PROCEEDINGS of the
23rd International Congress on Acoustics
9 to 13 September 2019 in Aachen, Germany

Investigation into Transaural System with Beamforming Using a Circular Loudspeaker Array Set at Off-center Position from the Listener.

Yu Ito[(1)], Yoichi Haneda[(2)]

[(1)]The University of Electro-Communications, Japan, i1930012@edu.cc.uec.ac.jp
[(2)]The University of Electro-Communications, Japan, haneda.yoichi@uec.ac.jp

Abstract

A transaural system based on a crosstalk canceller is effective for virtual acoustic imaging when two loudspeakers are arranged in front of the listener. However, when the two loudspeakers are arranged off-center from the listener, the crosstalk canceling performance at high frequencies drastically declines, particularly in a stereo dipole system. This occurs because, when both loudspeakers are located on the left side of the listener, the sound pressure level in the left ear is higher than that in the right ear. In this study, we first investigate the relationship between the cross-talk level and sound localization accuracy for off-center positioned stereo dipole systems. We then introduce two directivity beamformers based on a matched filter using a circular loudspeaker array to reproduce equal sound pressure levels in both ears. Herein, we consider two beams from the array to the listener's left and right ears as being equal to the left and right channel loudspeakers in the stereo dipole. To confirm the performance of the proposed method, we compared the off-center stereo dipole system and the proposed method through computer simulations. The proposed method was shown to improve the difference in sound pressure level between the left and right ears, particularly at high frequencies.

Keywords: Transaural System, Beamforming, Off-center Position

1 INTRODUCTION

A transaural system provides the listener with virtual acoustic images by matching the listener's ear signals and the signals of an auditory environment [1] [2]. This system typically uses two loudspeakers and a cross-talk canceller. A stereo dipole system, which uses two closely spaced loudspeakers, is known to be a robust approach for head rotations and slight movements [3]. An actual transaural system including a stereo dipole is effective when two loudspeakers are arranged in front of the listener; however, when loudspeakers are arranged in an off-center position from the listener, the performance of the sound localization decreases significantly [2]. In this study, we aim to investigate the cause of a performance degradation and propose a method that is effective even when loudspeakers are arranged in an off-center position.

First, we investigated the effects of the loudspeaker's position on the sound localization using a stereo dipole system through subjective experiments. We then introduced the beamforming of a circular loudspeaker array before applying a transaural system instead of two loudspeakers in a stereo dipole system to solve the problem found in the previous experiment. Finally, we confirmed the performance after applying beamforming and a transaural system through computer simulations.

2 TRANSAURAL SYSTEM AND STEREO DIPOLE

A transaural system enables the listener to perceive virtual acoustic images. Fig. 1 shows the configuration of a typical transaural system. In this figure, $S(\omega)$ is a source signal in the original sound field; $G_l(\omega)$ is a head-related transfer function (HRTF) from the source signal in the original sound field to the left ear; $P_l(\omega)$ and $P_r(\omega)$ are the listener's ear signals in the original sound field; $X_l(\omega)$ and $X_r(\omega)$ are the driving signals input into the loudspeakers in the reproduction field; $W_{\mathrm{rr}}(\omega), W_{\mathrm{rl}}(\omega), W_{\mathrm{lr}}(\omega)$, and $W_{\mathrm{ll}}(\omega)$ are the cross-talk canceller filters; $G_{\mathrm{ll}}(\omega)$ denotes the HRTF from the left speaker to the listeners left ear; $\hat{P}_{\mathrm{l}}(\omega)$ and $\hat{P}_{\mathrm{r}}(\omega)$ are the listener's ear signals in the reproduction field; and ω is the angular frequency. A transaural system usually applied widely spaced loudspeakers, typically spanning an angle of 60°, as seen by the listener. However, in the case of a stereo dipole, the two loudspeakers are set close together such that their span is only 10°. A stereo dipole is extremely robust with head movements without introducing any excessive artifacts [3]. The filters used in the transaural system are calculated using the method described below. For the case of two loudspeakers, the sound pressures $\hat{P}_l(\omega)$ and $\hat{P}_r(\omega)$ on the left and right ears in a reproducing system can be described using the following matrix forms:

$$\hat{\mathbf{P}} = \begin{pmatrix} \hat{P}_{\mathrm{l}}(\omega) \\ \hat{P}_{\mathrm{r}}(\omega) \end{pmatrix} = \begin{pmatrix} G_{\mathrm{ll}}(\omega) & G_{\mathrm{rl}}(\omega) \\ G_{\mathrm{lr}}(\omega) & G_{\mathrm{rr}}(\omega) \end{pmatrix} \begin{pmatrix} X_{\mathrm{l}}(\omega) \\ X_{\mathrm{r}}(\omega) \end{pmatrix} = \mathbf{G}(\omega)\mathbf{X}(\omega), \tag{1}$$

where

$$\mathbf{G}(\omega) = \begin{pmatrix} G_{\mathrm{ll}}(\omega) & G_{\mathrm{rl}}(\omega) \\ G_{\mathrm{lr}}(\omega) & G_{\mathrm{rr}}(\omega) \end{pmatrix}, \mathbf{X}(\omega) = \begin{pmatrix} X_{\mathrm{l}}(\omega) \\ X_{\mathrm{r}}(\omega) \end{pmatrix}.$$

Because the filter matrix $\mathbf{W}(\omega)$ needs to satisfy $\hat{P}_{\mathrm{l}} = P_{\mathrm{l}}$ and $\hat{P}_{\mathrm{r}} = P_{\mathrm{r}}$, $\mathbf{W}(\omega)$ is the inverse of the transfer function matrix of $\mathbf{G}(\omega)$.

$$\mathbf{W}(\omega) = \begin{pmatrix} W_{\mathrm{ll}}(\omega) & W_{\mathrm{rl}}(\omega) \\ W_{\mathrm{lr}}(\omega) & W_{\mathrm{rr}}(\omega) \end{pmatrix} = \frac{1}{G_{\mathrm{ll}}(\omega)G_{\mathrm{rr}}(\omega) - G_{\mathrm{rl}}(\omega)G_{\mathrm{lr}}(\omega)} \begin{pmatrix} G_{\mathrm{rr}}(\omega) & -G_{\mathrm{rl}}(\omega) \\ -G_{\mathrm{lr}}(\omega) & G_{\mathrm{ll}}(\omega) \end{pmatrix} \tag{2}$$

Because the inverse matrix of $\mathbf{G}(\omega)$ is often unstable, the inverse matrix is calculated using the least squares method in our study.

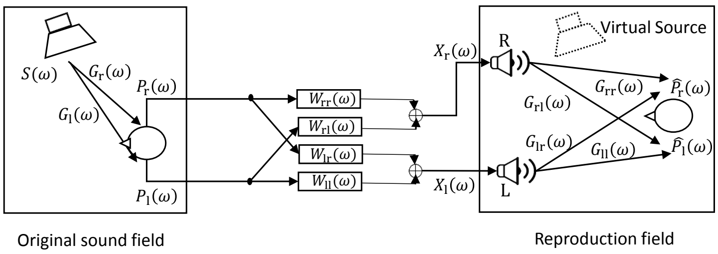

Figure 1. Configuration of transaural system

3 INVESTIGATION OF LOUDSPEAKERS PLACEMENT FOR STEREO DIPOLE

To investigate the effects of the loudspeaker arrangement on the performance of sound localization in a stereo dipole, we conducted a subjective experiment undertaken in an usual room with a reverberation time of 0.10 ms. We then compared the results of a subjective test with the cross-talk levels.

3.1 Procedure of subjective localization experiment

Fig. 2 shows the arrangement of the subjective localization experiment. The subjects sat in front of 18 loudspeakers. Nine stereo dipole systems were each prepared using two adjacent sets of loudspeakers (sets 1–9) with ten loudspeakers on the left as viewed from the listener. These two loudspeakers were 0.7 m from the subject's head, and they spanned 8° as seen by the listener, similar to that of a usual stereo dipole. Nine different subjects took part in the experiments. The virtual source direction was set to 36° and 68° (the directions of the 14th and 18th loudspeakers) to the front-right of the subject. In addition to the virtual source reproduction, we presented a normal white noise band-passed at 0.8-3 and 0.2-16 kHz from the 14th and 18th loudspeakers as references, respectively. The subjects selected their perceived sound direction among the 18 loudspeaker positions. The responses were limited to the horizontal plane at angle locations marked on the 18 loudspeakers.

Band-limited white noises band-passed at 0.8-3 and 0.2-16 kHz were used as the source signals. The frequency band was chosen to include frequencies that humans predominantly localize based on an interaural time difference (ITD) of below approximately 1.6 kHz and an interaural level difference (ILD) of above approximately 1.6 kHz)[4]. The latter band was chosen to include all frequencies sufficient for music listening. The length of the white noise was 3 s, which is sufficient for the listener to recognize the perceived direction [4].

The cross-talk canceling filters were calculated using the measured head-related impulse response (HRIR), which was applied in the same room using a sampling frequency of 48 kHz. Fig. 3 shows the conditions used in the HRIR measurements. We measured all HRIRs using a NEUMANN KU100 dummy head.

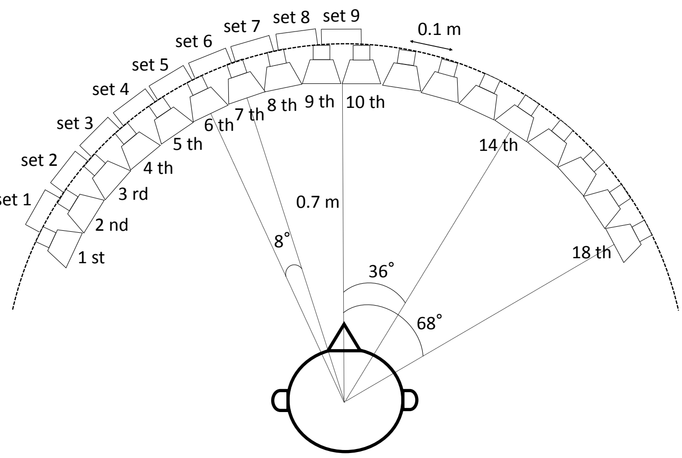

Figure 2. Arrangement of the subjective experiments

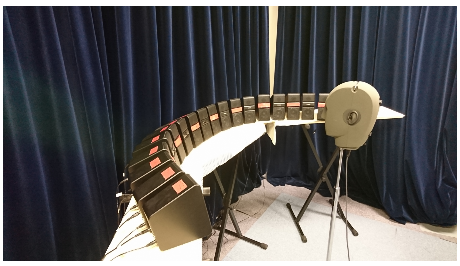

Figure 3. Measurement of HRIR for cross-talk cancellation filters

3.2 Results

Fig. 4 shows the averaged perceived direction of the subject's responses when we presented the virtual source to the listener by applying a stereo dipole with two loudspeakers among sets 1–9. The horizontal axis is the set number of loudspeakers used, and the vertical axis is the average number of answered loudspeakers. For example, when the set number of loudspeakers was 6, the transaural system used two loudspeakers placed at the 6th and 7th positions. In addition, REF indicates the results of the reference condition. The gray lines in Fig. 4 denote the position of the loudspeaker, which is the correct direction of the virtual source. Fig. 4 (a) shows the results of the virtual source direction of 36°, and Fig. 4 (b) shows the results of the virtual source direction of 68°. In these figures, we can see that the sound localization performance at a narrow band (0.8-3 kHz) was better than the nearly full band (0.2-16 kHz) at the off-center positioned stereo dipole. Moreover, if the stereo dipole is set on the left side of the listener, it becomes difficult to perceive the sound image on the right side when the angle between the actual loudspeakers and the virtual source is too wide. For set number 6 in Fig. 4 (a), the value at 0.8-3 kHz is close to the desired value, whereas the value at 0.2-16 kHz is not. This may be due to the high level of original crosstalk at high frequencies before applying the transaural system.

Fig. 5 shows the HRTF of both ears before applying the transaural system. Fig. 5 (a) denotes the HRTF from the right channel loudspeaker of set 9, and (b) indicates the HRTF from the right channel loudspeaker of set 6. Comparing (a) and (b), the amount of original crosstalk is increased in (b), particularly at high frequencies. An increase in the amount of crosstalk before applying the taransaural system may affect the results after the system is applied. This is because, if the amount of crosstalk increases, the denominator of Eq. (2) $G_{\mathrm{ll}}G_{\mathrm{rr}} - G_{\mathrm{rl}}G_{\mathrm{lr}}$ decreases, and the cross-talk cancelation filters become unstable. Fig. 6 shows the computer simulations of both ear signals after applying the transaural system at 0.2-16 kHz when using loudspeakers of (a) set 9 (center-positioned stereo dipole) and (b) set 6 (off-center-positioned stereo dipole). According to the simulation results, it was confirmed that, when two loudspeakers are arranged off-center from the listener (Fig. 6 (b)), the amount of crosstalk suppression after applying the transaural system is degenerated at high frequencies. This corresponds to the increase in the amount of the original crosstalk at high frequencies, as shown in Fig. 5 (b). As these results indicate, it is necessary to reduce the amount of crosstalk before applying the transaural system to achieve a sufficient performance within the full frequency band using off-center arranged loudspeakers, as with set 6.

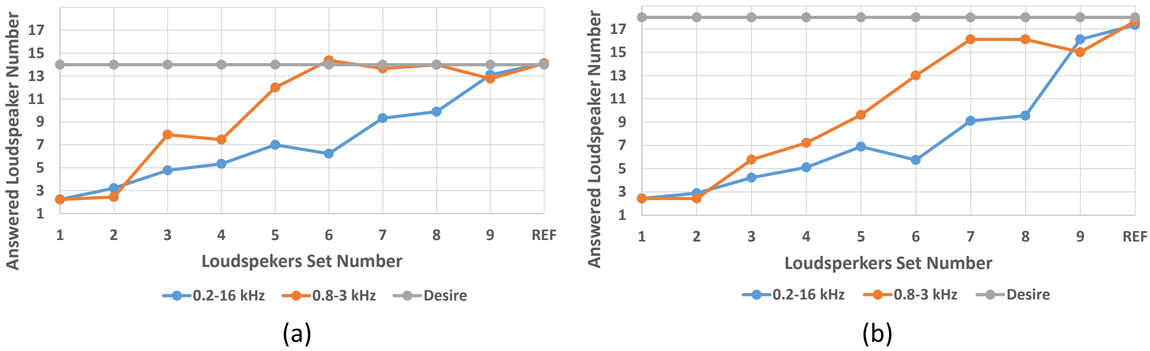

Figure 4. Results of subject responses for the virtual source localization tests. The virtual source was applied in the directions of (a)36° (14 ch) and (b) 68° (18 ch).

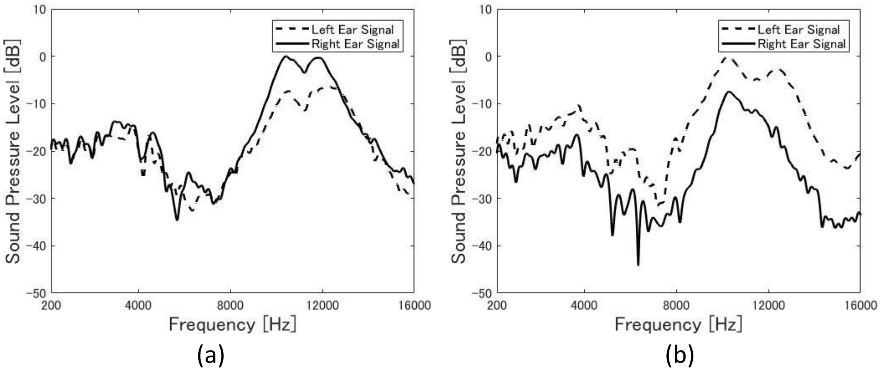

Figure 5. The HRTF from the right channel loudspeaker of (a) set 9 and (b) set 6 for both ears before applying the transaural system. The bold lines indicate the right ear signals, and the dashed lines indicate the left ear signals (cross-talk signals).

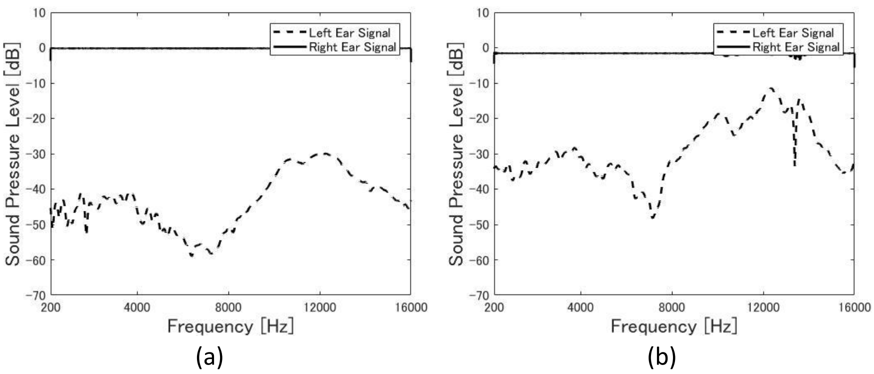

Figure 6. The computer simulations of both ear signals after applying the transaural system at 0.2-16 kHz when using loudspeakers of (a) set 9 (center-positioned stereo dipole) and (b) set 6 (off-center-positioned stereo dipole).

4 TRANSAURAL SYSTEM WITH BEAMFORMING

In the previous section, we found that it is necessary to reduce the original crosstalk before applying the transaural system for a sufficient sound localization performance. To reduce the original crosstalk and reproduce equal sound pressure levels in both ears for the off-center system, we introduce a method that forms two directivity beams using a circular loudspeaker array [5][6]. Here, we consider two beams from the array to the listener's left and right ears instead of the left- and right-channel loudspeakers in the stereo dipole. In this section, we consider the case in which the circular loudspeaker array is placed at the positions of sets 9 (center-positioned) and 6 (off-center-positioned) shown in Fig. 2.

4.1 Beamforming Using Matched Filtering

The matched filtering beamformer compensates the delay of the transfer functions between the loudspeakers and the focal point. Therefore, the filter is calculated as a complex conjugate of the transfer functions by dividing their absolute value [7][8]. The matched filter $\mathbf{q}(\omega)$ is calculated as follows:

$$\mathbf{q} = \left[\frac{G_1(\omega)^*}{|G_1(\omega)|}, \frac{G_2(\omega)^*}{|G_2(\omega)|}, \cdots, \frac{G_\mathrm{N}(\omega)^*}{|G_\mathrm{N}(\omega)|} \right]^\mathrm{T}, \tag{3}$$

where $G_n(\omega)$ denotes the transfer function from the nth loudspeaker in the array to the focal point, and $G_n^*(\omega)$ is the complex conjugate of $G_n(\omega)$. In addition, N is the number of loudspeakers in the array. This beamformer has the advantage of being easily combined with other systems because its white noise gain is constant [9].

4.2 Computer Simulations of the Ear Signals after Applying the Beamforming

We simulated the ear signals after applying a matched filtering beamformer to confirm the crosstalk level differences between the left and right ears. In this simulation, we used a circular array of 15 loudspeakers, and the matched filter was calculated at 0.2-16 kHz using the measured HRIR from each loudspeaker to both ears in the same room. Both ear signals $\mathbf{G}(\omega)$ were calculated through the following formula.

$$\mathbf{G} = \begin{pmatrix} G_\mathrm{ll}(\omega) & G_\mathrm{rl}(\omega) \\ G_\mathrm{lr}(\omega) & G_\mathrm{rr}(\omega) \end{pmatrix} = \begin{pmatrix} G_\mathrm{1l}(\omega) & G_\mathrm{2l}(\omega) & \cdots & G_\mathrm{Nl}(\omega) \\ G_\mathrm{1r}(\omega) & G_\mathrm{2r}(\omega) & \cdots & G_\mathrm{Nr}(\omega) \end{pmatrix} \begin{pmatrix} q_\mathrm{1l}(\omega) & q_\mathrm{1r}(\omega) \\ q_\mathrm{2l}(\omega) & q_\mathrm{2r}(\omega) \\ \vdots & \vdots \\ q_\mathrm{Nl}(\omega) & q_\mathrm{Nr}(\omega) \end{pmatrix} \tag{4}$$

where $G_{n\mathrm{l}}(\omega)$ denotes the measured transfer function from the loudspeaker n in the array to the left ear, and $q_{n\mathrm{l}}(\omega)$ is the matched filter focused on the left ear. Fig. 7 shows the array arrangement of this simulation. The simulation described herein was conducted using two configurations, with the loudspeaker array in front of the listener (same position as set 9 in Fig. 2) and off-center (same position as set 6). Fig. 8 shows the frequency response of both ear signals after applying the beamforming. Fig. 8 (a) denotes the case of the loudspeaker array arranged in the front, and Fig. 8 (b) denotes the case of the off-center arrangement. Compared with Fig. 5, the amount of crosstalk decreased in both arrangements (a) and (b) owing to the beamforming. Even for the arrangement of (b), the same level of sound pressure was observed in the left and right ears at high frequencies. With this method, the level difference at low frequencies is not improved because the wavelength is larger than the distance between the left and right ears, although this is not too important because sound localization is possible at low frequencies (0.8-3 kHz) even in an off-center loudspeaker arrangement without beamforming, as shown in Fig. 4 (a).

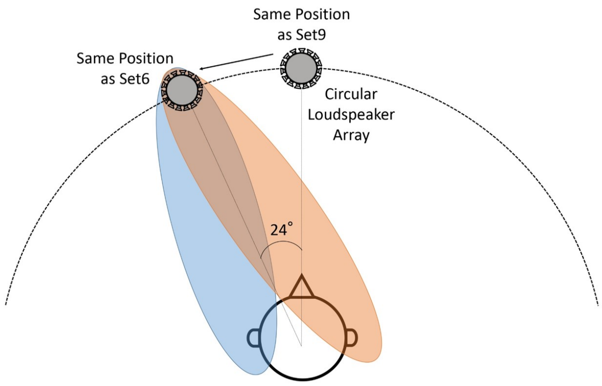

Figure 7. Arrangement of circular loudspeaker array in the computer simulations

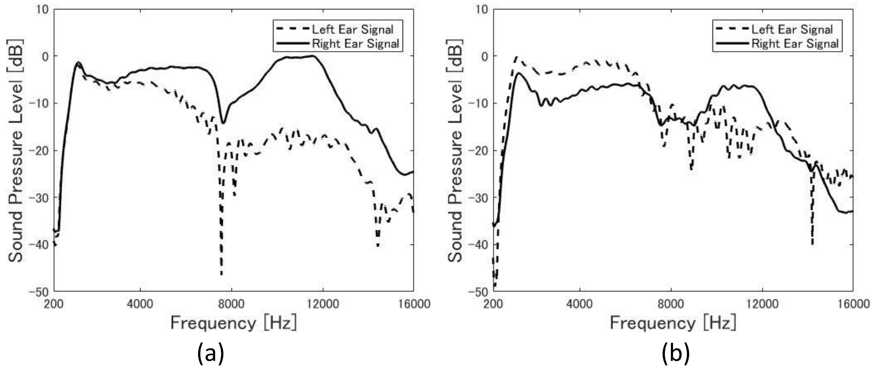

Figure 8. Computer simulations of both ear signals after applying beamforming using a loudspeaker array arranged in the positions of (a) set 9 (center-positioned) and (b) set 6 (off-center positioned).

4.3 Computer Simulations of Ear Signals After Applying the Beamforming and Transaural System

Finally, we simulated both ear signals after applying the beamforming and transaural system. Fig. 9 (a) shows the ear signals in the case of a center arrangement, and Fig. 9 (b) shows the signals for an off-center arrangement. The figures indicate that the amount of cross-talk is sufficiently suppressed in both (a) and (b), as compared with Fig. 6, particularly at high frequencies. Hence, it is expected that sound localization in the full frequency band is realized using this method.

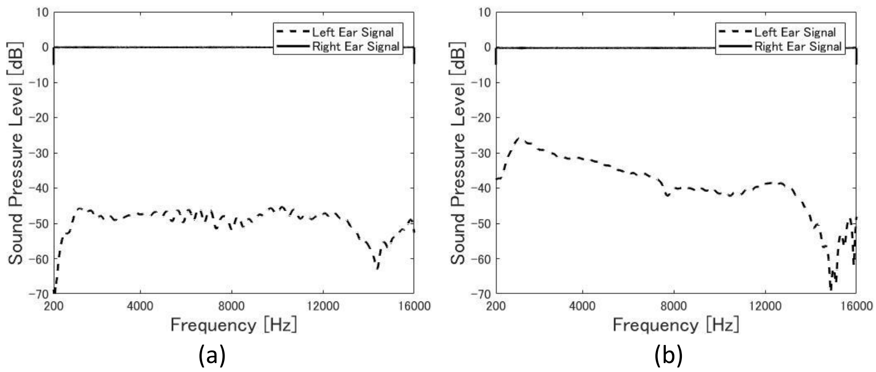

Figure 9. Computer simulations of both ear signals after applying beamforming and the transaural system at loudspeaker array positions of (a) set 9 (center-positioned) and (b) set 6 (off-center positioned).

5 CONCLUSIONS

We confirmed through a subjective experiment that a stereo dipole system using an off-center loudspeaker arrangement does not achieve a sufficient performance for sound localization, particularly at high frequencies. This is thought to be due to an increase in the amount of original crosstalk at high frequencies before applying the transaural system. Therefore, we introduced a method that forms directivity beams using a circular loudspeaker array to reduce the cross-talk before applying the transaural system. As a result, the amount of original crosstalk at high frequencies was reduced using beamforming. Moreover, the amount of crosstalk after applying a cross-talk cancelation filter was suppressed. Our future investigation will be to examine the performance of this method through a subjective experiment in a room with typical reverberation.

REFERENCES

[1] M. R. Schroeder, B. S. Atal, "Computer simulation of sound trancemission in rooms, " *IEEE Int. Conv. Rec*, part 7, pp.150-155,1963.

[2] J. Rose,P. A. Nelson,B. Rafaely,T. Takeuchi, "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations, " *J. Acoust. Soc. Am*,112 (5), Pt. 1, November. 2002.

[3] O. Kirkeby,P. A. Nelson, H. Hamada, "The "Stereo Dipole" - A Virtual Source Imaging System Using Two Closely Spaced Loudspeakers, " *J. Audio Eng. Soc.* , 46 (5), May 1998.

[4] J. Blauert, *Spatial Hearing, The psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA, 1997).

[5] K. Sato, Y. Haneda, "Directivity control of a compact circular loudspeaker array based on selected orders of circular harmonics expansion, " *J. Acoust. Soc. Am*,140 (4), Pt. 2-2, 2aSPb31, October. 2016.

[6] K. Sato, Y. Haneda, "Sidelobe suppression by desired directivity pattern optimization for a small circular loudspeaker array, " *Acoust. Sci. and Tech.* 39, (3), p.243-251, December 2018.

[7] S. Yon, M. Tanter, and M. Fink, "Sound focusing in rooms: The time-reversal approach, " *J. Acoust. Soc. Am. 113 (3), March 2003*

[8] H. Sakamoto, Y. Haneda, "Sound Localization of Beamforming-Controlled Reflected Sound from Ceiling in Presence of Direct Sound, " *Audio Eng. Soc. conv 144th,* Milan, 2018.

[9] M. Guldenschuh, A. Sontacchi, "TRANCEAURAL STEREO IN A BEAMFORMING APPROACH, " *Proc. conf. DAFx-09 12th, Como, Italy, September 1-4, 2009.*