inter.noise 2014
MELBOURNE AUSTRALIA
16-19 NOVEMBER

Linear optimal source distribution mapping for binaural sound reproduction

Jianwen ZHENG[1]; Jing LU[2]; Xiaojun QIU[3]

[1, 2, 3] Key Lab of Modern Acoustics, Institute of Acoustics, Nanjing University, China

ABSTRACT

Binaural playback over loudspeakers usually leads to crosstalk. The most effective strategy of general crosstalk cancellation systems is system inversion which, nevertheless, gives rise to several problems such as the loss of dynamic range and the lack of robustness. The optimal source distribution (OSD) method has been utilized to compromise the playback performance and the problems of the crosstalk cancellation; however, the method was based on the circular loudspeaker array and is hard to be implemented in some practical applications. The other problem associated with the method is that the frequency-span discretization only concentrates on the dynamic range loss. This paper proposes a novel mapping method by adjusting the circular loudspeaker array to the linear array, where the optimization of the discretization is discussed by taking full consideration of the crosstalk cancellation performance and the frequency response at the sweet spot. Both simulations and experiments demonstrate the efficacy of the proposed method. The proposed system has potential applications in both sound field manipulation and subjective noise evaluation.

Keywords: Binaural playback, crosstalk cancellation, optimal source distribution.
I-INCE Classification of Subjects Number(s): 03.1

1. INTRODUCTION

When binaural sound signals containing directional cues are presented with loudspeakers, the listener is likely to perceive the localization of sound images and experience realistic three-dimensional sound environment. However, the performance of the system is usually affected by the crosstalk. Therefore, it is necessary to preprocess the binaural signals by using the crosstalk cancellation system (CCS). In the past decades, a lot of methods have been proposed to simplify the implementation and to improve the performance of CCS after Bauer firstly introduced the idea [1]. The robustness of the system in addition to the reproduction performance has attracted researchers' interest [2]. The optimal loudspeaker position for CCS was discussed [3] and an effective system called "stereo dipole" was developed against the head movement away from the sweet spot [4]. The optimization procedure for the application of more than two loudspeakers and the analysis of optimal loudspeaker configuration were proposed in References 5 and 6, among which the optimal source distribution (OSD) was of particular interest in this paper [6]. Theoretically this approach effectively reconciles the crosstalk cancellation performance, the loss of dynamic range and the robustness of the whole playback system.

The OSD method is usually realized in circular array form, nevertheless, the circular configuration limits its practical application, and it is of significance to investigate the linear mapping method of OSD. Furthermore, the discretization strategy only concentrates on increasing the dynamic range while neglecting the performance of crosstalk cancellation and frequency response. In this paper, a theoretical model on the linear array with consideration of head scattering is established, and a numerical method for regulating the frequency range of different source pairs under the limited array scale is proposed, which takes consideration of both the crosstalk cancellation performance and the frequency response at the sweet spot. To objectively investigate the performance of the binaural reproduction system, the commonly used method is to calculate the

[1] james_zheng89@163.com
[2] lujing@nju.edu.cn
[3] xjqiu@nju.edu.cn

channel separation. However, it does not take consideration of the localization cues of human hearing systems. In this paper, a typical objective parameter interaural time difference (ITD), which is well related to the localization cues of human auditory systems, is used apart from the normally used crosstalk cancellation calculation.

2. THEORY AND METHODS

2.1 Principles of Binaural Reproduction System

The principle of binaural reproduction over loudspeakers based on OSD is described by a 2×2 CCS as illustrated in Figure 1. The goal of this system is to present the 2-channel binaural signals to each ear of the listener independently while the signals of each loudspeaker are fed to both ears (including crosstalk).

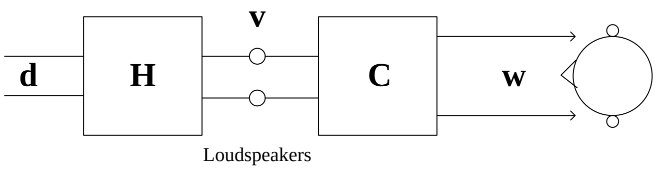


Figure 1 – Block diagram for binaural reproduction based on OSD

The signals received by two ears are given by **w**,

$$\mathbf{w} = \mathbf{Cv} \quad (1)$$

where **C** is a 2×2 matrix of transfer functions between loudspeakers and receivers, **v** is the source strengths of two monopole loudspeakers. The desired signals **d** to be synthesized at the receivers can be any stereo sound tracks. The inverse of the transfer function matrix, i.e., the CCS matrix **H,** is introduced to form the loudspeakers driving signal as

$$\mathbf{v} = \mathbf{Hd} \quad (2)$$

so that

$$\mathbf{w} = \mathbf{CHd} \quad (3)$$

Therefore, the crosstalk cancellation performance matrix **R** is defined by

$$\mathbf{R} = \mathbf{CH} \quad (4)$$

In general, **C** is noninvertible because it is usually ill-conditioned and its condition number $\kappa(\mathbf{C})$ is large [3]. To mitigate this problem and design practical filters, a regularization procedure is introduced [7], where the pseudo-inverse matrix **H** is given by

$$\mathbf{H} = [\mathbf{C}^{\mathrm{H}}\mathbf{C} + \beta\mathbf{I}]^{-1}\mathbf{C}^{\mathrm{H}} \quad (5)$$

where β is a regularization parameter. Similarly, the CCS filter can be designed using the time domain method [8]. The CCS matrix **H** can be designed so that the binaural signals **w** can approximate the desired signals **d** with a certain delay [8]. Since various errors are in the process, the matrix **R** deviates from the identity matrix so that the diagonal term and the nondiagonal term of matrix **R** show the portion of desired signal transmission and the crosstalk components respectively [9]. The inversion of ill-conditioned system usually leads to loss of dynamic range and lack of robustness. The OSD method can mitigate this problem by means of a pair of conceptual monopole loudspeakers whose span varies continuously as a function of frequency [6]. However, the circular configuration of the CCS based on OSD theory limits its practical application, and it is of significance to investigate the linear mapping method of OSD.

2.2 Linear Mapping Method

An intuitive and feasible linear mapping method is shown in Figure 2, where the major problem is that the source span angle cannot be set very large due to the linear array scale limit, In our example a span of 2.8 m is used, and the distance between the center of the array and the head (denoted as L) is 1.4 m. As can be seen, the sound propagation distance between the source and the receiver is $L/\cos\theta$. Accordingly, condition number $\kappa(\mathbf{C})$ can be plotted as a function of frequency and source pair spacing, as shown in Figure 3, where the transfer functions are calculated by using a rigid sphere model [10], which can well approximate the head related transfer function (HRTF) without interpolation.

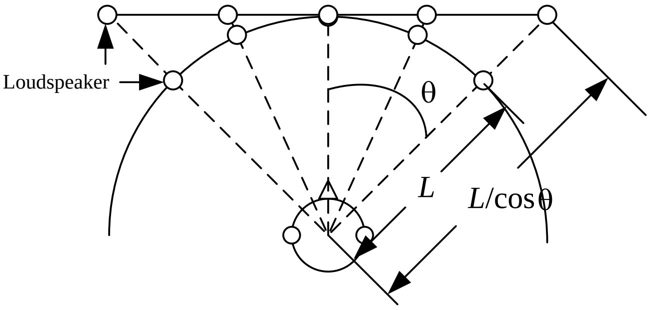


Figure 2 – Circular loudspeaker array to linear array

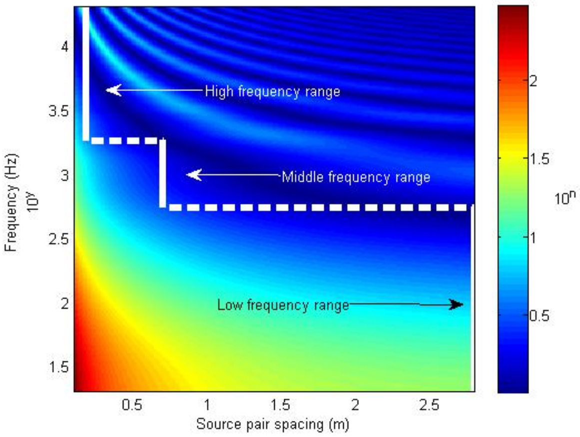


Figure 3 – Condition number $\kappa(\mathbf{C})$ of the matrix **C** as a function of frequency and source pair spacing

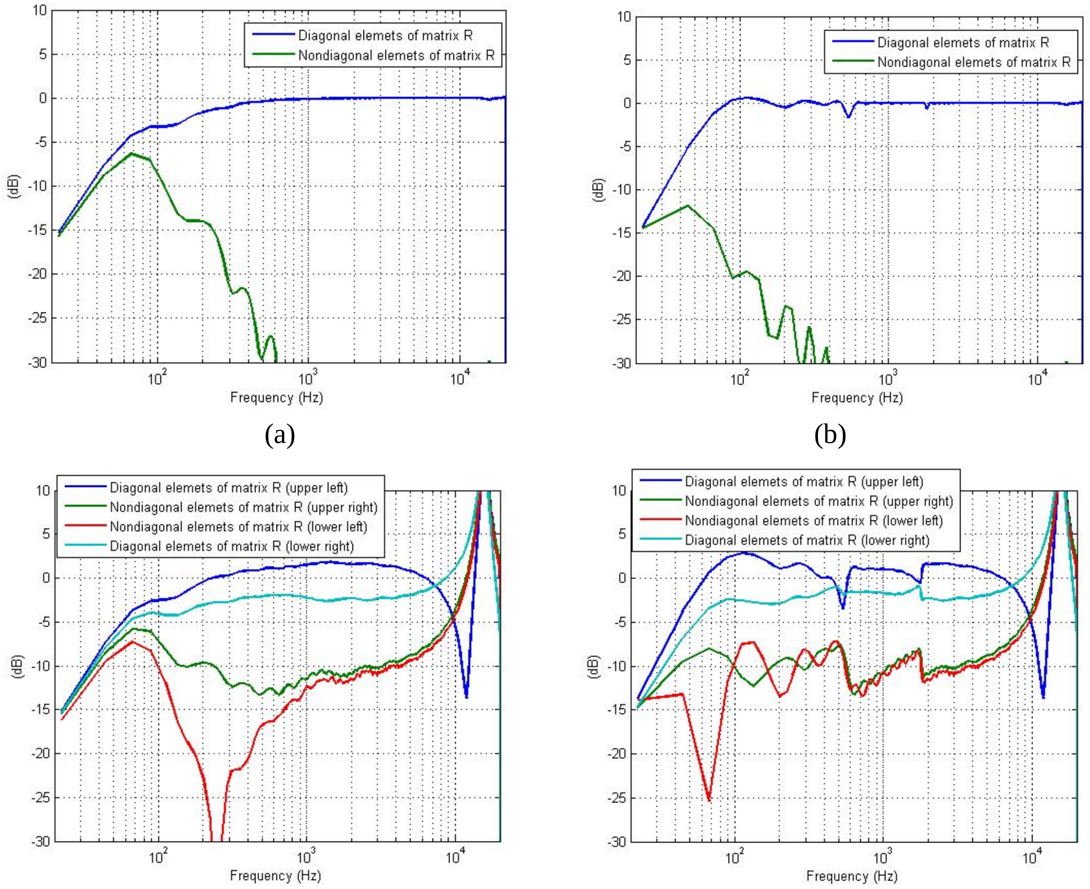
2.3 Frequency Range and Source Distribution

The discretization of the frequency range and the distribution of the sources in the CCS are related to the condition number $\kappa(\mathbf{C})$ as illustrated in Figure 3. There are three basic principles to fulfill the discretization task: (1) the source pair spacing and the frequency range should lead to a low $\kappa(\mathbf{C})$; (2) overlap of frequency range of different source pairs should be avoided; (3) the frequency range of all the source pairs should cover the whole audible range. However, due to the limited linear array scale and the complex $\kappa(\mathbf{C})$ distribution, it is hard to meet all the above three principles simultaneously. Therefore the source pair positions and the frequency range of different source pairs need to be regulated carefully through proper numerical simulations, which take full consideration of both the crosstalk cancellation performance and the frequency response at the sweet spot, as well as the dynamic range loss.

The crosstalk cancellation performance can be measured by the matrix **R** from equation (4), and the frequency response is generally measured by the received signal at the sweet spot when the desired signal is white noise. When the matrix **R** deviates from the identity matrix significantly, or the performance of the frequency response degrades around the crossover frequencies between different source pairs, the frequency range and the source distribution should be adjusted to a better combination.

Take a three-pair array for example. A pair of high-frequency units with a spacing of 0.18 m is chosen to cover the frequency range from 1.8 kHz to 20 kHz, while a pair of low-frequency units with a spacing of 2.8 m is chosen to cover the frequency range below 550 Hz. The spacing for the mid-frequency units filling the frequency gap is 0.7 m. The distributions are shown as white lines in Figure 3. As a comparison, stereo dipole is also introduced and the spacing is the same as that of the high-frequency units. For simplicity in the simulation, the loudspeakers are assumed to be point sources and the room reflection is neglected. The distance between the array and the dummy head is 1.4 m. The CCS matrix of inverse filtering is calculated by using equation (5) in the time domain, where β is 0.001. The sampling rate is 44.1 kHz and the length of inverse filter is 512.

The crosstalk cancellation performance obtained with stereo dipole and linear array are shown in Figure 4. The diagonal and nondiagonal elements of matrix **R**=**CH** are depicted, where **C** is the transfer function matrix obtained with rigid sphere model. As can be seen in Figures 4(a) and (b), the linear array crosstalk cancellation performance is significantly better than the dipole in low frequency range. Figures 4(c) and (d) show the cancellation performance when the center of the head is 0.05 m left from the sweet spot. The performance degrades significantly but the benefit of the linear array is still obvious in low frequency range.

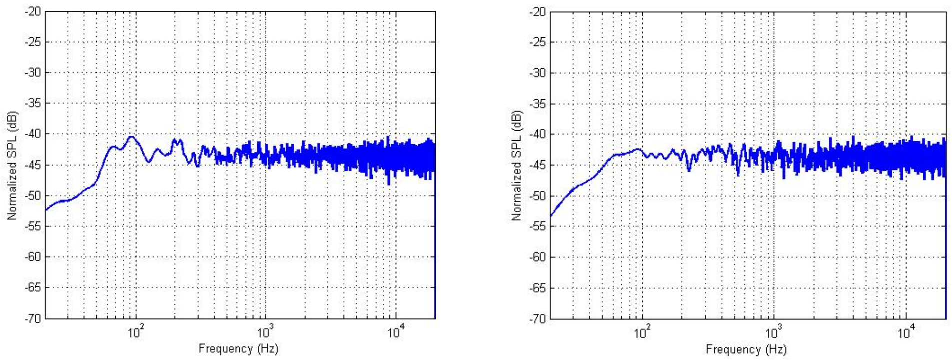


(a) (b)

(c) (d)

Figure 4 – Crosstalk cancellation performance. (a) Stereo dipole (sweet spot). (b) Linear array (sweet spot). (c) Stereo dipole (5 cm displacement). (d) Linear array (5 cm displacement).

The frequency response of both systems is shown in Figure 5. The input binaural signals are identical white noise signals. The output signal is received by one ear of the head at the sweet spot. Both systems show similar flat frequency response.



(a) (b)

Figure 5 – Frequency response. (a) Stereo dipole (sweet spot). (b) Linear array (sweet spot).

The norm of matrix **H** indicating the dynamic range loss, as depicted in Figure 6. It can be found that the linear array has less loss of the dynamic range than the stereo dipole in middle and low frequency range.

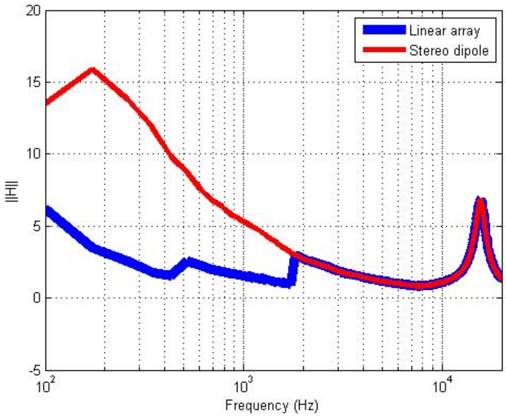


Figure 6 – The norm of matrix **H**.

2.4 Interaural Time Difference Evaluation

To evaluate the performance of the proposed system, the interaural time difference (ITD) [11], which is well related to the localization cues of human auditory systems, is used apart from the normally used cross-talk cancellation calculation. ITD brings preferable spatial impression for the sound field and is calculated by the position of the interaural cross-correlation peak [12] within the maximum possible interaural delay time of 1 ms. It is defined as the following,

$$\mathrm{ITD}(\tau) = \mathrm{argmax}(\mathrm{E}\{p_l(t)p_r(t+\tau)\}) \tag{6}$$

where $p_l(t)$ and $p_r(t)$ are sound signals received at the position of left and right ears respectively and filtered by a 700 Hz low-pass filter, since the frequency range below 700 Hz is considered as the ITD dominant range [11].

The difference between the expected and reproduced ITD received at the sweet spot is plotted in Figures 7(a). The ITD difference obtained with linear array is less than that obtained with stereo dipole, since the cancellation performance of linear array is better than that of stereo dipole below 700 Hz. The ITD difference of both systems increases when the center of the head is 0.05 m left from the sweet spot, as can be seen in Figures 7(b), and the linear array still has better performance.

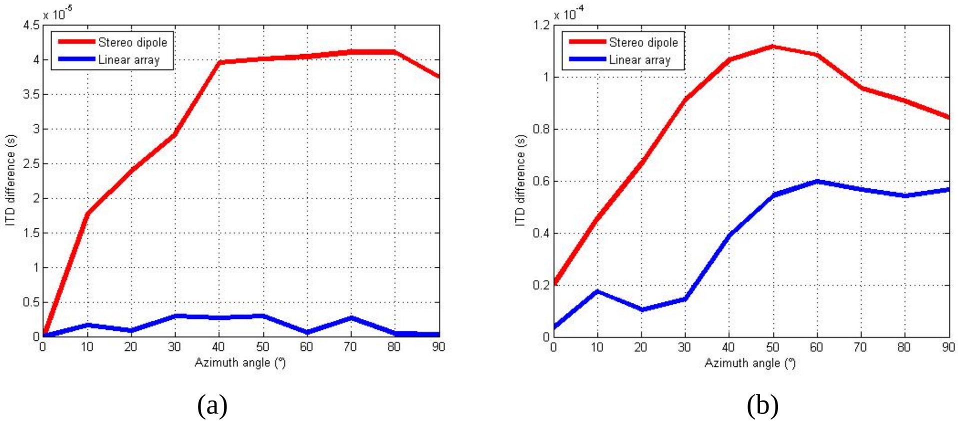


Figure 7 – Difference between expected and reproduced ITD. (a) At the sweet spot. (b) 5 cm displacement.

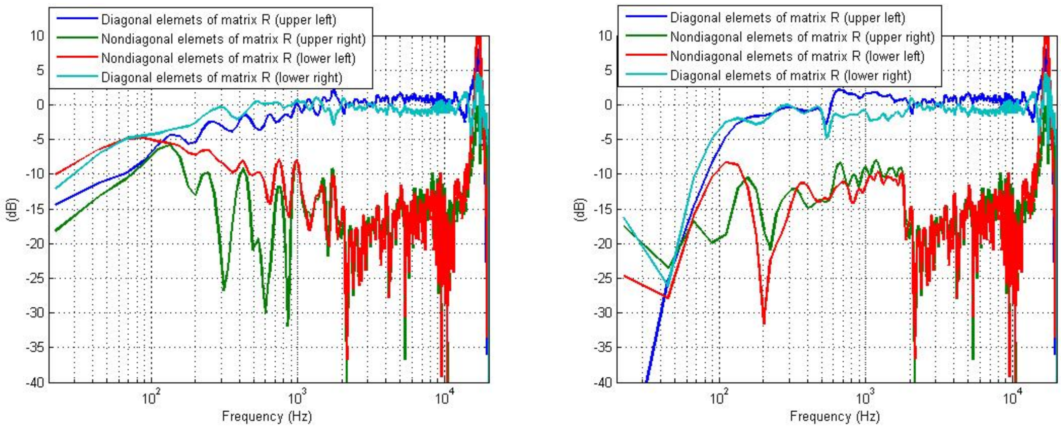
3. EXPERIMENTAL INVESTIGATIONS

The experimental arrangement is shown in Figure 8. The configuration of the proposed three-source-pair linear array is placed in a listening room. The loudspeaker array is kept at the same height as the ears of the dummy head, where two microphones are embedded respectively.

Figure 8 – The photo of the experimental arrangement in a listening room.

3.1 Objective Tests

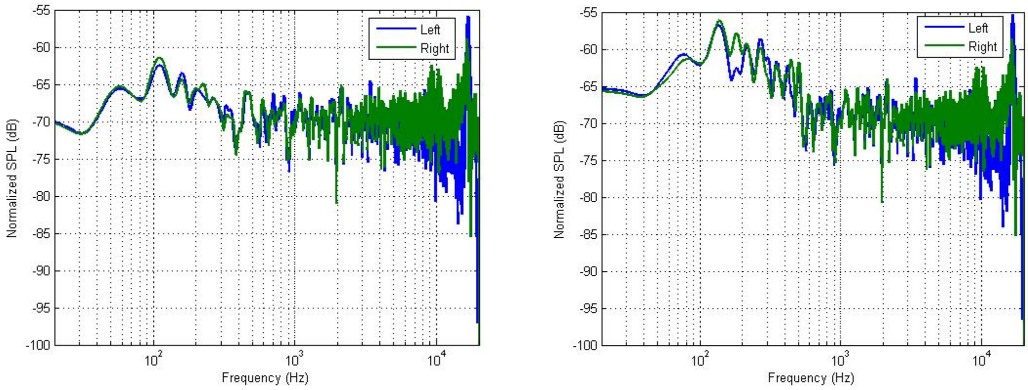
The crosstalk cancellation performance, the frequency response and the ITD are utilized as the objective evaluation indices. Figure 9 shows the cancellation performance obtained with the stereo dipole and the linear array at the sweet spot. As can be seen, the performance of the linear array is better than that of the stereo dipole at low frequencies as predicted in the simulation.



(a) (b)

Figure 9 – Crosstalk cancellation performance. (a) Stereo dipole (sweet spot). (b) Linear array (sweet spot).

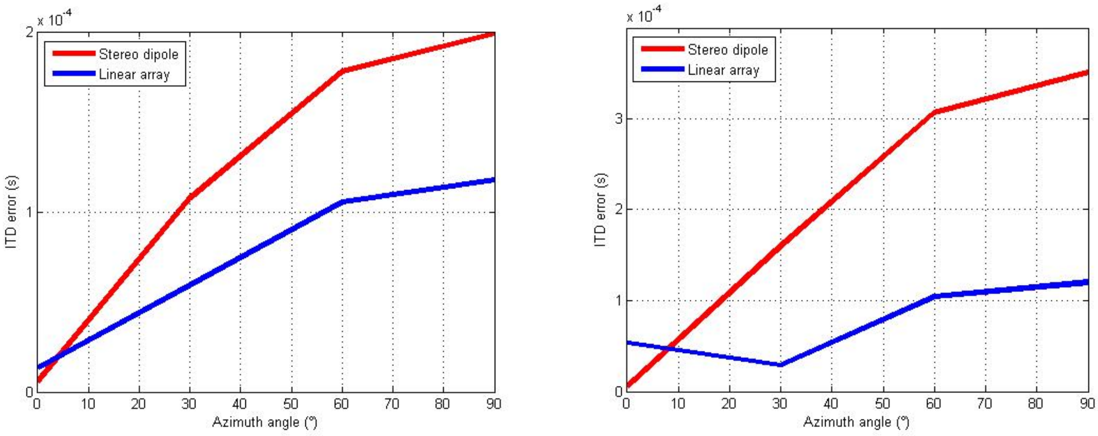
Figure 10 depicts the frequency response of both systems at the sweet spot. The variation is significantly larger than those in simulations especially in low and high frequency range since $\kappa(\mathbf{C})$ cannot be kept low throughout the whole frequency range (as depicted in Figure 3) due to the limited array scale.



(a) (b)

Figure 10 – Frequency response. (a) Stereo dipole (sweet spot). (b) Linear array (sweet spot).

Figure 11 demonstrates the difference between the expected and reproduced ITD of both systems at the sweet spot and when the head is 0.05 m left from the sweet spot. Similar to the simulations, the ITD difference of the linear array is less than that of the stereo dipole. The ITDs measurement results show that the proposed system is more robust to the head movement, which is helpful in enlarging the sweet spot area.



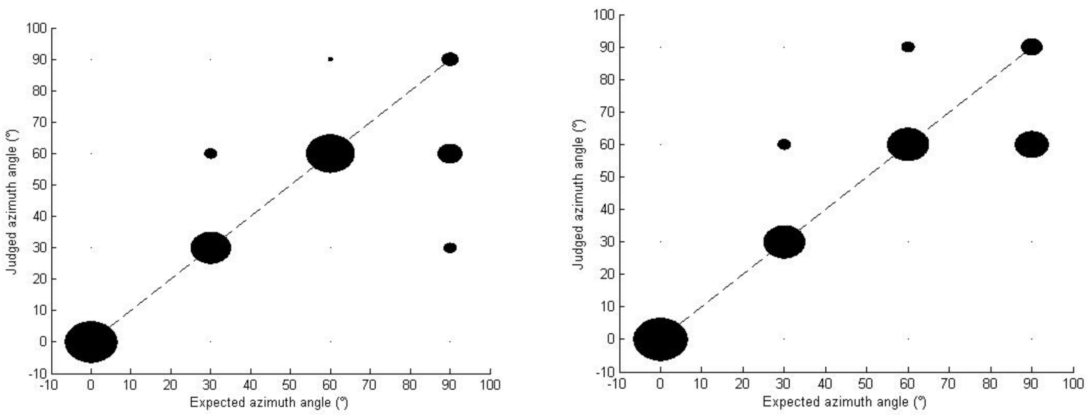
(a) (b)

Figure 11 – Difference between expected and reproduced ITD. (a) Sweet spot. (b) 5 cm displacement.

3.2 Subjective Listening Tests

Subjective evaluation experiments were also carried out in the listening room. The localization test signal is a 44-second-long speech. Virtual sound images at 4 directions on the horizontal plane with increment of 30° azimuth are generated. Thirteen young adults with normal hearing participated in the experiment. They sit at the sweet spot with no head movements.

The experimental results of the judged azimuth angles versus the target azimuth angles in the localization tests are shown in Figure 12. The area of each circle is proportional to the number of the listeners who localized the same perceived angle. The 45° dash line represents the perfect localization.



(a) (b)

Figure 12 – Azimuth localization results of the subjective test. (a) Stereo dipole. (b) Linear array.

Table 2 illustrates the evaluations of the surround sound tests, in which the test stimuli are three different surround sound periods. The mean opinion score (MOS) as shown in Table 1 is utilized. The results demonstrate that the linear array has better surround sound listening experience.

Table 1 – Mean opinion score for surround sound tests

MOS	Quality
5	Excellent with perfect surround sound feeling
4	Good with perceptible distortion
3	Fair with slightly annoying distortion
2	Poor with annoying distortion
1	Bad without any correct surround sound feeling

Table 2 – Surround sound tests (average MOS)

Signal	Stereo dipole	Linear array
1	3.9	4.3
2	4.0	4.3
3	3.9	4.5

4. CONCLUSIONS

In this paper, binaural playback system based on the OSD method is proposed. The system is in linear array form mapped from the circular OSD approach. The optimization of the source pair positions and frequency-span discretization is discussed with consideration of the crosstalk cancellation performance and the frequency response as well as the loss of the dynamic range. The experiments have demonstrated that the proposed system has better surround sound experience than the commonly used stereo dipole approach.

ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation of China No. 11374156 and No. 11104141.

REFERENCES

1. Bauer B B. Stereophonic earphones and binaural loudspeakers. J. Audio Eng. Soc., 1961, 9(2): 148-151.
2. Kyriakakis C. Fundamental and technological limitations of immersive audio systems. Proceedings of the IEEE, 1998, 86(5): 941-951.
3. Ward D B, Elko G W. Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation. Signal Processing Letters, IEEE, 1999, 6(5): 106-108.
4. Kirkeby O, Nelson P A, Hamada H. The 'Stereo Dipole': A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers. J. Audio Eng. Soc., 1998, 46(5): 387-395.
5. Bai M R, Tung C W, Lee C C. Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm. J. Acoust. Soc. Am., 2005, 117(5): 2802-2813.
6. Takeuchi T, Nelson P A. Optimal source distribution for binaural synthesis over loudspeakers. J. Acoust. Soc. Am., 2002, 112(6): 2786-2797.
7. Schuhmacher A, Hald J, Rasmussen K B, et al. Sound source reconstruction using inverse boundary element calculations. J. Acoust. Soc. Am., 2003, 113(1): 114-127.
8. Kirkeby O, Nelson P A, Orduna - Bustamante F, et al. Local sound field reproduction using digital signal processing. J. Acoust. Soc. Am., 1996, 100(3): 1584-1593.
9. Takeuchi T, Teschl M, Nelson P. Subjective evaluation of the Optimal Source Distribution system for virtual acoustic imaging. Proc. 19th Audio Engineering Society Conference, 2001.
10. Brungart D S, Rabinowitz W M. Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am., 1999, 106(3): 1465-1479.
11. Howard D M, Angus J. Acoustics and psychoacoustics. Taylor & Francis, 2009.
12. Blauert J. Communication acoustics. Springer, 2005.