Loudspeaker distributions suitable for crosstalk cancellers robust to head rotation

Cheolsu Han1,2,*, Takuma Okamoto1,3,†, Yukio Iwaya1,2,‡ and Yoši Suzuki1,2,§

1Research Institute of Electrical Communication, Tohoku University, 2–1–1, Katahira, Aoba-ku, Sendai, 980–8577 Japan
2Graduate School of Information Sciences, Tohoku University, 6–3–09, Aramaki Aza Aoba, Aoba-ku, Sendai, 980–8577 Japan
3Graduate School of Engineering, Tohoku University, 2–1–1, Katahira, Aoba-ku, Sendai, 980–8577 Japan

(Received 2 December 2011, Accepted for publication 7 February 2012)

Keywords: Robust crosstalk canceller, Transaural sound reproduction, Loudspeaker distribution, Head rotation

PACS number: 43.60.Pt [doi:10.1250/ast.33.266]

1. Introduction

We can localize sound sources by listening with only two ears because of the various localization cues of binaural sound. This phenomenon is applied to a type of auditory display, which reproduces/synthesizes three-dimensional (3-D) spatial sound. This type of system, sometimes called virtual auditory display (VAD), reproduces/synthesizes the sound pressure levels at both ears of the listener [1,2]. While headphones are often used as the sound output devices of VADs to simplify the system architecture, they can be cumbersome and uncomfortable [3]. However, when loudspeakers are used for VADs, sounds from all loudspeakers reach both ears of the listener, which causes an undesirable phenomenon known as crosstalk. The crosstalk issue must be carefully considered in designing VADs. Crosstalk is usually eliminated using a set of linear inverse filters, known as a crosstalk canceller [4]. The most common crosstalk canceller designs assume that the listener is located precisely at a specified position and orientation relative to the loudspeakers. This assumption, depending on the configuration, may lead to crosstalk cancellers that are sensitive to changes in the position of the listener [3,5]. In real world scenarios, head movement of the listener is not restricted, and previous studies have shown it to be very important to improve the precision of the perceived auditory space [6–8]. Therefore, a crosstalk canceller should be designed to be as robust as possible to the head rotation of the listener. For this reason, identification of loudspeaker arrangements that result in crosstalk cancellers robust to head rotation is very important for the design and implementation of VADs.

The objective of the present study was to identify loudspeaker distributions resulting in crosstalk cancellers robust to head rotation. Previous studies have examined symmetric loudspeaker distributions with different loudspeaker spans and elevation angles to improve the robustness of crosstalk cancellers to small changes in the transfer functions from the loudspeakers to the listener’s ears [9,10]. They have assessed the robustness by evaluating the condition numbers of the matrices consisting of theoretically modeled transfer functions for different loudspeaker distributions. They argued that a crosstalk canceller robust to small changes in the transfer functions can be obtained when the matrix has a low condition number. If the changes in the transfer functions caused by head rotation were the same for the various loudspeaker distributions, the optimal loudspeaker arrangements for the crosstalk canceller robust to head rotation could be properly evaluated from the condition numbers alone. Nevertheless, the changes induced by head rotation are highly dependent on the loudspeaker distributions even if the condition number is the same. Therefore, it is also very important to evaluate the output binaural signals resulting from filtering the input signals with both, the designed crosstalk cancellers and the new transfer functions after head rotation. Thus, in this study, we conducted a series of computer simulations to evaluate the resulting binaural signals of crosstalk cancellers for various loudspeaker configurations. We considered 52,650 different two-channel loudspeaker arrangements, including asymmetric cases. Finally, we evaluated the robustness of the inverse filters, derived from free-field impulse responses measured in an anechoic chamber, for crosstalk cancellation.

2. Crosstalk canceller design in the time domain

We calculated the crosstalk cancellation filters using the time-domain least-squares method [11]. This method results in a crosstalk canceller consisting of a set of causal and stable finite impulse response (FIR) filters. Figure 1 shows a block diagram of a typical two-channel crosstalk canceller. In Fig. 1, vectors \( \mathbf{x}_L \) and \( \mathbf{x}_R \) denote the left and right input signals, respectively. Vectors \( \mathbf{y}_L \) and \( \mathbf{y}_R \) represent the resulting binaural signals at the left and right ears, respectively. The crosstalk cancellation filters should be designed to reproduce the left and right input signals at the corresponding ear. Such a system can be described in matrix form as follows:

\[
\begin{bmatrix}
  d_m \\
  0
\end{bmatrix}
= 
\begin{bmatrix}
  H_{1L} & H_{2L} \\
  H_{1R} & H_{2R}
\end{bmatrix}
\begin{bmatrix}
  c_{L1} \\
  c_{R1}
\end{bmatrix}
+ 
\begin{bmatrix}
  d_m \\
  0
\end{bmatrix}
\]  
(1)

* e-mail: cshan@ais.riec.tohoku.ac.jp
† e-mail: okamoto@ais.riec.tohoku.ac.jp
‡ e-mail: iwaya@riec.tohoku.ac.jp
§ e-mail: yoh@riec.tohoku.ac.jp
HiDi

speaker distributions were evaluated in this study.

two loudspeakers is as follows:

\[
\begin{bmatrix}
    h_{ID}(0) & 0 & 0 \\
    \vdots & h_{ID}(0) & \vdots \\
    h_{ID}(N_h - 1) & \vdots & 0 \\
    0 & h_{ID}(N_h - 1) & h_{ID}(0) \\
    0 & 0 & h_{ID}(N_h - 1)
\end{bmatrix},
\]

where vectors \( h_{ID} \) are the \( N_h \)-length impulse responses from the loudspeakers to the left and right ears for \( i = 1, 2 \), and \( D = L, R \). Vectors \( e_{Di} \) are the crosstalk cancellation filters of length \( N_s \). Vector \( d_{ai} \) is the unit impulse signal delayed by \( m \) samples, \( \delta(t - m) \), of length \( N_s + N_c - 1 \). The delay \( m \) represents the modeling delay of the system. Equation (1) can be written in compact form as follows:

\[
D = HC.
\]  

Therefore, crosstalk cancellation filters can be calculated by inverting the matrix \( H \) as follows:

\[
C = (H^T H)^{-1} H^T D.
\]  

3. Computer simulation

A series of computer simulations to evaluate the error in crosstalk cancellation for various loudspeaker distributions was conducted to evaluate the robustness to changes in the head orientation of the listener.

3.1. Simulation conditions

We assumed that the loudspeakers were located at the positions described below. Sampled loudspeaker elevation angles, measured from the horizontal plane upward, ranged from 0 degrees (the horizontal plane) to 90 degrees (directly overhead) in multiples of 10 degrees. Loudspeaker azimuth angles, measured from the front and increasing in the leftward (anticlockwise) direction, lay between 0 and 350 degrees. The azimuth angles corresponding to elevation angles between 0 and 80 degrees were sampled at multiples of 10 degrees. Meanwhile the elevation angle of 90 degrees was only sampled at one point (directly overhead). Therefore, the number of possible loudspeaker locations, \( N_{\text{spk}} \), is \( 36 \times 9 + 1 = 325 \), and the number of possible distributions of the two loudspeakers is as follows:

\[
N_{\text{dis}} = \binom{N_{\text{spk}}}{2}.
\]

Thus, crosstalk cancellers for \( N_{\text{dis}} = 52,650 \) different loudspeaker distributions were evaluated in this study.

We selected \( h_{ID} \) as the head-related impulse responses (HRIR) from the loudspeakers to the left and right ears of the SAMRAI dummy head (Koken Co., Ltd.), designed to approximate the average Japanese male head. The HRIR was measured in an anechoic room of the Research Institute of Electrical Communication, Tohoku University. Loudspeakers used to measure the HRIR were located 1.5 m away from the center of the head.

3.2. Simulation method

First, crosstalk cancellers were designed for the various loudspeaker distributions using the time-domain method. This method, described in the previous section, requires several parameters. In this study, the length of the impulse response \( (N_h) \) and the length of the inverse filter \( (N_c) \) were set to 512 samples and 1,024 taps, respectively. The sampling frequency was 48 kHz, and the modeling delay \( m \) was 400 taps.

Next, the signal generated at each ear was calculated by the sequential filtering of the input signals using the crosstalk canceller and the HRIR. As input signals, an impulse for the left ear and a null signal (silence) for the right ear were employed. Moreover, these input signals were filtered using a low-pass filter (LPF) with a cutoff frequency of 20 kHz when designing the crosstalk cancellers. This signal combination is thought to be sufficient for the present purpose because human heads can be regarded as being left-right symmetrical. To judge the robustness of the crosstalk canceller to changes in the listener’s head orientation, the errors in the crosstalk cancellation were computed under the assumption that the listener would rotate his/her head by 5 degrees away from the ideal orientation to the left and right. Therefore, the HRIRs were selected by assuming said displacement.

The distortion power (DP) was calculated at each ear by comparing the desired output signal \( s(n) \), equal to the \( m \)-sample-delayed ideal input signal, with the actual output signal \( \hat{s}(n) \) as follows:

\[
DP = \sum |s(n) - \hat{s}(n)|^2.
\]

In this study, the DP was used as the metric of the robustness of crosstalk cancellers.

Finally, the total DP was calculated for each loudspeaker distribution as follows:

\[
DP_t = \sqrt{|DP_{\|}|^2 + |DP_{\|t}|^2 + |DP_{\perp}|^2 + |DP_{\perp t}|^2},
\]

where \( DP_{\|} \) and \( DP_{\perp} \) are, respectively, the DP values of the left and right channels, assuming that the listener’s head would rotate five degrees to the left, away from the ideal orientation; meanwhile, \( DP_{\|t} \) and \( DP_{\perp t} \) are, respectively, the DP values of the left and right channels, assuming that the listener would rotate his/her head five degrees to the right. The loudspeaker arrangement with the smallest \( DP_t \) was regarded in this study as the best to implement the most robust crosstalk canceller in the presence of head rotation.

4. Results and discussion

We evaluated robustness against head rotation of 52,650 loudspeaker distributions based on the total DP (\( DP_t \)) of Eq. (7). Figure 2 shows the plot of calculated \( DP_t \) values. The magnitudes of \( DP_t \) for the 52,649 loudspeaker arrangements...
are shown as a function of the ranking. The DP values could not be calculated for one of the 52,650 loudspeaker distributions within our simulation conditions.

A part of the ranking of the examined loudspeaker distributions shown in Fig. 2 is listed in Table 1. When the results were analyzed according to the elevation angles of the loudspeakers, those in the upper part of the sphere (50–80 deg.) were found to result in more robust crosstalk cancellers than other elevation angles. This result can be explained from the observation that the impulse responses \( h_{iD} \) for the loudspeakers positioned at high elevations were generally less susceptible to head rotation than those for loudspeakers positioned near the horizontal plane. This result is consistent with that reported by Takeuchi et al. [12]. They conducted localization experiments using a multi-channel system with two different loudspeaker distributions and found that distributing loudspeakers above the listener increases robustness to head rotation. Moreover, robustness to head rotation is further improved if the two loudspeakers are positioned at azimuth angles close to those of the listener’s ears (100–110 deg. and 210–260 deg.).

Figure 3 shows the relative amplitudes of the resulting impulse signals at the left ear, while Fig. 4 shows the relative amplitudes of the resulting null signals at the right ear. The resulting signals shown in the left and right side panels, in the both figures, are for the cases that the listener rotated his/her head five degrees to the left and right, respectively. The resulting signals from the top panel in order show the resulting signals for the 1st, 10th, 100th, 1,000th, 10,000th, and 50,000th ranked loudspeaker distributions, respectively. When the head orientation was not exact, the resulting signals for the higher ranked loudspeaker arrangements were found to be more similar to the ideal impulse and null signals than those for the lower ranked loudspeaker configurations. Therefore, crosstalk cancellers that use the higher ranked loudspeaker arrangements are regarded to be more robust to head rotation than those designed for other loudspeaker configurations.

5. Conclusions

Crosstalk cancellers are usually designed to precisely present arbitrary binaural sounds to a listener’s two ears, assuming a fixed head orientation relative to the loudspeakers. This constraint can result in a large deviation from the desired sound caused by even a small rotation of the listener’s head. Therefore, in this study, we investigated loudspeaker distributions resulting in crosstalk cancellers robust to changes in the listener’s head orientation via computer simulations. The best loudspeaker arrangement according to our results corresponded to one loudspeaker placed at an elevation angle of 50 degrees and an azimuth angle of 100 degrees, and the second loudspeaker situated at an elevation of 80 degrees and azimuth of 220 degrees. Furthermore, we found that positioning the loudspeakers above the horizontal plane generally resulted in crosstalk cancellation systems that are less sensitive to head rotation.
Acknowledgment

Parts of this research were supported by a Grant-in-Aid for Specially Promoted Research (no. 19001004) from MEXT, a Grant-in-Aid for Scientific Research (B) (no. 23300043) from JSPS, and the Tohoku University GCOE program CERIES.

References