mathbf{r}_l) = A^2\gamma_{mk}(\theta)e^{j2k_0 d_{mk}\sin\xi_{mk}}\overline{\mathbf{n}}_k \quad (5)$$

where $\gamma_{mk}(\theta) = \Gamma_k(\theta)\Gamma_m(\theta)$ and,

$$\begin{aligned} d_{mk} &= \sin\left[\frac{(m-k)\pi}{N}\right]\sqrt{r_a^2 + r_l^2 + 2r_a r_l\cos(\psi - \theta)}, \\ \xi_{mk} &= \theta - \frac{(k+m)\pi}{N} + \tan^{-1}\left[\frac{r_l\sin(\psi-\theta)}{r_a + r_l\cos(\psi-\theta)}\right]. \end{aligned}$$

The active intensity is then given by:

$$\mathbf{I}_{a,mk}(\mathbf{r}_l) = A^2\gamma_{mk}(\theta)\cos\left(2k_0 d_{mk}\sin\xi_{mk}\right)\overline{\mathbf{n}}_k. \quad (6)$$

and, the total active intensity is:

$$\mathbf{I}_a(\mathbf{r}_l) = \frac{1}{2\rho c}\sum_{m=1}^{N}\sum_{k=1}^{N}\mathbf{I}_{a,mk}(\mathbf{r}_l) \quad (7)$$

In other words, the active intensity is related to not only the active intensities, $\mathbf{I}_{a,kk}(r_l)$, due to individual loudspeakers, but also the *cross-talk* active intensities $\mathbf{I}_{a,mk}(r_l),\ m \neq k$.

In order for the multichannel audio system to reproduce direction of the plane wave correctly $\mathbf{I}_a(\mathbf{r}_l)$ should be codirectional with the direction of propagation. The strength of the directionality of the sound field is determined by the magnitude of the active intensity. Therefore, in order to reproduce the plane wave correctly active intensity should have a large magnitude and also be codirectional with the propagation direction of the plane wave.

3. DIRECTIVITY FUNCTION DESIGN

The aim of the proposed multichannel system is to have at most two loudspeakers active for a *single plane wave*. For example, if the plane wave is incident from an angle, θ, such that $\frac{2\pi k}{N} \leq \theta \leq \frac{2\pi(k+1)}{N}$, only the loudspeakers k and $k+1$ should be active. This constraint allows using stereophonic panning laws for designing the common microphone directivity pattern . Two rules are employed for this purpose: i) cross-terms, $\gamma_{mk}(\theta)$ for non-consecutive microphones, m and k, should be minimized, and ii) directivity function should approximate stereophonic panning laws for directions of incidence between consecutive microphones.

Assuming a smooth directivity function, $\Gamma(\theta)$, the cross-talk terms can be minimized by designing the directivity function to have its zeros at $\theta = 2\pi k/N$ for $k \neq m$. In this manner, a plane wave incident from an angle between two consecutive microphones will be reproduced by the two corresponding loudspeakers only. The shape of the directivity function for $-2\pi/N \leq \theta \leq 2\pi/N$ can be designed based on the tangent panning law that is known to provide a good level of localization acuity in stereophonic reproduction [1]. This allows each plane wave forming the sound field to be panned *naturally* without any additional processing. Tangent panning law relates the gains of two loudspeakers to the target direction of the panned source and the angular separation between them as:

$$\frac{\tan\phi}{\tan\left(\phi_0/2\right)} = \frac{g_1 - g_2}{g_1 + g_2} \quad (8)$$

where $0 < \phi_0 < \pi$ is the separation between the loudspeakers, $-\phi_0/2 \leq \phi \leq \phi_0/2$ is the direction of the panned source defined from the midline of the two loudspeakers, and $0 \leq g_1, g_2 \leq 1$ are the amplitude gains of the loudspeakers. Additionally sound power can be normalized so that $g_1^2 + g_2^2 = 1$. These expressions can be simplified such that:

$$g_1 = \sqrt{\frac{T(\phi)}{1 + T(\phi)}}. \quad (9)$$

where

$$T(\phi) = \left[\frac{\tan\phi + \tan(\phi_0/2)}{\tan(\phi_0/2) - \tan\phi}\right]^2. \quad (10)$$

For the proposed system with N elements, the angular separation between consecutive microphones/loudspeakers is $\phi_0 = 2\pi/N$, and the amplitude panning gain factors are $g_1 = \Gamma(\pi/N - \phi)$ and $g_2 = \Gamma(\pi/N + \phi)$. The directivity function can then be expressed using (9) as:

$$\Gamma(\theta) = \sqrt{\frac{T(\pi/N - \theta)}{1 + T(\pi/N - \theta)}}. \tag{11}$$

A directional microphone with a prescribed directivity pattern can be realized using a differential microphone array consisting of a number of omnidirectional microphone elements. The directivity function of an M^{th}-order differential array can be expressed as:

$$\Gamma(\theta) = \sum_{m=0}^{M} a_m \left[\cos(\theta)\right]^m . \tag{12}$$

Therefore, the design process involves obtaining coefficients, a_m, that determine the inter-element delays that should be used [8].

For the considered system, each microphone in the recording array is replaced with a differential microphone array emulating stereophonic panning. In other words, the recording array is realized as a *circular array of differential microphones*. Each of these differential microphones emulate stereophonic panning laws. In order to obtain the microphone directivity that emulates the tangent panning function for the given azimuth range, and minimizes the cross-talk between non-consecutive channels, the coefficients, a_m, should be calculated by evaluating (11) at P discrete angles $0 \leq \theta_p \leq 2\pi/N$ and setting the nulls of the directivity function at $\theta = 2n\pi/N$, $n \neq k$. For odd number of channels another null at $\theta = \pi$ should also be imposed in order to reduce cross-talk further. The resulting set of linear equations can be expressed in matrix form as:

$$\mathbf{G} = \mathbf{C}\mathbf{a}, \tag{13}$$

where

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & \cos\theta_1 & \dots & \cos^M \theta_1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\theta_P & \dots & \cos^M \theta_P \\ 1 & \cos\frac{2\pi}{N} & \dots & \cos^M \frac{2\pi}{N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \cos\frac{2\pi(N-1)}{N} & \dots & \cos^M \frac{2\pi(N-1)}{N} \\ 1 & \cos\pi & \dots & \cos^M \pi \end{bmatrix},$$

$$\mathbf{G} = \left[1\ \Gamma(\theta_1)\ \dots\ \Gamma(\theta_p)\ 0\ \dots 0\right]^T, \tag{14}$$

$$\mathbf{a} = \left[a_0\ a_1\ \dots\ a_M\right]^T .$$

This constitutes an overdetermined system of linear equations.

The least squares solution for the gain factors is given by

$$\mathbf{a} = \mathbf{C}^+\mathbf{G} \tag{15}$$

where $\mathbf{C}^+ = \left(\mathbf{C}^T\mathbf{C}\right)^{-1}\mathbf{C}^T$ is the pseudoinverse of $\mathbf{C}$.

Fig. 3 shows the directivity functions of the microphones for a system with $N = 5$ channels. The directivity functions were obtained using the panning function calculated at $P = 10$ points for $M = 5$. The coefficients of the plotted directivity functions are $a_0 = -0.0402$, $a_1 = -0.0697$, $a_2 = 0.6771$, $a_3 = 1.2247$, $a_4 = -0.1314$ and $a_5 = -0.6622$.

Note that the proposed directivity function does not cause amplitude panning of a sound source between two loudspeakers. On the contrary, each plane wave constituting a complex sound field will be reproduced by two loudspeakers, and all loudspeakers will be active during run-time.

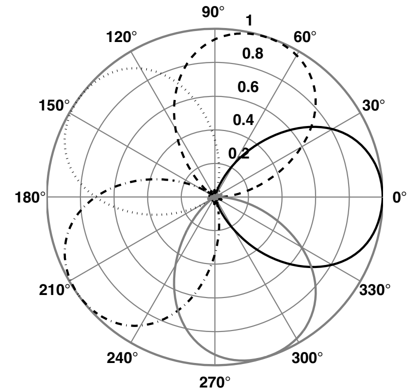


Figure 3: Directivity patterns of the differential microphones in a 5-channel recording array.

4. EVALUATION OF THE SYSTEM

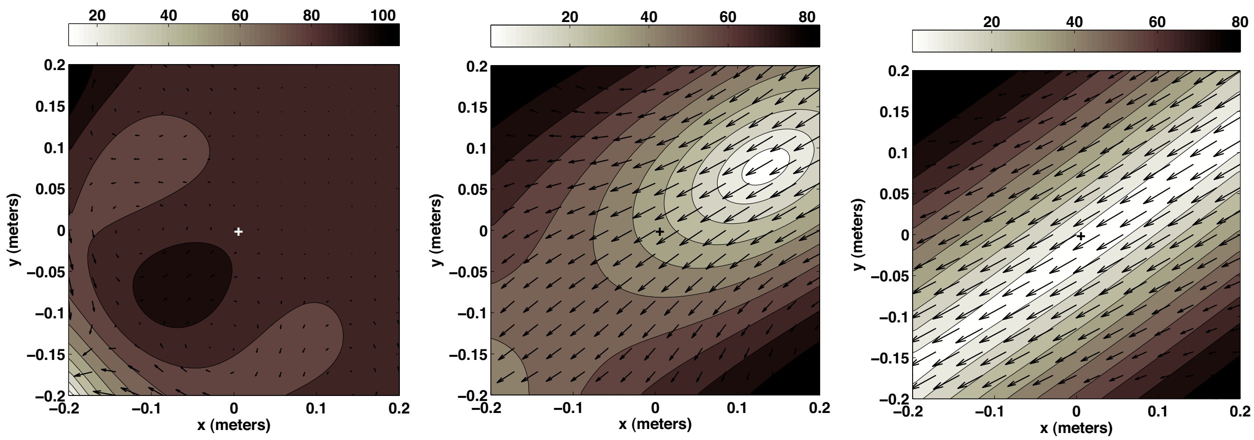
The directional accuracy of the multichannel system for a single monochromatic plane wave in the listening area can be used to objectively evaluate the performance of the proposed system [9]. While this kind of analysis considers a single plane wave, it gives a good indication of the behavior of the system for come complex sound fields consisting of a linear combination of many plane waves. For this purpose, let us assume that the loudspeakers are positioned in the acoustical far-field and behave as plane wave sources as described in Sec. 2. The actual active intensity due to a plane wave, $\mathbf{I}_{a,pw}(\mathbf{x})$, and the active intensity for the proposed system, $\mathbf{I}_{a,mc}(\mathbf{x})$ can be numerically calculated at discrete points. Let us define the directional error as:

$$\epsilon(\mathbf{x}) = \arccos\left(\frac{\mathbf{I}_{a,pw}(\mathbf{x}) \cdot \mathbf{I}_{a,mc}(\mathbf{x})}{|\mathbf{I}_{a,pw}(\mathbf{x})||\mathbf{I}_{a,mc}(\mathbf{x})|}\right) \tag{16}$$

which corresponds to the angle between the actual and reproduced active intensities. For ideal reproduction $\epsilon(\mathbf{x}) = 0$.

The proposed system was evaluated for $N = 5$ channels, and for a microphone array radius of $r_a = 15.5$ cm. Three directivity functions for the individual elements of the recording array were evaluated: omnidirectional, ideal cardioid, and the described differential microphone directivity. The directivity pattern of the differential microphone was designed as proposed in the previous section. Figure 4 shows the error and the active intensity vectors for a square region of 0.2 m size around the sweet spot. The simulation was made for a plane wave incident from $\theta = 30°$ with a frequency of $f_0 = 500$ Hz. The contour plots show the error, $\epsilon(\mathbf{x})$ at each point around the central listening position according to (16). Lighter colours correspond to smaller directional errors. The vector plots show the direction and the magnitude of the active intensity at the respective points.

It may be observed that the omnidirectional microphone array fails to provide any consistent and useful directional reproduction. In addition, the error can be higher than $\pi/2$. The results for the cardioid microphone array are much better when compared with



(a) (b) (c)

Figure 4: The directional error $\epsilon(\mathbf{x})$, and the active intensity vectors for a multichannel system with $N = 5$ channels, for a plane wave with $f_0 = 500$ Hz incident from $\phi = 30^\circ$ using (a) omnidirectional, (b) cardioid, and (c) fourth-order differential microphones. The cross sign denotes the centre of the listening area. Note the difference between colormap scales.

the omnidirectional array. This confirms the original assumption in [5] that using directional microphones is essential in the investigated multichannel audio system. However, the area where the reproduction is more accurate is shifted from the central position. It may also be observed that slight shifts from the central listening position will cause the directional properties of the reproduced sound field to deteriorate significantly. The proposed directivity function provides a good and consistent directional reproduction in a wider listening area. In addition, the error is distributed more homogenously. The mean directional errors are, 6.25°, 15.30°, and 83.52° for the fifth-order differential, cardioid, and omnidirectional arrays, respectively.

The proposed system was also evaluated via informal listening sessions. Directional room impulse responses for four source positions around a five channel microphone array were simulated in a digital waveguide mesh (DWM) room model as described in [10]. Anechoic recordings of an ensemble of a guitar, a flute, a bass, and percussion were convolved with these and played back in an acoustical booth over a 5-channel system. The initial results show that it is possible to obtain excellent subjective localization accuracy, and a high level of realism with the proposed method.

5. CONCLUSIONS

A multichannel recording and reproduction system for achieving a panoramic auditory reality was studied in this paper. The system is based on the recording of a sound field by multiple directional microphones arranged on a circle with equal spacing as originally proposed by Johnston and Lam [5]. The reproduction is made by a loudspeaker array with the same number of loudspeakers. The design of microphone directivity that emulates stereophonic tangent panning law was presented. Objective evaluation based on the concept of active intensity was given. It was shown that the proposed method provides good directional reproduction for a wide region. Informal listening tests also indicated that the proposed system provides excellent localization and a high level of realism. A formal subjective evaluation is planned.

6. REFERENCES

[1] V. Pulkki, "Virtual sound source positioning using vector-base amplitude panning," *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, June 1997.

[2] W. G. Gardner, "Reverberation algorithms," in *Applications of digital signal processing to audio and acoustics*, M. Kahrs and K. Brandenburg, Eds. Kluwer Academic Publishers, 2002, pp. 85–131.

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

[4] M. M. Boone, U. Horbach, and W. P. J. Bruijn, "Spatial sound-field reproduction by wave-field synthesis," *J. Audio Eng. Soc.*, vol. 43, no. 12, pp. 1003–1012, December 1995.

[5] J. D. Johnston and Y. H. Lam, "Perceptual soundfield reconstruction," Presented at the AES 109th Convention, Los Angeles, USA, Preprint #2399, 22-25 September 2000.

[6] J. Hall and Z. Cvetković, "Coherent multichannel emulation of acoustic spaces," in *Proc. AES* 28^{th} *Conference, Piteå, Sweden*, 30 June - 2 July 2006.

[7] R. D. Heyser, "Instantaneous intensity," in *Presented at the AES 81st Convention, Los Angeles, USA, Preprint #2399*, 12-16 November 1986.

[8] G. W. Elko, "Differential microphone arrays," in *Audio signal processing for next-generation multimedia communication systems*, Y. Huang and J. Benesty, Eds. Boston, USA: Kluwer Academic Publishers, 2004.

[9] M. A. Poletti, "A unified theory of horizontal holographic sound systems," *J. Audio Eng. Soc.*, vol. 48, no. 12, pp. 1155–1182, December 2000.

[10] H. Hacıhabiboğlu, B. Günel, and Z. Cvetković, "Simulation of directional microphones in 3-D digital waveguide mesh-based models of room acoustics," *IEEE Trans. on Audio, Speech and Language Process.*, 2009, (in press).