LETTER

Performance Enhancement of Cross-Talk Canceller for Four-Speaker System by Selective Speaker Operation

Su-Jin CHOI†, Jeong-Yong BOO†, Ki-Jun KIM†, Nonmembers, and Hochong PARK*†, Member

SUMMARY We propose a method of enhancing the performance of a cross-talk canceller for a four-speaker system with respect to sweet spot size and ringing effect. For the large sweet spot of a cross-talk canceller, the speaker layout needs to be symmetrical to the listener’s position. In addition, a ringing effect of the cross-talk canceller is reduced when many speakers are located close to each other. Based on these properties, the proposed method first selects the two speakers in a four-speaker system that are most symmetrical to the target listener’s position and then adds the remaining speakers between these two to the final selection. By operating only these selected speakers, the proposed method enlarges the sweet spot size and reduces the ringing effect. We conducted objective and subjective evaluations and verified that the proposed method improves the performance of the cross-talk canceller compared to the conventional method.

key words: cross-talk canceller, four-speaker system, ringing effect, sweet spot

1. Introduction

When a stereo sound is played from loudspeakers, the sound transmitted to the listener varies with listening position, which yields different sound perceptions in different listening positions. In order to transmit the intended sound to the listener, regardless of the listening position, a cross-talk canceller (CTC) is widely used [1]–[5]. A CTC for stereo sound cancels any undesired cross-talk in a given target listening position, so that each ear of the listener in the target position receives only the signal of each channel, similar to the way sound is heard when using a stereo headphone. Therefore, CTC can provide the intended stereo sound perception to the listener in any position. For example, in a teleconference system, CTC perceptually places the speech signal at the same position as the visual representation.

In an actual listening situation, however, a listener in the target position of CTC may not hear the desired sound because his/her head moves around frequently and the actual listening position deviates from the intended target position. The region around the target position where the listener can hear sound of an acceptable quality is defined as a sweet spot [2]. Therefore, for the practical usage of CTC, the sweet spot needs to be large enough to eliminate the effect of minute head movements. In addition, due to incorrect modeling of the acoustic path from the speaker to the ear in an actual environment, CTC cannot completely cancel the cross-talk, and the signal level in the target position may erroneously increase in some frequency components [3]. This phenomenon of CTC is called a ringing effect, and an effort to reduce the ringing effect is necessary for practical usage of CTC.

A great deal of research on high-performance CTC has been carried out for stereo systems [1], [2], [4]. Multi-speaker systems with more than two speakers have also attracted interest for providing enhanced stereophonic sound image in large rooms. However, applying a CTC to a multi-speaker system has not yet been studied in any detail. In this paper, we propose a method of enhancing CTC performance for a four-speaker system, as the first example of a multi-speaker system, with respect to sweet spot size and ringing effect. The main concept of the proposed method is to operate the speakers in a selective way according to the listener’s position. To this end, we develop a method for selecting the operating speakers among four speakers such that the CTC performance with respect to both sweet spot size and ringing effect is improved. We conduct objective and subjective evaluations to confirm the enhanced performance of the proposed method compared to the conventional method without speaker selection.

2. Proposed Method of Cross-Talk Canceller

2.1 CTC for Four-Speaker System

Figure 1 shows a configuration of a four-speaker system equipped with CTC in a two-dimensional space, where four speakers, denoted by $SP_i$, are located in a row in the $x$ direction [5]. The position of $SP_i$ is $(x_{pi}, y_p)$, and the target listening position is $(x_0, y_0)$. The radius of the head is as-
sumed to be 0.08m.

A signal vector \( \mathbf{S} = [S_L(f) \ S_R(f)]^T \) is the original stereo input spectra, and \( S_L(f) \) and \( S_R(f) \) are the desired signals at the left and right ears of the listener, respectively. A \( 4 \times 4 \) CTC filter matrix \( \mathbf{C} \) in (1) is applied to \( \mathbf{S} \), and an input signal to \( \mathbf{SP}_l \) becomes the \( i \)-th element of \( \mathbf{CS} \).

\[
\mathbf{C} = \begin{bmatrix}
C_{L1}(f) & C_{R1}(f) \\
C_{L2}(f) & C_{R2}(f) \\
C_{L3}(f) & C_{R3}(f) \\
C_{L4}(f) & C_{R4}(f)
\end{bmatrix}
\] (1)

Each ear receives the sum of four signals from four speakers through different acoustic paths. An acoustic transfer function of each path from speaker to ear is denoted by a \( 2 \times 4 \) matrix \( \mathbf{H} \) in (2). In this paper, we use a free-field approximation for \( \mathbf{H} \), where \( H_{ij} \) is a function of frequency and travel distance [2].

\[
\mathbf{H} = \begin{bmatrix}
H_{11}(f) & H_{12}(f) & H_{13}(f) & H_{14}(f) \\
H_{21}(f) & H_{22}(f) & H_{23}(f) & H_{24}(f)
\end{bmatrix}
\] (2)

Assuming that the speaker bypasses the input, a speaker output vector is \( \mathbf{CS} \), and the final signal at the ear becomes \( \mathbf{HCS} \). Then, the aim of CTC is to make \( \mathbf{HCS} = \mathbf{S} \) in the target listening position \((x_0, y_0)\). Therefore, such \( \mathbf{C} \) is determined by \( \mathbf{C} = \mathbf{H}^+ = (\mathbf{HH}^T)^{-1} \), a pseudo inverse of \( \mathbf{H} \). Since both \( \mathbf{H} \) and \( \mathbf{C} \) are functions of \((x_0, y_0)\), we can express them as \( \mathbf{H}(x_0, y_0) \) and \( \mathbf{C}(x_0, y_0) \), respectively.

2.2 Sweet Spot of Conventional CTC

If a CTC was designed based on \( \mathbf{H}(x_0, y_0) \) but the actual listening position is changed to \((x_0 + \Delta x, y_0 + \Delta y)\), then the final signal at the ear is not \( \mathbf{S} \) because the working CTC filter \( \mathbf{C}(x_0, y_0) = \mathbf{H}^+(x_0, y_0) \) differs from \( \mathbf{H}^+(x_0 + \Delta x, y_0 + \Delta y) \), which is a pseudo inverse of the true acoustic path. The performance degradation by position error \((\Delta x, \Delta y)\) is measured from the channel separation ratio (CSR) defined by (3), where \( G_{ij} \) is an element of \( \mathbf{G} = \mathbf{H}(x_0 + \Delta x, y_0 + \Delta y) \mathbf{C}(x_0, y_0) \) [4].

\[
\begin{align*}
\text{CSR}_L &= 10 \log_{10} \frac{\int |G_{11}(f)|^2 df}{\int |G_{12}(f)|^2 df}, \\
\text{CSR}_R &= 10 \log_{10} \frac{\int |G_{21}(f)|^2 df}{\int |G_{22}(f)|^2 df}
\end{align*}
\] (3)

In an ideal case when \((\Delta x, \Delta y) = (0, 0)\), CSR\(_L\) and CSR\(_R\) become infinity because \( G_{11}(f) = G_{22}(f) = 0 \). If both CSR\(_L\) and CSR\(_R\) at \((x_0 + \Delta x, y_0 + \Delta y)\) are larger than 10dB, \((x_0 + \Delta x, y_0 + \Delta y)\) is accepted as the sweet spot of CTC [4].

Figure 2 shows the theoretical sweet spot size of the conventional CTC as a function of \( x_0 \) when applied to the four-speaker system shown in Fig. 1 with \(x_p1 = -0.75m, x_p2 = -0.25m, x_p3 = 0.25m, x_p4 = 0.75m, y_p = 3m, \) and \( y_0 = 0 \). When the target position moves away from the center, which makes the speaker layout asymmetrical to the target position, the sweet spot size of CTC significantly decreases. Therefore, a new method is required which provides the constant sweet spot size, regardless of target position, for a four-speaker system.

2.3 Ringing Effect of Conventional CTC

When the target position \((x_0, y_0)\) is given, we first determine the acoustic path \( \mathbf{H}(x_0, y_0) \) using a given acoustic model, and then design a CTC filter such as \( \mathbf{C}(x_0, y_0) = \mathbf{H}^+(x_0, y_0) \). In this way, the speaker output is \( \mathbf{CS} = \mathbf{H}^+ \mathbf{S} \) and the undesired cross-talk is cancelled out when the output signal travels to the ear through the acoustic path.

In an actual listening environment, however, the true acoustic path may differ from the determined \( \mathbf{H}(x_0, y_0) \) due to the complex acoustic operation such as room reflection and diffraction caused by the head and ear. Then, the expected cross-talk cancellation does not occur. Furthermore, the signal received at the ear may contain some erroneous large frequency components, because the large frequency components in the speaker output \( \mathbf{H}^+ \mathbf{S} \) caused by inverse operation in \( \mathbf{H}^+ \) are not correctly compensated by the true acoustic path. In an ideal case, of course, even these large components in \( \mathbf{H}^+ \mathbf{S} \) will be completely compensated by the matched acoustic path. This phenomenon is called a ringing effect, and the overall impact of the ring effect reduces as the number of speakers increases and the distance between speakers decreases [3]. Therefore, when running a CTC, we must include a number of special operations to reduce the ringing effect by ensuring that speaker output \( \mathbf{CS} = \mathbf{H}^+ \mathbf{S} \) does not have large frequency components, in preparation for incorrect compensation.

2.4 Proposed Selective Speaker Operation

After analyzing the sweet spot size of the conventional CTC for the four-speaker system, we found in Fig. 2 that a large sweet spot occurs when the speakers are symmetrical to the target position \((x_0, y_0)\). In addition, we know that the ringing effect reduces when using the maximum number of speakers with a short distance between them [3]. Based on these analyses, we propose an enhanced CTC for a four-speaker system based on a selective speaker operation, and develop a speaker selection method which enables the CTC operating
only the selected speakers to provide a large sweet spot and reduced ringing effect, compared to the conventional CTC.

For a large sweet spot, the speakers need to be symmetrical to \((x_0, y_0)\) as much as possible. Therefore, at the first step, we select two speakers among the four speakers that are most symmetrical to \((x_0, y_0)\). The degree of symmetry of the speaker pair is measured from \(d_i/d_j\), where \(d_i = |x_0 - x_p|\) is the \(x\)-directional distance between \((x_0, y_0)\) and \(SP_i\), and \(d_i > d_j\). The speaker pair with \(d_i/d_j\) closest to 1.0 is selected, and if multiple speaker pairs satisfy this condition, one with the largest distance between speakers is selected. At the next step, we add all speakers between the selected speaker pair to the final speaker selection. The purpose of this step is to reduce the ringing effect by increasing the number of speakers and decreasing the distance between them.

After selecting \(P\) speakers for a given \((x_0, y_0)\) based on the above steps, we determine a \(2 \times P\) acoustic path matrix \(H(x_0, y_0)\), excluding the acoustic paths for the unsellected speakers. Then, we implement the final CTC by setting \(C(x_0, y_0) = H^+(x_0, y_0)\), and the speaker input matrix \(CS\) becomes a \(P \times 1\) matrix. In this way, input to the \(P\) selected speakers is determined from \(CS\) and that to the unsellected speakers becomes null.

### 3. Performance Evaluation

We evaluated the performance of the proposed method for the test system with \(x_p1 = -0.75\), \(x_p2 = +0.25\), \(x_p3 = 0.25\), \(x_p4 = 0.75\), \(y_p = 3\), \(-0.75 < x_0 < 0.75\), and \(y_0 = 0\). Table 1 shows the selected speakers when applying the proposed method to the test system.

Figure 2 shows the sweet spot size of the proposed CTC and the conventional CTC for the test system. The conventional CTC always operates all four speakers, and the proposed CTC selects the operating speakers as in Table 1. Both CTC’s design the CTC filters in the same way such as \(C(x_0, y_0) = H^+(x_0, y_0)\) for a given \((x_0, y_0)\). It can be seen that the proposed CTC provides a wider sweet spot than the conventional CTC when the listener moves away from the center, and has almost a constant sweet spot size with varying target position.

We verify the ringing effect reduction of the proposed method by analyzing the output signal of the speaker. Figure 3 shows the spectral magnitude of \(SP_4\) output for the three cases of speaker operation when the original input \(S\) is a white signal and \((x_0, y_0) = (0.35, 0)\). Figure 3(a) shows the result when \(SP_2\) and \(SP_3\) are selected as in the proposed method, while Fig. 3(b) shows the result when only \(SP_2\) and \(SP_4\) are selected without an intermediary speaker \(SP_3\). Figure 3(c) shows the result of the conventional method operating all four speakers. We can see that the proposed method reduces the ringing effect, compared to the case without an intermediary speaker, with fewer boosting frequency components and a smaller boosting level. Also, even with fewer operating speakers than the conventional method, the proposed method has a ringing effect that is almost equivalent to that of the conventional method. The output of other speakers shows the same ringing effect reduction. These analyses confirm that the proposed selection method effectively reduces the ringing effect by selecting intermediary speakers as well as selecting symmetric speakers.

We also evaluated the subjective performance of CTC for the test system by an experiment of sound location estimation [6], when \((x_0, y_0) = (-0.35, 0)\) as a typical example. The purpose of this experiment is to determine the CTC that provides the best stereophonic image to the listener. The experimental conditions are summarized as follows:

- **Room:** Quiet office room of \(7m \times 7.9m \times 2.7m\)
- **Speaker:** KRK RP8 G3
- **Subjects:** 5 males, 1 female, age of 23 ~ 29
- **Target position:** \((x_0, y_0) = (-0.35, 0)\)
- **Listening positions:** \((x_0 + \Delta x, y_0 + \Delta y) = (-0.35, \pm 0.5)\)

A stereo pink noise of 10sec with virtual angles of \(0^\circ, \pm 30^\circ, \pm 60^\circ\) and \(\pm 60^\circ\) is generated using a virtual path from the sound location to the ear based on a free-field approximation [2]. Then, the signal is inputted to the test system equipped with two different CTC’s; one is the proposed CTC with selective speaker operation, and the other is the conventional CTC in which all four speakers are operating.

The subjects hear the sound from randomly selected locations among \(0^\circ, \pm 30^\circ, \pm 60^\circ\), and estimate the perceived sound location as one of \(0^\circ, 30^\circ, 60^\circ, -30^\circ\) and \(-60^\circ\). The subjects are allowed to repeatedly hear each sound until they can decide the perceived sound location. The same experiment is conducted three times for each subject for all test signals in each position.

Theoretically, for the proposed CTC, \((-0.35, 0.5)\) and \((-0.35, -0.5)\) are inside the sweet spot because each position has a CSR larger than 10dB. In contrast, for the conventional CTC, two positions are outside the sweet spot. There-

**Table 1**

<table>
<thead>
<tr>
<th>Target position in (x)-axis, (x_0)(m)</th>
<th>Selected speakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>(-0.75) ~ (-0.417)</td>
<td>(SP_1, SP_2)</td>
</tr>
<tr>
<td>(-0.417) ~ (-0.150)</td>
<td>(SP_1, SP_2, SP_3)</td>
</tr>
<tr>
<td>(-0.150) ~ (0.150)</td>
<td>(SP_1, SP_2, SP_3, SP_4)</td>
</tr>
<tr>
<td>(0.150) ~ (0.417)</td>
<td>(SP_2, SP_3, SP_4)</td>
</tr>
<tr>
<td>(0.417) ~ (0.750)</td>
<td>(SP_3, SP_4)</td>
</tr>
</tbody>
</table>

![Fig. 3](image-url)  
**Fig. 3** Speaker output spectrum for different speaker selections. (a) Selection using the proposed method. (b) Selection using the proposed method without intermediary speaker. (c) All four speakers using the conventional method.
Table 2  Angle estimation ratio (%) in the form of confusion matrix.

<table>
<thead>
<tr>
<th></th>
<th>-60</th>
<th>-30</th>
<th>0</th>
<th>30</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δy = 0.5</td>
<td>27.8</td>
<td>16.7</td>
<td>5.6</td>
<td>5.6</td>
<td>5.6</td>
</tr>
<tr>
<td>Estimated angle (°)</td>
<td>0.0</td>
<td>5.6</td>
<td>16.7</td>
<td>27.8</td>
<td></td>
</tr>
<tr>
<td>Δy = -0.5</td>
<td>55.6</td>
<td>16.7</td>
<td>11.1</td>
<td>11.1</td>
<td>11.1</td>
</tr>
<tr>
<td>Estimated angle (°)</td>
<td>0.0</td>
<td>11.1</td>
<td>16.7</td>
<td>55.6</td>
<td></td>
</tr>
<tr>
<td>Proposed method</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δy = 0.5</td>
<td>38.9</td>
<td>22.2</td>
<td>7.8</td>
<td>7.8</td>
<td>7.8</td>
</tr>
<tr>
<td>Estimated angle (°)</td>
<td>0.0</td>
<td>7.8</td>
<td>22.2</td>
<td>55.6</td>
<td></td>
</tr>
<tr>
<td>Δy = -0.5</td>
<td>77.8</td>
<td>11.1</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Estimated angle (°)</td>
<td>0.0</td>
<td>0.0</td>
<td>11.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 4  Average estimated angle. (a) Conventional method for Δy = 0.5. (b) Proposed method for Δy = 0.5. (c) Conventional method for Δy = -0.5. (d) Proposed method for Δy = -0.5.

Table 3  Average of absolute estimation error.

<table>
<thead>
<tr>
<th></th>
<th>Δy = 0.5</th>
<th>Δy = -0.5</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conventional</td>
<td>21.7°</td>
<td>37.0°</td>
<td>29.3°</td>
</tr>
<tr>
<td>Proposed</td>
<td>12.3°</td>
<td>21.0°</td>
<td>16.7°</td>
</tr>
</tbody>
</table>

fore, if the proposed CTC shows better performance in location estimation at two positions than the conventional CTC, we can conclude that the proposed CTC for four-speaker system has a wider sweet spot than the conventional method.

Table 2 shows the angle estimation ratio for all test cases in the form of a confusion matrix, where the diagonal cells correspond to the correct estimation ratio. In most cases, the proposed method has higher correct estimation ratio than the conventional method. Figure 4 shows the average estimated angle for each true angle with 95% confidence interval, and Table 3 shows the overall average of absolute estimation error between the true angle and the estimated angle in each position. From Fig. 4 and Table 3, we can see that the proposed method has better location estimation performance on average than the conventional method.

Finally, we conducted a Student’s t-test to verify the statistical significance of the experiment; if \( p < 0.05 \), the statistical performances of the two groups significantly differ [6]. In our experiment, we have \( p = 0.0102 \) for Δy = 0.5 and \( p = 0.0003 \) for Δy = -0.5. Therefore, we can conclude that at Δy = ±0.5, the proposed CTC yields a significantly better stereophonic image than the conventional CTC, which subjectively confirms the larger sweet spot by the proposed CTC when \((x_0, y_0) = (-0.35, 0)\).

4. Conclusion

In this paper, we propose a method to enhance the performance of CTC for a four-speaker system by selective speaker operation. In order to ensure a wider sweet spot, we select two speakers that are symmetrical to the target position. We then add all intermediary speakers to the final selection, which reduces the ringing effect of CTC by increasing the number of speakers and decreasing the distance between speakers. Finally, we compute the CTC filter after deleting the acoustic paths of the unselected speakers, and apply it to the input signal. The objective and subjective evaluations confirm that the proposed method enhances the performance of CTC for the four-speaker system with respect to sweet spot size and ringing effect, compared to the conventional method operating all four speakers.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (NIPA-2014-H0301-14-1021) supervised by the NIPA (National IT Industry Promotion Agency).

References