

Recording and Reproducing High Order Surround Auditory Scenes for Mixed and Augmented Reality

Zhiyun Li, Ramani Duraiswami, Larry S. Davis
Perceptual Interfaces and Reality Laboratory, UMIACS
University of Maryland, College Park, MD 20742
zli@cs.umd.edu; {ramani; lsd}@umiacs.umd.edu

Abstract

Virtual reality systems are largely based on computer graphics and vision technologies. However, sound also plays an important role in human's interaction with the surrounding environment, especially for the visually impaired people. In this paper, we will develop the theory of recording and reproducing real-world surround auditory scenes in high orders using specially designed microphone and loudspeaker arrays. It is complementary to vision-based technologies in creating mixed and augmented realities. Design examples and simulations are presented.

1 Introduction

Virtual reality systems are finding increasing applications in design, simulation, remote control, teleconference and training. Since the human sight accounts for more than 90 percent of all perceived information, most such systems are based on computer graphics and vision technologies. Examples include head-mounted displays, CAVE [3] and tiled displays [9][4][7][6], etc. In addition, to augment human-computer interactions, they are usually equipped with joysticks, ultrasonic pointers, data gloves, and hand-held wands.

Other than that, sound also provides rich information about the surrounding world. With the advance in microphone, loudspeaker and digital signal processing technologies, it is possible and desirable to build a sound-augmented immersive reality system which records and reproduces real-world surround auditory scenes accurately.

1.1 Previous Work

While there exist algorithms to create virtual sound sources using loudspeaker array such as in [11], our concern is to recreate the realistic auditory scenes from real-world

recordings. Several schemes have been proposed to build the microphone-loudspeaker array system. In [1], based on the Kirchhoff-Helmholtz integral on a plane, the sound field is captured by a directive microphone array, and recreated by a loudspeaker array. That is called the *Wave Field Synthesis* (WFS) method. While the system in [1] works well in an auditorium environment where the listening area can be separated from the primary source area by a plane, it is hard to render an immersive perception of a surround sound field. In [10], a general framework was proposed, which uses linear microphone arrays to identify, localize and track the sound sources, then uses the loudspeaker array to recreate them with the correct spatial cues via WFS. To work properly for complex auditory scenes with multiple moving sound sources, however, it requires a robust and accurate localizing and tracking system and a highly directive beamformer which are usually very expensive, if available. In [5], we have developed a unified and simple approach for capturing and recreating high order 3D auditory scenes using the reciprocity principle that is satisfied between the two processes. It uses a spherical microphone array mounted on a rigid sphere to capture 3D auditory scenes and a spherical loudspeaker array in free space to recreate the recorded scenes. This approach is designed to be independent of sound sources.

1.2 Present Work

However, the capture and recreation of full 3D auditory scenes using spherical microphone-loudspeaker arrays are apparently redundant and uneconomic for surround auditory scenes which are usually the cases in real life. In such cases, sound sources are roughly on a plane such as in a cocktail party, roundtable conference, surround music recording, auditory traffic scenes, etc. In this paper, we will parallel the design methodology in [5] and develop a recording and reproducing system for high order surround auditory scenes. We first design a circular microphone array mounted on a sound-rigid cylinder, with its axis perpendic-

ular to the sound source plane when recording. Using our cylindrical beamforming algorithm, this array will be used for recording and spatial filtering high order surround auditory scenes. Then we use a specially arranged circular loudspeaker array to reproduce the recorded surround auditory scenes. The main advantages of our system are:

1. It is easier to build and more practical to use for surround auditory scenes, especially compared with the spherical arrays;
2. It is highly efficient and effective for recording and reproducing surround auditory scenes;
3. Its performance is robust and optimal;
4. It is highly scalable.

1.3 Applications in Mixed and Augmented Reality

The applications of our system in mixed and augmented reality are bestly illustrated by a teleconference setup, such as “the office of the future” described in [9]. In that setup, distant and virtual environments are integrated seamlessly with the real office to create an immersive platform for distant individuals to collaborate. This platform provides a visually immersive effect using distributed cameras and projectors, however, it ignores the immersive auditory information. Our system is analogous to this platform in the sense we use microphone and loudspeaker arrays to integrate distant surround auditory scenes into the real and virtual auditory environment in the real room. On the other hand, our system is also complementary to this platform since we can combine the immersive visual and auditory effects together to create an even more effective one.

2 Recording Surround Auditory Scenes

In this section, we will first introduce the theory of acoustic scattering from a rigid cylinder. We then develop the theory of cylindrical beamforming in section 2.2. In addition, we explain why we need a cylinder instead of using the microphone array in free space. To demonstrate our design, we analyze the performance of a practical example in section 2.3. We address the practical design issues such as discreteness, practical beampattern, robustness, and its performance in 3D sound field. Both theoretical analyses and simulation results are presented.

2.1 Wave Scattering from Rigid Cylinder

Considering a general scenario as shown on the left side in Figure 1, where a unit magnitude plane sound wave of wave number \mathbf{k} is incident from (θ_k, φ_k) , an infinite

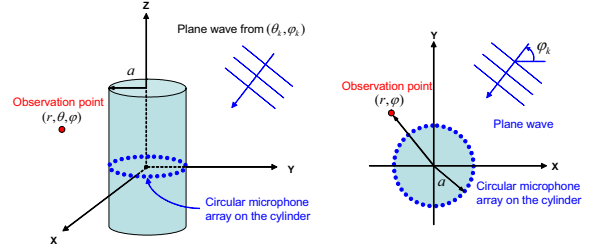


Figure 1. A plane wave incident on a rigid cylinder.

length rigid cylinder of radius a is positioned vertically at the origin of the coordinate system, and the continuous circular microphone array is on the cross-section circle at $z = 0$ plane. The complex pressure at the observation point $\mathbf{r} = (r, \pi/2, \varphi)$ on the $z = 0$ plane is:

$$p_i = e^{i\mathbf{k}\cdot\mathbf{r}}. \quad (1)$$

If \mathbf{k} is perpendicular to the axis of the cylinder (z -axis), i.e. $\theta_k = \pi/2$, we then have:

$$p_i = e^{ikr \cos(\varphi - \varphi_k)}, \quad (2)$$

where $k = |\mathbf{k}|$. Now it is reduced to a 2D problem as shown on the right side in Figure 1. We will come back to the 3D issues in section 2.3.3.

This plane wave can be expressed in terms of cylindrical waves [8]:

$$p_i = \sum_{n=0}^{\infty} \epsilon_n i^n J_n(kr) \cos[n(\varphi - \varphi_k)], \quad (3)$$

where $J_n(x)$ is the n th-order Bessel function of x , and ϵ_n are Neumann symbols defined as:

$$\epsilon_n = \begin{cases} 1, & n = 0; \\ 2, & n \geq 1; \end{cases} \quad (4)$$

The scattered wave is:

$$p_s = - \sum_{n=0}^{\infty} \epsilon_n i^n \frac{J'_n(ka) H_n(kr)}{H'_n(ka)} \cos[n(\varphi - \varphi_k)], \quad (5)$$

where $H_n(x)$ is the n th-order Hankel function of x . The prime denotes the derivative with respect to the argument.

The total complex pressure at the circle on the cylinder surface is:

$$\begin{aligned} p_t(\varphi, \varphi_k) &= (p_i + p_s)|_{r=a} \\ &= \sum_{n=0}^{\infty} B_n(ka) \cos[n(\varphi - \varphi_k)], \end{aligned} \quad (6)$$

where:

$$B_n(ka) = \epsilon_n i^n \left[J_n(ka) - \frac{J'_n(ka)H_n(ka)}{H'_n(ka)} \right]. \quad (7)$$

2.2 Cylindrical Beamforming

The basic principle of cylindrical beamforming is to make orthogonal decompositions of the recorded surround sound field, and the decomposed components are combined to approximate a desired beampattern.

If we assign the weight $\frac{\cos(n'\varphi)}{\pi B_{n'}(ka)}$ to each continuous point on the circle and make an integral over φ , we have:

$$\begin{aligned} & \int_0^{2\pi} p_t \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi \\ &= \int_0^{2\pi} \sum_{n=0}^{\infty} B_n(ka) \cos[n(\varphi - \varphi_k)] \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi. \end{aligned} \quad (8)$$

Since

$$\cos[n(\varphi - \varphi_k)] = \cos(n\varphi) \cos(n\varphi_k) + \sin(n\varphi) \sin(n\varphi_k), \quad (9)$$

and using the following integral identities:

$$\int_0^{2\pi} \cos(n\varphi) \cos(n'\varphi) d\varphi = \pi \delta_{nn'}, \quad (10)$$

$$\int_0^{2\pi} \sin(n\varphi) \cos(n'\varphi) d\varphi = 0, \quad (11)$$

$$\int_0^{2\pi} \cos(n'\varphi) d\varphi = 0, \quad (12)$$

for $n, n' \neq 0$, where $\delta_{nn'}$ is the Kronecker delta function defined as:

$$\delta_{nn'} = \begin{cases} 1, & n = n' \\ 0, & n \neq n' \end{cases}. \quad (13)$$

We then have:

$$\int_0^{2\pi} p_t \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \cos(n'\varphi_k), \quad (n' > 0), \quad (14)$$

That means the wave from φ_k direction has the gain of $\cos(n'\varphi_k)$.

Similarly, if the weight is $\frac{\sin(n'\varphi)}{\pi B_{n'}(ka)}$, we have:

$$\int_0^{2\pi} p_t \frac{\sin(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \sin(n'\varphi_k), \quad (n' > 0), \quad (15)$$

In addition, when the weight is $\frac{1}{2\pi B_0(ka)}$, we have:

$$\begin{aligned} & \int_0^{2\pi} p_t \frac{1}{2\pi B_0(ka)} d\varphi \\ &= \frac{1}{2\pi B_0(ka)} \left\{ \int_0^{2\pi} B_0(ka) d\varphi \right. \\ & \quad \left. + \int_0^{2\pi} \sum_{n=1}^{\infty} B_n \cos[n(\varphi - \varphi_k)] d\varphi \right\} \\ &= 1. \end{aligned} \quad (16)$$

If the ideal beampattern is a peak at φ_0 and zero everywhere else, it can be modeled as the delta function $\delta(\varphi - \varphi_0)$ ¹. Its Fourier series expansion is:

$$\begin{aligned} \delta(\varphi - \varphi_0) &= \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} [\cos(n\varphi_0) \cos(n\varphi) \\ & \quad + \sin(n\varphi_0) \sin(n\varphi)]. \end{aligned} \quad (17)$$

Therefore, to achieve the ideal beampattern $f(\varphi, \varphi_0) = \delta(\varphi - \varphi_0)$, the weight for the microphone at point φ is:

$$\begin{aligned} w(\varphi, \varphi_0) &= \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi} \sum_{n=1}^{\infty} \left[\cos(n\varphi_0) \frac{\cos(n\varphi)}{\pi B_n(ka)} \right. \\ & \quad \left. + \sin(n\varphi_0) \frac{\sin(n\varphi)}{\pi B_n(ka)} \right] \\ &= \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi^2} \sum_{n=1}^{\infty} \frac{\cos[n(\varphi - \varphi_0)]}{B_n(ka)}. \end{aligned} \quad (18)$$

Some careful readers may have noticed the similarity between the formulas (3) and (6) in the sense that both are the infinite series of cosine functions except the coefficients are different. We rewrite (3) as:

$$p_i = \sum_{n=0}^{\infty} A_n \cos[n(\varphi - \varphi_k)], \quad (19)$$

where

$$A_n = \epsilon_n i^n J_n(kr). \quad (20)$$

It is plausible that the cosine components recorded by a circular microphone array in free space can also be extracted similarly as we did in section 2.2. So why bother to scatter the plane wave by a rigid cylinder?

The intuition is: the scattering makes the pressure distribution along the circular microphone array more complicated, especially for higher frequency wave, so more spatial information is captured. This is illustrated more clearly in Figure 2 and Figure 3. When ka is relatively small, the cylinder is approximately transparent to the sound wave,

¹It is also called "Dirac's delta function" or the "impulse symbol". Refer [8, pp. 31] for the formal definition. Don't be confused with the Kronecker delta function as (13).

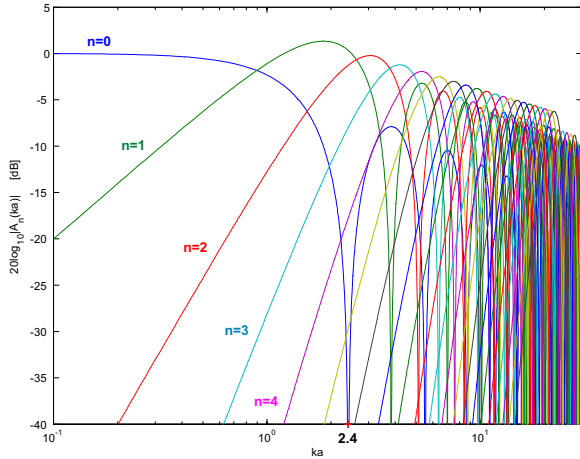


Figure 2. $A_n(ka)$ for n from 0 to 20.

so $A_n(ka)$ and $B_n(ka)$ have similar behaviors. When ka increases, the scattering effect is stronger, so the irregular pressure distribution along the circular array contains richer spatial frequency content. For example, in Figure 2, when ka lies in the first notch of $A_0(ka)$ which is about 2.4, apparently the spatial frequency component of $n = 0$ cannot be extracted robustly since it is too weak compared to the $n = 1$ component and hidden by noise. While at the same ka in Figure 3, the component of $n = 0$ is boosted to about -6 dB which is much easier to be extracted. There are more notches for larger ka in Figure 2 although $A_n(ka)$ finally converges with respect to n for given ka . On the contrary, in Figure 3, for fixed ka , $B_n(ka)$ exhibits smoother convergence with the increase of n .

In summary, with the help of the scattering from the cylinder, it is much easier to extract the desired spatial components.

2.3 Practical Design and Analysis

As an example, suppose we have a cylinder with radius of 10 cm, our array has 32 microphones positioned uniformly on the circle. We will analyze the performance of this system. Since we only care about the pressure on the microphones which is on the side surface of the cylinder, in practice, the cylinder can be of finite length and the microphone array can be placed away from both endpoints. The scattered sound field distortion is negligible at the microphone positions.

2.3.1 Discreteness

So far we have assumed a continuous circular microphone array. In practice, however, we have to use finite number of

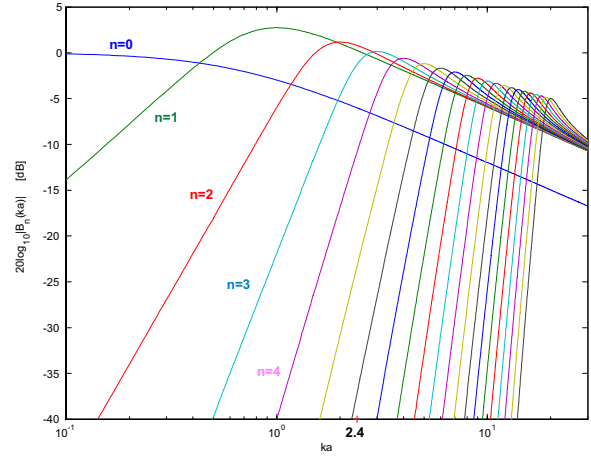


Figure 3. $B_n(ka)$ for n from 0 to 20.

microphones positioned discretely on a circle. According to the well-known sampling theorem, to fully reconstruct a band-limited signal, it must be sampled at a rate at least twice as fast as the highest frequency. This sampling rate is called the *Nyquist frequency*. This theorem is for both spatial and temporal sampling.

In our system, we assume the temporal sampling rate is already fast enough. Our concern is the spatial sampling rate with respect to the number of microphones on a circle. Using N microphones positioned uniformly on the circle, the sampling rate is N . So we can fully capture the sound field with highest spatial frequency of $n = N/2$. For example, in Figure 3, when $ka = 1$, the amplitude of the spatial frequency component of $n = 4$ is about -40 dB which is negligible in practice, so we may just say the highest spatial frequency is $n = 3$ and we need $N = 6$ microphones.

For the array with N microphones, if the highest spatial frequency with significant amplitude is M ($M \leq N/2$), the weight is:

$$w_M(\varphi, \varphi_0) = \frac{1}{4\pi^2 B_0(ka)} + \frac{1}{\pi^2} \sum_{n=1}^M \frac{\cos[n(\varphi - \varphi_0)]}{B_n(ka)}. \quad (21)$$

With fixed a , the spatial frequency content depends on the temporal frequency. This is also illustrated in Figure 3 when ka increases, the sound field along the circular array contains higher spatial frequencies. Spatial aliasing occurs when the sound field contains spatial frequency greater than $N/2$.

2.3.2 Practical Beampattern

The ideal beampattern is defined as (17). In the real array with finite number of microphones, using the practical

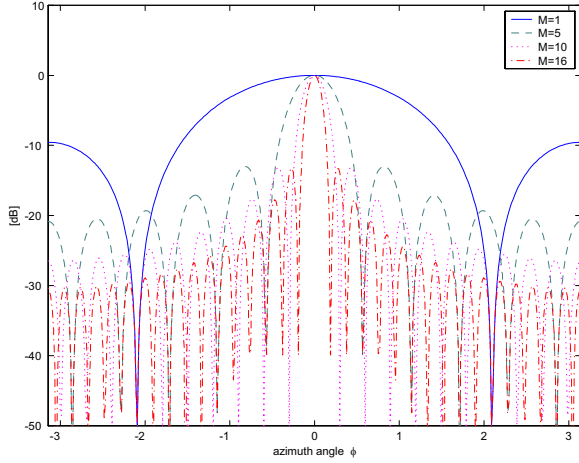


Figure 4. Beampatterns for $M = 1, 5, 10, 16$.

weight as (21), the actual beampattern is an M -truncated version of (17):

$$f_M(\varphi, \varphi_0) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^M \cos[n(\varphi - \varphi_0)]. \quad (22)$$

We say this beampattern is of order M . In Figure 4, we plot a few beampatterns of different orders. To show the spatial aliasing effects, we compare the beampatterns at high frequencies against the theoretical beampattern of order 16, which is the maximum order for our example. The plot is for $\theta = \pi/2$ and $\varphi = [-\pi, \pi]$ as shown in Figure 5. In practice, the maximum order of beampattern also depends on the ka value. As discussed in section 2.2, only the spatial frequency components with significant strength can be decomposed robustly.

2.3.3 Beampattern in 3D Sound Field

Our system is designed to make spatial filtering of two dimensional auditory scenes where sound sources are roughly on the same plane specified by the circular microphone array. In the real world, however, undesired sound can come from all 3D directions other than on this plane, such as the room reverberations or other noise.

In Figure 6, we plot the polar beampatterns for different elevation angles. It clearly shows when the wave direction is moving away from the plane specified by the microphone array, the beampattern tends to be less directive. This can be explained theoretically. Suppose the sound wave is from (θ_k, φ_k) and the component of the wave number \mathbf{k} perpendicular to the cylinder axis (z -axis) is \mathbf{k}_\perp , the incident wave is:

$$p_i = e^{i\mathbf{k}\cdot\mathbf{r}} = e^{i\mathbf{k}_\perp\cdot\mathbf{r}} = e^{ik_\perp r \cos(\varphi - \varphi_k)}, \quad (23)$$

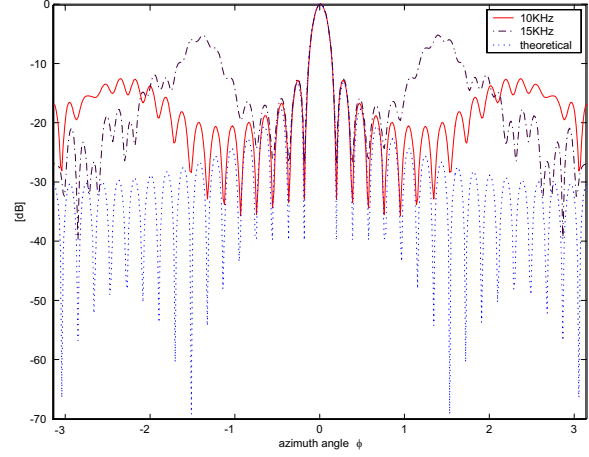


Figure 5. Spatial aliasing.

where:

$$k_\perp = k \sin \theta_k. \quad (24)$$

Accordingly, (14), (15) and (16) become:

$$\int_0^{2\pi} p_t \frac{\cos(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \frac{B_{n'}(k_\perp a)}{B_{n'}(ka)} \cos(n'\varphi_k), \quad (25)$$

$$\int_0^{2\pi} p_t \frac{\sin(n'\varphi)}{\pi B_{n'}(ka)} d\varphi = \frac{B_{n'}(k_\perp a)}{B_{n'}(ka)} \sin(n'\varphi_k), \quad (26)$$

$$\int_0^{2\pi} p_t \frac{1}{2\pi B_0(ka)} d\varphi = \frac{B_0(k_\perp a)}{B_0(ka)}, \quad (27)$$

$(n' > 0).$

Therefore, the actual M -truncated beampattern for θ_k is:

$$f_M(\varphi, \varphi_0, \theta_k) = \frac{1}{2\pi} \frac{B_0(k_\perp a)}{B_0(ka)} + \frac{1}{\pi} \sum_{n=1}^M \frac{B_n(k_\perp a)}{B_n(ka)} \cos[n(\varphi - \varphi_0)]. \quad (28)$$

From Figure 3, we notice that except $B_0(ka)$ is monotonically decreasing, each $B_n(ka)$ ($n > 0$) has only one peak. On the left side of its peak, it is monotonically increasing and on the right side monotonically decreasing². Since $k_\perp a < ka$ for $\theta_k \neq \pi/2$, there exists a threshold integer m ($0 < m < M$) such that:

$$\begin{cases} \frac{B_n(k_\perp a)}{B_n(ka)} \geq 1, & 0 \leq n \leq m \\ \frac{B_n(k_\perp a)}{B_n(ka)} < 1, & m < n \leq M \end{cases}. \quad (29)$$

That means the higher spatial frequency components are suppressed while the lower spatial frequency components

²Strict theoretical proof omitted here.

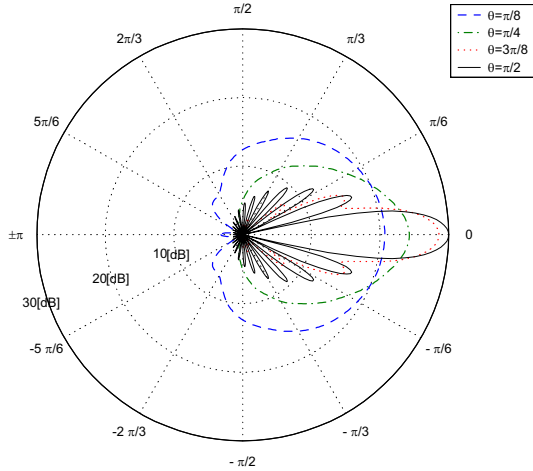


Figure 6. Polar Beampatterns for different elevation angles at 5 kHz. The beampattern for $\theta = \pi/2$ is of order 13.

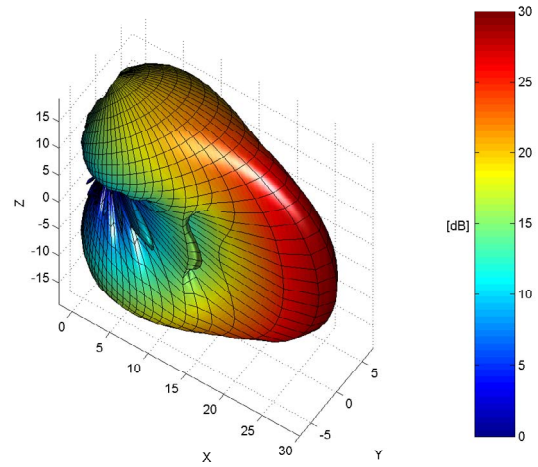


Figure 7. 3D beampattern of order 13 at 5 kHz, looking at $\varphi = 0$. (Lighted to show 3D effect.)

are amplified. Apparently, this low-pass filter makes the resulted beampattern less directive.

The full 3D beampattern is shown in Figure 7. It has a strawberry-like shape. Therefore, our array can also be used in 3D auditory scenes where spatial filtering in the azimuthal dimension is desired.

3 Reproducing the Recorded Surround Auditory Scene

Suppose we have recorded the surround auditory scene using our circular microphone array, the next step is to reproduce it with a circular loudspeaker array. In this array, L loudspeakers are positioned uniformly around a circle in free space. To achieve the optimal performance given the number of microphones (N) used in recording, we let L equal to N . For clarity, we will still use the notation L for the number of loudspeakers. In this section, we will first consider the plane wave case where the radius of the circular loudspeaker array is large enough to produce plane waves at the observation points. We will derive the theoretical condition of reproduction, and give a simple solution to ensure the optimal performance. Then we will extend this into the spherical wave case where the circle of the loudspeaker array is small so that every loudspeaker can be modeled as a point sound source.

3.1 Plane Wave Case

We denote the L loudspeaker positions as $\varphi_l, l = 1, \dots, L$, which are also the plane wave incident directions. Suppose we feed every loudspeaker with the identical signal of wave number k , and the output of the l -th loudspeaker is weighted by a complex number $a_l(k)$, the reproduced sound field at the observation point φ is a weighted combination of plane waves:

$$p_c = \sum_{n=0}^{\infty} \epsilon_n i^n J_n(kr) \sum_{l=1}^L a_l(k) \cos [n(\varphi - \varphi_l)]. \quad (30)$$

To perfectly reproduce the original sound field, it requires (3) equal (30). Thus the condition for reproduction is:

$$\cos [n(\varphi - \varphi_k)] = \sum_{l=1}^L a_l(k) \cos [n(\varphi - \varphi_l)], \quad (31)$$

$$(n = 0, \dots, \infty).$$

That is:

$$\cos (n\varphi_k) = \sum_{l=1}^L a_l(k) \cos (n\varphi_l), \quad (32)$$

$$\sin (n\varphi_k) = \sum_{l=1}^L a_l(k) \sin (n\varphi_l), \quad (33)$$

$$(n = 0, \dots, \infty).$$

However, since we only can record the sound field up to order M , the reproduction can be achieved to order M at

most. If we let:

$$a_l(k) = f_M(\varphi_k, \varphi_l) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^M \cos[n'(\varphi_k - \varphi_l)], \quad (34)$$

then (32) and (33) can be satisfied to order M as following:

$$\begin{aligned} & \sum_{l=1}^L a_l(k) \cos(n\varphi_l) \\ &= \sum_{l=1}^L \left\{ \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^M \cos[n'(\varphi_k - \varphi_l)] \right\} \cos(n\varphi_l) \\ &= \cos(n\varphi_k) \quad (n = 0, \dots, M), \end{aligned} \quad (35)$$

and

$$\begin{aligned} & \sum_{l=1}^L a_l(k) \sin(n\varphi_l) \\ &= \sum_{l=1}^L \left\{ \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^M \cos[n'(\varphi_k - \varphi_l)] \right\} \sin(n\varphi_l) \\ &= \sin(n\varphi_k) \quad (n = 0, \dots, M). \end{aligned} \quad (36)$$

The physical meaning is clear: if each loudspeaker plays back the beamformed signal in this direction, the loudspeaker array will recreate the original surround sound field to order M .

3.2 Spherical Wave Case

Thus far, we have assumed the loudspeakers are far enough from the observation points to generate plane waves. In this section, we will extend the algorithm to the general point-source model where the reproduced sound field is a weighted combination of spherical waves.

Suppose the radius of the circular loudspeaker array is r_0 . If the loudspeakers can be modeled as point sources, the sound field at the observation point (r, θ, φ) reproduced by the circular loudspeaker array becomes [2]:

$$\begin{aligned} p_c &= 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) R_n(kr_0) \\ &\quad \times \sum_{m=-n}^n \sum_{l=1}^L a_l(k) Y_n^m(\theta_l, \varphi_l) Y_n^{m*}(\theta, \varphi), \end{aligned} \quad (37)$$

where

$$R_n(kr_0) = -ikr_0 e^{ikr_0} i^{-n} h_n(kr_0). \quad (38)$$

Normally, the listener is at the center of the circular ring of

the loudspeakers, so $\theta_l = \pi/2$ and $\theta = \pi/2$. Since

$$\begin{aligned} Y_n^m\left(\frac{\pi}{2}, \varphi\right) &\equiv \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(\cos \frac{\pi}{2}) e^{im\varphi} \\ &= \sqrt{\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!}} P_n^m(0) e^{im\varphi}, \end{aligned} \quad (39)$$

the recreated sound field is:

$$\begin{aligned} p_c &= 4\pi \sum_{n=0}^{\infty} i^n j_n(kr) R_n(kr_0) \sum_{l=1}^L a_l(k) \\ &\quad \times \sum_{m=-n}^n \left[\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} \right] [P_n^m(0)]^2 e^{im(\varphi_l - \varphi)}. \end{aligned} \quad (40)$$

Using the identity relating associated Legendre polynomials with negative m to the corresponding functions with positive m :

$$P_n^{-m}(x) = (-1)^m \frac{(n-m)!}{(n+m)!} P_n^m(x), \quad (41)$$

we have:

$$\begin{aligned} & \sum_{m=-n}^n \left[\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} \right] [P_n^m(0)]^2 e^{im(\varphi_l - \varphi)} \\ &= \sum_{m=0}^n \epsilon_m \left[\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} \right] [P_n^m(0)]^2 \cos[m(\varphi_l - \varphi)]. \end{aligned} \quad (42)$$

let

$$E_n = 4\pi i^n j_n(kr) R_n(kr_0), \quad (43)$$

$$F_n^m = \epsilon_m \left[\frac{2n+1}{4\pi} \frac{(n-m)!}{(n+m)!} \right] [P_n^m(0)]^2. \quad (44)$$

So we have:

$$\begin{aligned} p_c &= \sum_{l=1}^L a_l(k) \sum_{n=0}^{\infty} E_n \sum_{m=0}^n F_n^m \cos[m(\varphi_l - \varphi)] \\ &= \sum_{l=1}^L a_l(k) \sum_{m=0}^{\infty} \left(\sum_{n=m}^{\infty} E_n F_n^m \right) \cos[m(\varphi_l - \varphi)] \\ &= \sum_{l=1}^L a_l(k) \sum_{n=0}^{\infty} \left(\sum_{m=n}^{\infty} E_m F_m^n \right) \cos[n(\varphi_l - \varphi)] \\ &= \sum_{n=0}^{\infty} \left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^L a_l(k) \cos[n(\varphi_l - \varphi)]. \end{aligned} \quad (45)$$

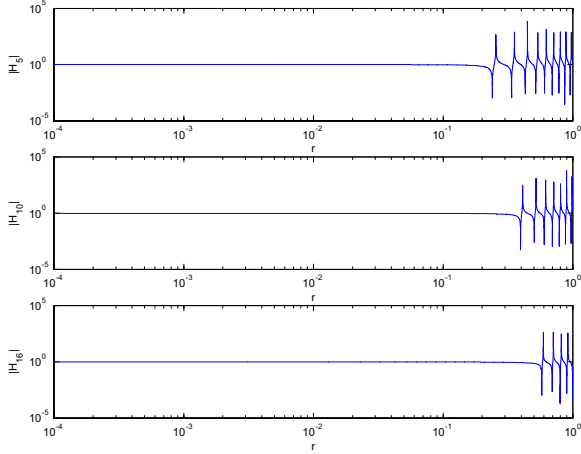


Figure 8. H_5, H_{10} and H_{16} at frequency 2 kHz.

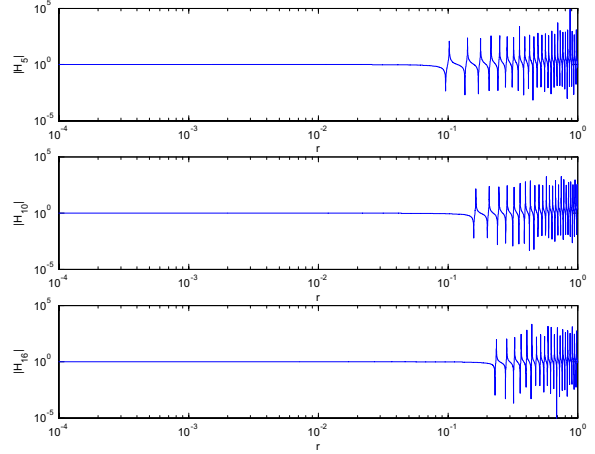


Figure 9. H_5, H_{10} and H_{16} at frequency 5 kHz.

Comparing (45) against (3) or (19), the condition for reproduction is:

$$\left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^L a_l(k) \cos[n(\varphi_l - \varphi)] = A_n \cos[n(\varphi - \varphi_k)], \quad (n = 0, \dots, \infty). \quad (46)$$

That is:

$$\left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^L a_l(k) \cos(n\varphi_l) = A_n \cos(n\varphi_k), \quad (47)$$

$$\left(\sum_{m=n}^{\infty} E_m F_m^n \right) \sum_{l=1}^L a_l(k) \sin(n\varphi_l) = A_n \sin(n\varphi_k), \quad (48)$$

$$(n = 0, \dots, \infty).$$

Similarly, (47) and (48) can be satisfied for $n = 0$ to M with the solution of $a_l(k)$ as:

$$a_l(k) = \frac{A_n}{\sum_{m=n}^{\infty} E_m F_m^n} f_M(\varphi_k, \varphi_l) = H_n(k, r, r_0) \left\{ \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n'=1}^M \cos[n'(\varphi_k - \varphi_l)] \right\}, \quad (49)$$

which is still the beampattern but multiplied by a variable coefficient:

$$H_n(k, r, r_0) = \frac{A_n}{\sum_{m=n}^{\infty} E_m F_m^n}. \quad (50)$$

To analyze its effects on the reproduced sound field, we plot $H_n(k, r, r_0)$ with $r_0 = 1$ m. Figure 8 plots H_5, H_{10} and

H_{16} at 2 kHz with respect to r . Figure 9 plots the same functions at 5 kHz. Several interesting observations are:

1. If r is small compared with r_0 , i.e. the observation points are in the far field area, $H_n \approx 1$. This is roughly the same as the plane wave case;
2. At fixed frequency, with larger n , the far field area stretches. This is because the first zero of H_n occurs at the same r as the first zero of $J_n(kr)$, the n -th order Bessel function of the first kind;
3. The far field area shrinks as the frequency increases, which is obvious since the first zero of $J_n(kr)$ for fixed n is a constant number, when k increases, r has to decrease to stay in the far field area.

3.3 Simulations

We simulate the reproductions of a plane wave of 5 kHz incident from $\varphi = \pi/4$. Figure 10-12 show the reproductions by plane waves to the order 5, 10, 16, respectively. Figure 13-15 show the reproductions by spherical waves at the same orders with $r_0 = 1$ m. Figure 16 shows the reproduction to order 5 at 2 kHz. As we have predicted, the reproduction is effective in a larger area than in Figure 13.

4 Implementation Issues and Future Work

In addition to the practical design we have performed in previous sections, some implementation issues need to be addressed in building an efficient real-world system. Specifically, we will come back to the teleconference example in section 1.3 to illustrate a few of them.

One of the most desirable features in teleconference is real-time interaction over a band-limited network. Therefore, we must have an efficient way to compress, transmit and decompress our surround auditory data. Since the auditory scenes are recorded using a microphone array, the multichannel signals are highly correlated with each other. That clearly requires a highly efficient codec. In addition, to accommodate various network resources and computing capacities, we also need a multi-resolution data structure and an adaptive data transmission algorithm. These issues will be included in the future work.

5 Summary and Acknowledgments

In this paper, we proposed an integrated approach to record and reproduce high order surround auditory scenes, which can be used in creating mixed and augmented reality systems. We designed a circular microphone array mounted on a rigid cylinder. We then used it in surround sound recording and beamforming. We derived the conditions for perfect reproduction. Given the recordings in finite orders, we designed a circular loudspeaker array to reproduce the original scenes optimally. Design examples and performance analyses are presented. A prototype is currently being built.

The authors would like to thank the four anonymous reviewers and our shepherd Dr. Kiyoshi Kiyokawa for the reviews and comments which helped to improve the overall quality of this paper. This work was partially supported by NSF Award 0205271.

References

- [1] A. Berkhout, D. Vries, and P. Vogel. Acoustic control by wave field synthesis. *Journal of the Acoustical Society of America*, 93(5):2764–2778, May 1993.
- [2] D. Colton and R. Kress. *Inverse Acoustic and Electromagnetic Scattering Theory*. Springer-Verlag, New York, 1997.
- [3] C. Cruz-Neira, D. Sandin, and T. Defanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In *Proceedings of ACM SIGGRAPH 1993*, pages 135–142, Anaheim, CA, Aug. 1993.
- [4] K. Li, H. Chen, Y. Chen, D. Clark, P. Cook, S. Damianakis, G. Essl, A. Finkelstein, T. Funkhouser, A. Klein, Z. Liu, E. Praun, R. Samanta, B. Shedd, J. Singh, G. Tzanetakis, and J. Zheng. Building and using a scalable display wall system. *IEEE Computer Graphics and Applications*, 20(4):29–37, 2000.

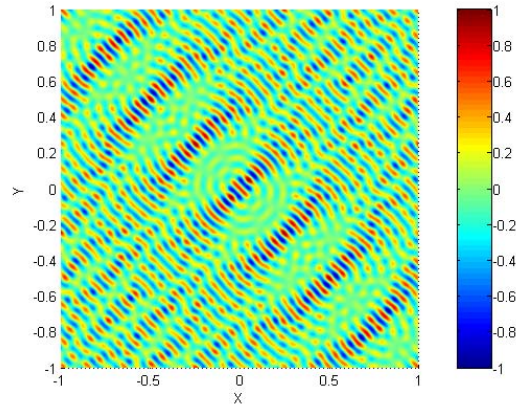


Figure 10. Reproduce the 5 kHz plane wave to order 5 using plane waves.

- [5] Z. Li, R. Duraiswami, and N. A. Gumerov. Capture and recreation of higher order 3d sound fields via reciprocity. In *Proceedings of the 10th International Conference on Auditory Display (ICAD2004)*, Sydney, Australia, July 6-9 2004.
- [6] Z. Li and A. Varshney. A real-time seamless tiled display for 3d graphics. In *Proceedings of 7th Annual Symposium on Immersive Projection Technology (IPT 2002)*, Orlando, FL, March 24 - 25 2002.
- [7] A. Majumder, Z. He, H. Towles, and G. Welch. Achieving color uniformity across multi-projector displays. In *IEEE Visualization 2000*, pages 117–124, Salt Lake City, Oct. 2000.
- [8] P. Morse and K. Ingaard. *Theoretical Acoustics*. McGraw Hill, New York, 1968.
- [9] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, and H. Fuchs. The office of the future : A unified approach to image-based modeling and spatially immersive displays. In *Proceedings of ACM SIGGRAPH'98*, pages 179–188, Orlando, Florida, July 1998.
- [10] H. Teutsch, S. Spors, W. Herbordt, W. Kellermann, and R. Rabenstein. An integrated real-time system for immersive audio applications. In *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, Oct. 2003.
- [11] D. B. Ward and T. D. Abhayapala. Reproduction of a plane-wave sound field using an array of loudspeakers. *IEEE Transactions on Speech and Audio Processing*, 9(6):697–707, Sept. 2001.

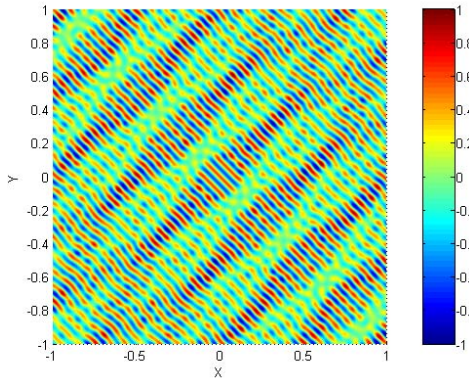


Figure 11. Reproduce the 5 kHz plane wave to order 10 using plane waves.

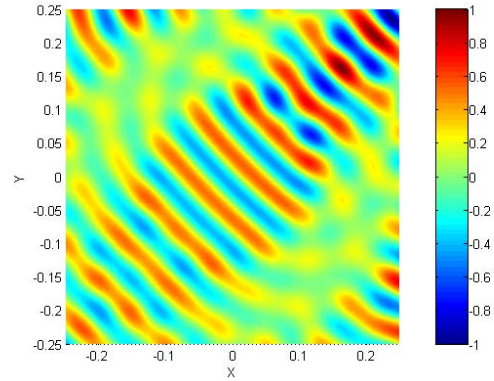


Figure 14. Reproduce the 5 kHz plane wave to order 10 using spherical waves.

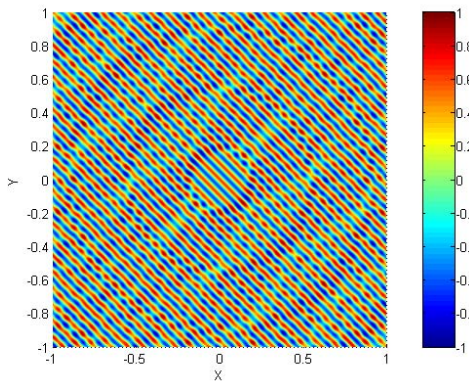


Figure 12. Reproduce the 5 kHz plane wave to order 16 using plane waves.

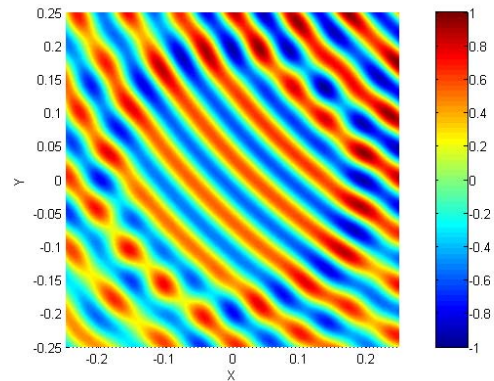


Figure 15. Reproduce the 5 kHz plane wave to order 16 using spherical waves.

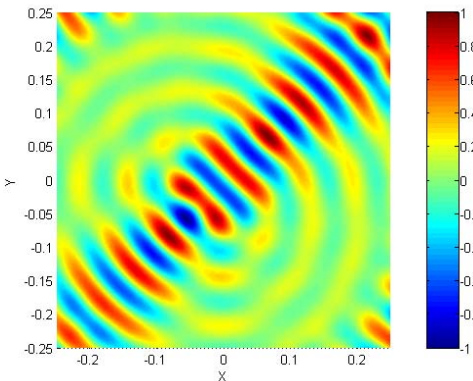


Figure 13. Reproduce the 5 kHz plane wave to order 5 using spherical waves.

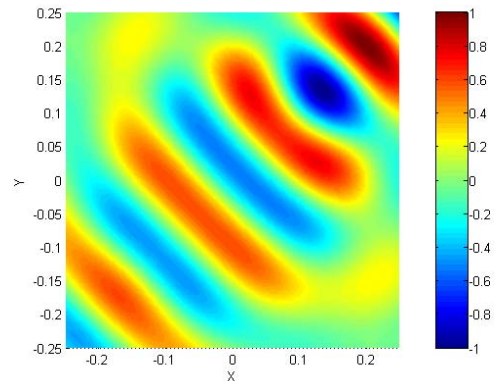


Figure 16. Reproduce the 2 kHz plane wave to order 5 using spherical waves.