OPTIMUM LOUDSPEAKER ARRANGEMENT OF DYNAMIC CROSSTALK CANCELLATION SYSTEM FOR TWO LISTENERS

Ryo Matsuda  
Graduate School of Engineering, Kyoto University, Kyoto, Japan  
email: matsuda.ryo.45s@st.kyoto-u.ac.jp

Makoto Otani  
Graduate School of Engineering, Kyoto University, Kyoto, Japan

Hiraku Okumura  
Graduate School of Engineering, Kyoto University, Kyoto, Japan  
Yamaha Corporation, Hamamatsu, Japan

Crosstalk cancellation (CTC) allows binaural reproduction through loudspeakers using inverse filter so called CTC filter which cancels transfer functions between loudspeakers and listener’s ears. A CTC system for single listener has been studied from various viewpoints, however CTC systems for multiple listeners has not been studied much yet. The authors have developed dynamic CTC system for multiple listeners and found that the sound localization performance greatly differs depending on the position of the listener relative to loudspeakers. Therefore, for development of multi-listener CTC system, it is necessary to clarify the influence of existence of other listeners and loudspeaker arrangement in dynamic CTC system. In this paper, we investigate the optimum loudspeaker arrangement of dynamic CTC system for two listeners by employing two dummy head microphones and 122ch loudspeaker array.

Keywords: Binaural reproduction, Crosstalk cancellation, Multi-listener

1. Introduction

Three-dimensional sound can be presented to a listener by reproducing sound pressures at both ears of a listener, so called binaural reproduction [1]. Binaural signals are obtainable by binaural recording using dummy head microphones or binaural synthesis based on head-related transfer functions (HRTFs). HRTFs are acoustic transfer functions from a sound source to the ear drums of a listener, which include interaural time difference (ITD), interaural level difference (ILD) and spectral cues.

Generally, binaural signals are presented to a listener using a set of headphones. On the other hand, crosstalk cancellation (CTC) [2][3][4] using loudspeakers is an alternative to realize binaural reproduction and it reduce fatigues of a listener, cause by wearing headphones. The CTC system requires inverse filter (CTC filter) to eliminate acoustic transmissions from loudspeakers to both ears, which destroy intended spatial perception. Crosstalk cancellation is considered to be sound pressure control when using more than two loudspeakers or for multiple listeners. The filter coefficients for CTC are determined by acoustic transfer functions between each loudspeaker and each ear, therefore
conventional static CTC has a disadvantage that it requires a listener to be still. It is seriously problematic because head movement plays an important role in sound localization [5][6].

The authors have developed a dynamic CTC system [7][8] to solve this problem using non-contact head tracking and demonstrated that dynamic CTC system outperforms static CTC in the aspect of sound localization [9][10][11]. On the other hand, we realized that CTC filter does not work perfectly because of system latency caused by head tracking and real-time convolution. Due to the system delay, there was a difference between the head position/orientation when designing the inverse filter and the actual head position/orientation. However, practically, system latency and head-tracking error are inevitable, therefore we have to optimize CTC system considering these problems.

Previous studies have examined the influence of loudspeaker arrangement on robustness using sweet spot size [12][13][14]. However, in this paper, we aim to choose an optimal arrangement of loudspeakers for two listeners which is robust to displacement of head position/rotation. Furthermore, a spherical loudspeaker array is employed in order to enable a more flexible selection of loudspeaker arrangements than previous studies, which is also useful for an actual implementation of multi-channel dynamic CTC system.

This paper is composed as follows: Section 2 presents a design of CTC filter for two listeners in frequency domain. Section 3 describes experimental conditions. In Section 4, we present and discuss the results of experiment. Section 5 summarizes and concludes this paper.

2. Crosstalk cancellation design

A schematic of the CTC system with $N$ loudspeakers is shown in Fig. 1. Assuming that the signals reproduced at both ears of the two listeners are equal to the input signals with delays of $\tau$, frequency domain representations of the reproduced signals, $Z_{1L}(\omega)$ and $Z_{1R}(\omega)$ for listener 1, $Z_{2L}(\omega)$ and $Z_{2R}(\omega)$ for listener 2, are expressed as

$$Z_{1L}(\omega) = X_{1L}(\omega)D_m(\omega),$$
$$Z_{1R}(\omega) = X_{1R}(\omega)D_m(\omega),$$
$$Z_{2L}(\omega) = X_{2L}(\omega)D_m(\omega),$$
$$Z_{2R}(\omega) = X_{2R}(\omega)D_m(\omega),$$

where $\omega$ is the angular frequency. $X_{1L}(\omega)$, $X_{1R}(\omega)$, $X_{2L}(\omega)$ and $X_{2R}(\omega)$ are the input signals to each channel, namely binaural signals to be reproduced at both ears of both listeners; $D_m(\omega)$ is a transfer function corresponding to an $m$-sample delay with a sampling frequency of $f_s$ ($m = \tau f_s$) in a discrete time system. Then, considering the total system, the following matrix expression is derived,

$$Z = GHX.$$  

The coefficient matrix $H$ for the CTC filters that satisfies Eqs. (1) and (2) is calculated as a least-norm-solution in the frequency domain [15] as

$$H = G^*(GG^*)^{-1}D$$

where

$$G = \begin{bmatrix} G_{11R}(\omega) & \ldots & G_{N1R}(\omega) \\ G_{11L}(\omega) & \ldots & G_{N1L}(\omega) \\ G_{12R}(\omega) & \ldots & G_{N2R}(\omega) \\ G_{12L}(\omega) & \ldots & G_{N2L}(\omega) \end{bmatrix}.$$
\[
D = \begin{bmatrix}
D_m(\omega) & 0 & 0 & 0 \\
0 & D_m(\omega) & 0 & 0 \\
0 & 0 & D_m(\omega) & 0 \\
0 & 0 & 0 & D_m(\omega)
\end{bmatrix}, \tag{5}
\]

\[
H = \begin{bmatrix}
H_{1R1}(\omega) & H_{1L1}(\omega) & H_{2R1}(\omega) & H_{2L1}(\omega) \\
\vdots & \vdots & \vdots & \vdots \\
H_{1RN}(\omega) & H_{1LN}(\omega) & H_{2RN}(\omega) & H_{2LN}(\omega)
\end{bmatrix}, \tag{6}
\]

where \( S_i \ (i = 1, \ldots, N; \text{from the listener’s right to left}) \) are sound sources; \( H_{jL_i}(\omega) \) and \( H_{jR_i}(\omega) \) are coefficients of the CTC filters for listener \( j \) which are inserted between \( X_{jL}(\omega) \) and \( X_{jR}(\omega) \) and \( S_i \); \( G_{ijL}(\omega) \) and \( G_{ijR}(\omega) \) are the transfer functions between \( S_i \) and both ears for listener \( j \). Generally, the inverse filter \( H \) can be unstable. Therefore, a stable inverse filter is obtainable as

\[
H = G^*(GG^* + \beta I)^{-1}D \tag{7}
\]

where \( \beta \) is the regularization parameter [16] and \( I \) is the \( 4 \times 4 \) unit matrix.

3. **Acoustical measurements**

3.1 **Measurements setup**

The acoustical measurements were performed with a spherical loudspeaker array (ViReal Dome [17]) at Yamaha Toyooka factory. A radius of the spherical array is 2.5 m. Acoustical transfer functions for generating CTC filter were measured using two sets of dummy head microphones (4128C, Brüel & Kjær). The measurement setup is shown in Fig. 2 and loudspeaker arrangements on the spherical loudspeaker array in Fig. 3. As shown in Fig. 3, One dummy head microphones (dummy head 1) was placed at the center of the loudspeaker array and another (dummy head 2) is located at 0.9 m to the right of dummy head 2. The loudspeakers placed on the floor were not used for this measurement. Therefore, we used 91 loudspeakers from the spherical loudspeaker array. The number of
loudbspeakers used in the measurement to calculate CTC is 4, therefore total 2,672,670 combinations of loudbspeakers were examined in total.

In order to measure the transfer function from each loudbspeaker to both ears of both dummy heads, time stretched pulse (TSP) of 131,072 pt was used with a sampling frequency of 48 kHz. Facing angle of dummy head 1 was between −75 and 75 degrees at 5-degree interval in a counterclockwise direction. On the other hand, the orientation of dummy head 2 is fixed. The transfer function was measured for each head direction to verify the influence on CTC filter due to head position/orientation error.

CTC filters were calculated from measured transfer function by Eq. 7. A band-pass filter was applied to CTC filters because the room used for measurement is not completely anechoic. The lower and upper frequency limits are 200 Hz and 16 kHz, respectively. In Eq. 7, larger values of regularization parameter $\beta$ suppress the gain of CTC filters, thereby increasing a stability of the system. However, at the same time, larger values of $\beta$ deteriorate the accuracy of CTC filters. Therefore, in this paper, we set the regularization parameter $\beta$ to 0 in order to exclude its effect.
3.2 Sweet spot size

Sweet spot size was employed as a parameter to evaluate robustness of CTC filter against misalignment of head movement/orientation. Sweet spot size is suitable to evaluate performance of CTC filter under misaligned head condition because it depends on channel separation. In order to calculate both channel separation and sweet spot size, frequencies from 20 Hz to 20 kHz were divided on a logarithmic scale and frequency bins from 200 Hz to 16 kHz were used.

We define plant matrix $R$ as the product of transfer function and CTC filter,

$$ R = GH = \begin{bmatrix} R_{11} & R_{12} & R_{13} & R_{14} \\ R_{21} & R_{22} & R_{23} & R_{24} \\ R_{31} & R_{32} & R_{33} & R_{34} \\ R_{41} & R_{42} & R_{43} & R_{44} \end{bmatrix}. $$

If there is no misalignment of head position/orientation, namely in the best case, $R$ is an identity matrix. The channel separation, which is the ratio of desired and undesired responses generated by CTC filter, is defined as follows:

$$ CHSP_{1R}(k) = \frac{R_{12}(k) + R_{13}(k) + R_{14}(k)}{R_{11}(k)}, $$

$$ CHSP_{1L}(k) = \frac{R_{21}(k) + R_{23}(k) + R_{24}(k)}{R_{22}(k)}, $$

and the average channel separation is defined as

$$ CHSP_i = \frac{1}{n_f - n_j + 1} \sum_{k=n_j}^{n_f} 20 \log_{10}(|CHSP_i(k)|), $$

where $k$ is the frequency bin index; $i$ is either $1L$, $1R$, $2L$ or $2R$, and $n_f$ and $n_j$ are frequency indices of the upper and lower limits, respectively.

The sweet spot size is defined as absolute sweet spot size in previous studies [12], i.e. two times the maximum one side displacement angle from the reference direction/position that results in a channel separation index below $-12$ dB when assuming symmetry with respect to the median plane. In this paper, we cannot assume median-plane symmetry because two dummy heads are employed; therefore, we redefine sweet spot size as addition of both side displacement angles. The reference head direction of dummy head 1 is set to 0 deg.

Figure 4 illustrates an example how to determine the sweet spot size is shown. In Fig. 4, the red and blue lines represent average channel separation for two loudspeaker combinations and the dash-dotted line represents $-12$-dB channel separation. In this case, the blue line shows 20 deg. as the maximum left side displacement from the reference direction/position and 10 deg. as the maximum right side displacement. Therefore, the sweet spot size for this loudspeaker pair with the blue line is estimated to be 30 deg. In the same way, the sweet spot size for the loudspeaker combination by the red line is estimated to be 10 deg. Table 1 summarizes the experimental conditions.
Table 1: Summary of experimental conditions.

<table>
<thead>
<tr>
<th>Transfer Function Measurement</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling rate</td>
<td>48 kHz</td>
</tr>
<tr>
<td>TSP length</td>
<td>131,072 pt</td>
</tr>
<tr>
<td>Head direction</td>
<td>-75 - 75 deg. (5-degree intervals)</td>
</tr>
<tr>
<td>The number of loudspeakers</td>
<td>91</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CTC Filter Design</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of listeners</td>
<td>2</td>
</tr>
<tr>
<td>The number of loudspeakers</td>
<td>4</td>
</tr>
<tr>
<td>The number of loudspeaker combinations</td>
<td>2,672,670 ($\binom{91}{4}$)</td>
</tr>
</tbody>
</table>

4. Results and Discussions

Table 2 shows the largest sweet spot size for the left, right and both ears respectively. The left and right terms in parenthesis in Table 2 demonstratate the horizontal angle and the elevation angle respectively. The horizontal angle is indicated in a counterclockwise direction from $-180$ deg. to $180$ deg. Sweet spot size for both ears are calculated from the sum of sweet spot size for left and right.

The maximum sweet spot sizes are 30, 40, and 60 deg. respectively for the left ear, right ear, and both ears. For the cases of single ear, the maximum sweet spot sizes are observed with the loudspeaker arrangement that is asymmetric with respect to the median plane. Furthermore, there is a difference in the maximum sweet spot sizes between the left and right ears, i.e. 30 and 40 deg. Such difference might be attributable to acoustical influences due to the existence of another listener (dummy head 2).

Considering sweet spot size for both ears, it would be preferable to locate loudspeakers symmetrically with respect to the median plane. In fact, one of the best loudspeaker arrangements for the case of both ears is median-plane symmetric.
Most of loudspeaker arrangements indicated in Table 2 include laterally positioned loudspeakers, namely, horizontal angles of $-108$, $-90$, $-85$, $-72$, $72$, $85$, $90$, and $108$ deg. This result likely reflects that laterally positioned loudspeakers contribute to natural channel separation between both ears due to head shadow effect.

![Table 2: Best loudspeaker combination.](image)

<table>
<thead>
<tr>
<th>sweet spot size</th>
<th>loudspeaker 1</th>
<th>loudspeaker 2</th>
<th>loudspeaker 3</th>
<th>loudspeaker 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>30 [deg.]</td>
<td>(0, 90)</td>
<td>(108, 72)</td>
<td>(108, 53)</td>
<td>(90, 0)</td>
</tr>
<tr>
<td>30 [deg.]</td>
<td>(36, 72)</td>
<td>(72, 58)</td>
<td>(72, 27)</td>
<td>(90, 0)</td>
</tr>
<tr>
<td>40 [deg.]</td>
<td>(0, 90)</td>
<td>(85, 40)</td>
<td>(90, 20)</td>
<td>(90, 0)</td>
</tr>
<tr>
<td>40 [deg.]</td>
<td>(108, 72)</td>
<td>(108, 53)</td>
<td>(90, 0)</td>
<td>(90, 0)</td>
</tr>
<tr>
<td>50 [deg.]</td>
<td>(72, 58)</td>
<td>(90, 20)</td>
<td>(90, 20)</td>
<td>(90, 0)</td>
</tr>
</tbody>
</table>

In this experiment, we employed the measured impulse response to calculate CTC filter and simulation. However, it is not practical to prepare impulse response database that takes the existence of other listeners into account. Therefore, it would be interesting to investigate how the use of pre-determined impulse responses deteriorate the CTC filters compared to the ones generated from the measured impulse responses. Other future works would include an investigation of effects of the number of loudspeakers, which is larger than four, in the robustness of dynamic CTC.

Acknowledgment

This work is partly supported by Grant-in-Aid from JSPS, Japan (No. 16H02857 and 17KT0137)

REFERENCES


