Audio Engineering Society

Convention Paper 10237

Presented at the AES 159th Convention
2025 October 23–25, Long Beach, CA, USA

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Increasing the Sweet Spot Size in Multi-channel Crosstalk Cancellation Systems

Xing Tian[1], Ning Cong[1], and Jianbin Yang[1]

[1]*Dynaudio Lab, Gammel Lundtoftevej 3B, Copenhagen, Denmark*

Correspondence should be addressed to Jianbin Yang (stanley.yang@goerdyna.com)

ABSTRACT

The sweet spot in a crosstalk cancellation system can be expanded by using more than two loudspeakers, but the improvement remains insufficient. To address this problem, this study proposes a weighted contrast control approach that employs multiple control points and an acoustic contrast strategy instead of conventional inverse filtering. The proposed method enhances robustness by reducing dependence on the condition number of transfer function matrices, enabling an effective increase in the sweet spot size. Experiments conducted using a loudspeaker line array demonstrate that the proposed approach outperforms multi-channel inverse filtering in both absolute and relative sweet spot sizes. Additionally, the proposed method achieves greater binaural channel separation at the target position and improved array output efficiency. These findings highlight the weighted contrast control approach as a robust and efficient solution for expanding the sweet spot in multi-channel crosstalk cancellation systems.

1 Introduction

Spatial audio technology is increasingly applied in fields such as home theaters, gaming, and virtual reality, providing users with an immersive listening experience. A technical approach to achieving spatial audio combines binaural rendering with crosstalk cancellation (CTC). Binaural rendering encodes audio objects from any position in three-dimensional space into one binaural signal; CTC ensures that the binaural signal, when played through loudspeaker systems, is transmitted accurately to the listener's ears while minimizing acoustic interference between them.

In recent decades, various CTC techniques have been developed [1–5], with optimal filters typically designed for a fixed position. As a result, the effective listening area, commonly known as the "sweet spot", is limited to a narrow region. Minor head movements by the listener, such as a few centimeters to either lateral side, may entirely compromise the crosstalk cancellation effect. To cope with this issue, some studies have explored alternative loudspeaker arrangements to expand the sweet spot, including stereo dipole setup [6], optimal source distribution [7, 8], and the addition of elevated loudspeakers [9]. Furthermore, with a fixed loudspeaker layout, existing improved approaches can be categorized into two main types. The first is the adaptive algorithm, which dynamically updates the CTC filters in real time by incorporating a head movement tracking device [10, 11]. The second is the regional control algorithm, which enlarges the sweet spot by increasing the number of control points [12, 13]. However, the adaptive algorithm requires additional hardware and increases computational complexity, making it unsuitable

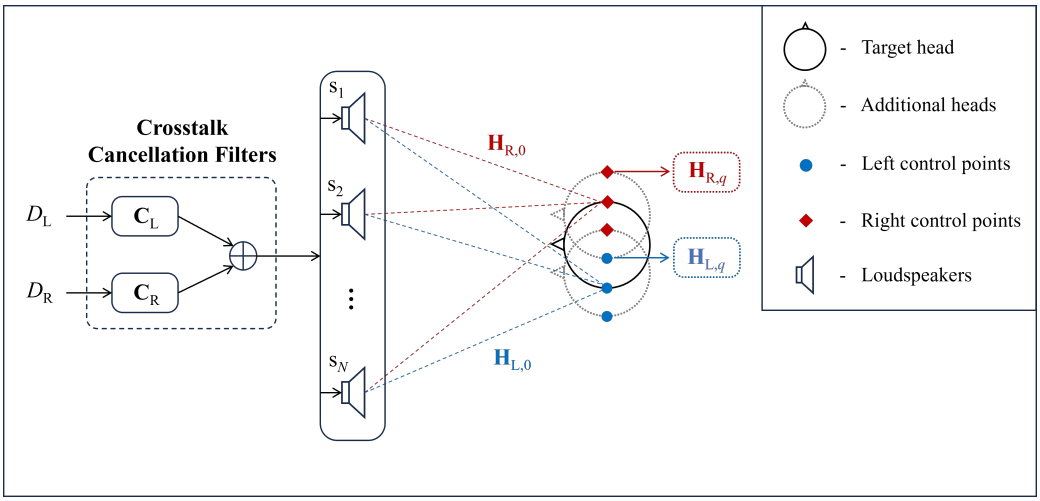


Fig. 1: Diagram of a multi-channel crosstalk cancellation system with multiple control points.

for low-cost applications. In contrast, the regional control algorithm, which typically relies on deconvolution methods [1], is easier to implement but suffers from reduced robustness as the number of control points increases, resulting in diminishing improvement to the sweet spot size.

Acoustic contrast optimization has been widely applied in sound zone control [14] and more recently in CTC [15]. The novelty of this work lies in combining weighted acoustic contrast with multiple control points for multi-channel loudspeaker systems, achieving improved robustness and an expanded sweet spot compared with conventional deconvolution-based regional control methods.

The remainder of this paper is organized as follows. Sec. 2 introduces the multi-channel crosstalk cancellation system and describes two approaches, including the proposed one. Sec. 3 details the experimental setup and presents the results of objective evaluations. Furthermore, the performance differences between methods are discussed in Sec. 4. Finally, Sec. 5 provides the conclusion.

2 Crosstalk Cancellation Methods

Fig. 1 illustrates the schematic diagram of a multi-channel crosstalk cancellation system. The system includes N loudspeakers, and the transfer function matrices from each loudspeaker to the left and right ears of the target head are denoted as $\mathbf{H}_{\mathrm{L},0}$ and $\mathbf{H}_{\mathrm{R},0}$, respectively. To expand the sweet spot, $Q+1$ pairs of control points, representing Q additional heads and one target head, can be introduced, with the additional heads positioned near the target head. Similarly, the transfer function matrices from each loudspeaker to the left and right ears of the q-th ($q = 1,2,\ldots,Q$) additional head are denoted as $\mathbf{H}_{\mathrm{L},q}$ and $\mathbf{H}_{\mathrm{R},q}$, respectively. Using these transfer function matrices, various methods can be employed to design the CTC filters $\mathbf{C}_\mathrm{L}$ and $\mathbf{C}_\mathrm{R}$. These two filters are applied to the binaural signals D_L and D_R, ensuring that the playback signals at the target head position closely match D_L and D_R while eliminating crosstalk between the two ears as possible.

Considering frequency-domain solutions, the dimensions of variables for the left ear at each discrete frequency bin are defined as follows,

$$D_\mathrm{L} \in \mathbb{C},\ \mathbf{C}_\mathrm{L} \in \mathbb{C}^{N\times 1},\ \mathbf{H}_{\mathrm{L},0} \in \mathbb{C}^{1\times N},\ \mathbf{H}_{\mathrm{L},q} \in \mathbb{C}^{1\times N}. \tag{1}$$

The dimensions of variables for the right ear are identical to those for the left ear. Playback signals at the left and right ears of the target head can be expressed as

$$E_\mathrm{L} = \mathbf{H}_{\mathrm{L},0}\left(\mathbf{C}_\mathrm{L}D_\mathrm{L} + \mathbf{C}_\mathrm{R}D_\mathrm{R}\right), \tag{2}$$

$$E_\mathrm{R} = \mathbf{H}_{\mathrm{R},0}\left(\mathbf{C}_\mathrm{L}D_\mathrm{L} + \mathbf{C}_\mathrm{R}D_\mathrm{R}\right). \tag{3}$$

2.1 Multi-channel Inverse Filter

The multi-channel inverse filter (MIF) based on deconvolution aims to minimize the mean squared er-

ror between the playback signal and the desired signal [16, 17], with the cost function defined as

$$J_{\mathrm{MIF}} = \begin{bmatrix} E_{\mathrm{L}} - D_{\mathrm{L}} & E_{\mathrm{R}} - D_{\mathrm{R}} \end{bmatrix} \begin{bmatrix} E_{\mathrm{L}} - D_{\mathrm{L}} \\ E_{\mathrm{R}} - D_{\mathrm{R}} \end{bmatrix}. \quad (4)$$

Minimizing the cost function in a least-squares sense yields the optimal CTC filter solution,

$$\overline{\mathbf{C}} = \left(\mathbf{H}^{*}\mathbf{H} + \beta\mathbf{I}\right)^{-1}\mathbf{H}^{*} \cdot \Lambda, \quad (5)$$

where $\overline{\mathbf{C}} = \begin{bmatrix} \mathbf{C}_{\mathrm{L}} & \mathbf{C}_{\mathrm{R}} \end{bmatrix}$, $*$ denotes the conjugate transpose operation, β represents the regularization factor, and $\mathbf{I} \in \mathbb{R}^{N\times N}$ represents the identity matrix. The expressions for $\mathbf{H} \in \mathbb{C}^{2(Q+1)\times N}$ and $\Lambda \in \mathbb{C}^{2(Q+1)\times 2}$ are given as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{\mathrm{L},0} \\ \mathbf{H}_{\mathrm{R},0} \\ \vdots \\ \mathbf{H}_{\mathrm{L},q} \\ \mathbf{H}_{\mathrm{R},q} \\ \vdots \end{bmatrix}, \quad \Lambda = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \ddots & \ddots \\ 1 & 0 \\ 0 & 1 \\ \ddots & \ddots \end{bmatrix}. \quad (6)$$

2.2 Proposed Approach

Using a strategy distinct from MIF, the approach proposed in this work defines the cost function based on the acoustic contrast between the left and right control points,

$$J = \frac{\mathbf{C}^{*}\mathbf{S}_{\mathrm{L}}\mathbf{C}}{\mathbf{C}^{*}\mathbf{S}_{\mathrm{R}}\mathbf{C}}, \quad (7)$$

and $\mathbf{C}$ represents the CTC filter for the left or right ear,

$$\mathbf{S}_{\mathrm{L}} = \alpha \cdot \mathbf{H}_{\mathrm{L},0}^{*}\mathbf{H}_{\mathrm{L},0} + (1-\alpha) \cdot \sum_{q=1}^{Q} \mathbf{H}_{\mathrm{L},q}^{*}\mathbf{H}_{\mathrm{L},q}, \quad (8)$$

$$\mathbf{S}_{\mathrm{R}} = \alpha \cdot \mathbf{H}_{\mathrm{R},0}^{*}\mathbf{H}_{\mathrm{R},0} + (1-\alpha) \cdot \sum_{q=1}^{Q} \mathbf{H}_{\mathrm{R},q}^{*}\mathbf{H}_{\mathrm{R},q}. \quad (9)$$

In Eq. (8) and (9), $0 < \alpha \leq 1$ represents the weight factor. The numerator of the cost function J denotes the weighted energy sum at all left control points, whereas the denominator represents the weighted energy sum at all right control points. Hence, the proposed approach is referred to as the weighted contrast control (WCC) method. Notably, when $\alpha = 1$, the cost function J is determined solely by the transfer function matrices from the target head position, i.e., $\mathbf{H}_{\mathrm{L},0}$ and $\mathbf{H}_{\mathrm{R},0}$.

The optimal CTC filters at the left and right ears involves maximizing and minimizing the cost function J. Specifically, this requires calculating the eigenvectors corresponding to the maximum and minimum eigenvalues of the matrix $\mathbf{S}_{\mathrm{R}}^{-1}\mathbf{S}_{\mathrm{L}}$. To ensure the filter stability, regularization must be applied to the matrix inversion. Accordingly, this work defines two separate cost functions,

$$J_{\mathrm{L}} = \frac{\mathbf{C}_{\mathrm{L}}^{*}\mathbf{S}_{\mathrm{L}}\mathbf{C}_{\mathrm{L}}}{\mathbf{C}_{\mathrm{L}}^{*}\left(\mathbf{S}_{\mathrm{R}} + \lambda_{\mathrm{R}}\mathbf{I}\right)\mathbf{C}_{\mathrm{L}}}, \quad (10)$$

$$J_{\mathrm{R}} = \frac{\mathbf{C}_{\mathrm{R}}^{*}\mathbf{S}_{\mathrm{R}}\mathbf{C}_{\mathrm{R}}}{\mathbf{C}_{\mathrm{R}}^{*}\left(\mathbf{S}_{\mathrm{L}} + \lambda_{\mathrm{L}}\mathbf{I}\right)\mathbf{C}_{\mathrm{R}}}, \quad (11)$$

incorporating two regularization factors, λ_{L} and λ_{R}, for calculating the left and right CTC filters. In this manner, the optimal solutions for $\mathbf{C}_{\mathrm{L}}$ and $\mathbf{C}_{\mathrm{R}}$, both corresponding to a maximization optimization problem, are given by

$$\mathbf{C}_{\mathrm{L}} = \Phi\left\{\left(\mathbf{S}_{\mathrm{R}} + \lambda_{\mathrm{R}}\mathbf{I}\right)^{-1}\mathbf{S}_{\mathrm{L}}\right\}, \quad (12)$$

$$\mathbf{C}_{\mathrm{R}} = \Phi\left\{\left(\mathbf{S}_{\mathrm{L}} + \lambda_{\mathrm{L}}\mathbf{I}\right)^{-1}\mathbf{S}_{\mathrm{R}}\right\}. \quad (13)$$

Here, $\Phi\{\mathbf{A}\}$ represents the eigenvector corresponding to the maximum eigenvalue of matrix $\mathbf{A}$.

The $\mathbf{C}_{\mathrm{L}}$ and $\mathbf{C}_{\mathrm{R}}$ obtained from Eq. (12) and (13) are both unit vectors and must be further normalized to ensure that the playback signals E_{L} and E_{R} at the target head position match D_{L} and D_{R} in amplitude and a constant delay in phase. Assuming complete elimination of crosstalk in the playback signals, we have

$$E_{\mathrm{L}} \simeq \mathbf{H}_{\mathrm{L},0}\mathbf{C}_{\mathrm{L}}D_{\mathrm{L}} = e^{j\varphi} \cdot D_{\mathrm{L}}, \quad (14)$$

$$E_{\mathrm{R}} \simeq \mathbf{H}_{\mathrm{R},0}\mathbf{C}_{\mathrm{R}}D_{\mathrm{R}} = e^{j\varphi} \cdot D_{\mathrm{R}}. \quad (15)$$

Where φ represents the common phase, which can be taken as the phase sum of $\mathbf{H}_{\mathrm{L},0}\mathbf{C}_{\mathrm{L}}$ and $\mathbf{H}_{\mathrm{R},0}\mathbf{C}_{\mathrm{R}}$; thus, the normalized equations for $\mathbf{C}_{\mathrm{L}}$ and $\mathbf{C}_{\mathrm{R}}$ are obtained as follows,

$$\mathbf{C}_{\mathrm{L}} \leftarrow \frac{\mathbf{C}_{\mathrm{L}}}{|\mathbf{H}_{\mathrm{L},0}\mathbf{C}_{\mathrm{L}}|} \cdot \frac{\mathbf{H}_{\mathrm{R},0}\mathbf{C}_{\mathrm{R}}}{|\mathbf{H}_{\mathrm{R},0}\mathbf{C}_{\mathrm{R}}|}, \quad (16)$$

$$\mathbf{C}_{\mathrm{R}} \leftarrow \frac{\mathbf{C}_{\mathrm{R}}}{|\mathbf{H}_{\mathrm{R},0}\mathbf{C}_{\mathrm{R}}|} \cdot \frac{\mathbf{H}_{\mathrm{L},0}\mathbf{C}_{\mathrm{L}}}{|\mathbf{H}_{\mathrm{L},0}\mathbf{C}_{\mathrm{L}}|}. \quad (17)$$

For both the MIF and WCC methods, selecting an appropriate regularization factor is crucial. A smaller regularization factor results in a weaker amplitude constraint on the CTC filters, which may lead to excessive

high-energy output and reduce the loudspeaker system's efficiency. Conversely, a larger regularization factor improves output efficiency but also increases the residual crosstalk in the playback signals. The array effort (AE) is commonly used to measure the output efficiency of a multi-channel system and to determine the value of the regularization factor. AE is defined as the logarithmic energy ratio of two filter sets [18],

$$\mathrm{AE} = 10\log_{10}\left(\frac{\mathbf{C}_{\mathrm{opt}}^{*}\mathbf{C}_{\mathrm{opt}}}{\mathbf{C}_{\mathrm{ref}}^{*}\mathbf{C}_{\mathrm{ref}}}\right), \tag{18}$$

where $\mathbf{C}_{\mathrm{opt}}$ represents the optimized filter $\mathbf{C}_{\mathrm{L}}$ or $\mathbf{C}_{\mathrm{R}}$; $\mathbf{C}_{\mathrm{ref}}$ represents the filter that drives all loudspeakers in-phase to produce the same average sound pressure level in the left or right ear as produced by using $\mathbf{C}_{\mathrm{opt}}$. It can be observed that a larger AE corresponds to lower output efficiency of the system. The values of the regularization factors β, λ_{L} and λ_{R} can be initially set to small values and then increased iteratively until the AE reaches a value lower than the specified maximum.

3 Experimental Results

Experiments are conducted in a listening room and utilize a loudspeaker line array with a total length of 1.7 m, consisting of 10 identical full-range loudspeaker drivers, as shown in Fig. 2. The irregular driver arrangement results from the specifications of our prototype. The target head position is located 2 m in front of the array center, and transfer functions from each loudspeaker to all control points are measured using a Neumann KU100 dummy head with a sampling rate of 44.1 kHz.

The experiments compare two control point layouts:

(a). $Q = 2$, the additional two heads are positioned at ± 2 cm laterally from the target head position, resulting in a total of 3-pair control points.

(b). $Q = 4$, the additional four heads are positioned at ± 2 cm and ± 4 cm laterally from the target head position, resulting in a total of 5-pair control points.

In these two layouts, the MIF and WCC methods are applied to design the CTC filters, respectively. The CTC filter length is set to 4096, and the maximum AE is constrained to 10 dB. For clarity, the methods corresponding to layout (a) are labeled as MIF-H3 and WCC-H3, while those for layout (b) are labeled as MIF-H5 and WCC-H5. Given that the sweet spot size at low frequencies is relatively large compared to high frequencies, further expansion is deemed unnecessary. Therefore, the weight factor in the WCC method is set to $\alpha = 1$ below 1 kHz and $\alpha = 0.5$ above 1 kHz.

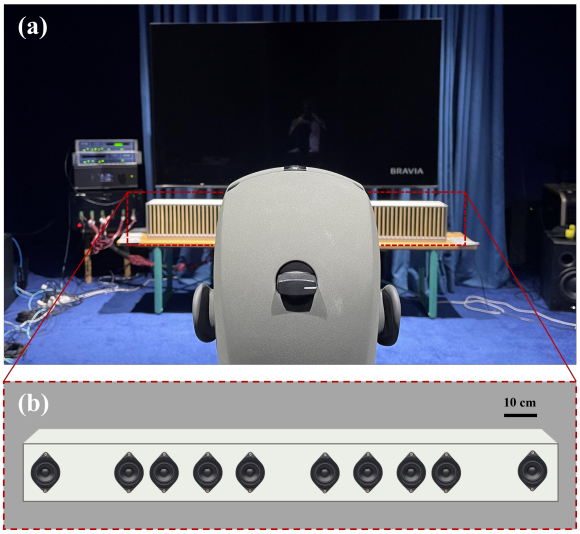


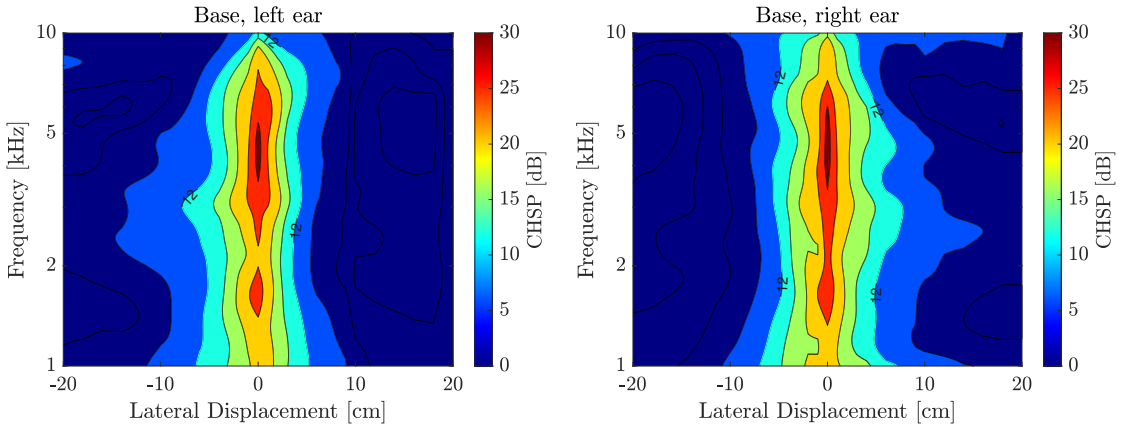
Fig. 2: (a) Experimental setup for evaluating a multi-channel crosstalk cancellation system using (b) a line array of ten identical 2.5-inch drivers.

For comparison, excluding control points at additional head positions, the MIF method is applied to control only the binaural control points at the target head position, serving as a baseline for the sweet spot size. This benchmark method is referred to as Base.

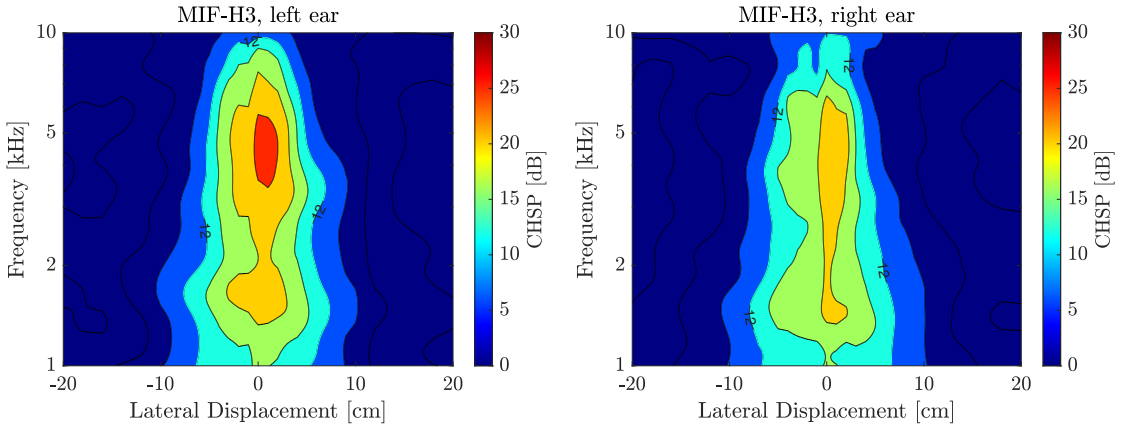
Channel separation (CHSP) has been proposed to evaluate the performance of a crosstalk cancellation system and defined as the logarithmic difference between the playback signals at the two ears. The CHSP at left ear is measured with $D_{\mathrm{L}} = 1$ and $D_{\mathrm{R}} = 0$, while the CHSP at right ear is measured with $D_{\mathrm{L}} = 0$ and $D_{\mathrm{R}} = 1$. By measuring the binaural playback signals E_{L} and E_{R} at the target head position twice, $\mathrm{CHSP}_{\mathrm{L}}$ and $\mathrm{CHSP}_{\mathrm{R}}$ can be calculated as follows,

$$\mathrm{CHSP}_{\mathrm{L}} = 20\log_{10}\left(\frac{|E_{\mathrm{L}}|}{|E_{\mathrm{R}}|}\right), \text{ with } D_{\mathrm{R}} = 0, \tag{19}$$

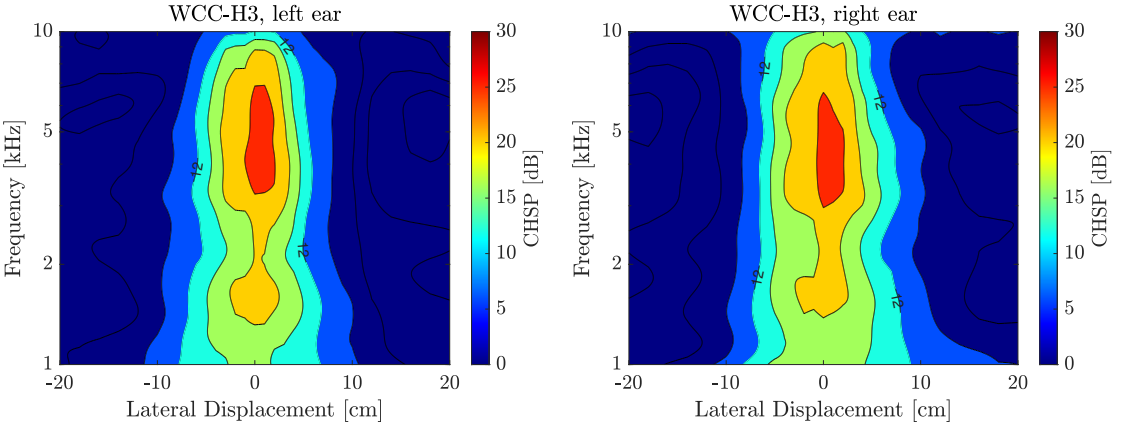
$$\mathrm{CHSP}_{\mathrm{R}} = 20\log_{10}\left(\frac{|E_{\mathrm{R}}|}{|E_{\mathrm{L}}|}\right), \text{ with } D_{\mathrm{L}} = 0. \tag{20}$$


Base, left ear
Base, right ear
Frequency [kHz]
Lateral Displacement [cm]
CHSP [dB]

(a)


MIF-H3, left ear
MIF-H3, right ear
Frequency [kHz]
Lateral Displacement [cm]
CHSP [dB]

(b)


WCC-H3, left ear
WCC-H3, right ear
Frequency [kHz]
Lateral Displacement [cm]
CHSP [dB]

(c)

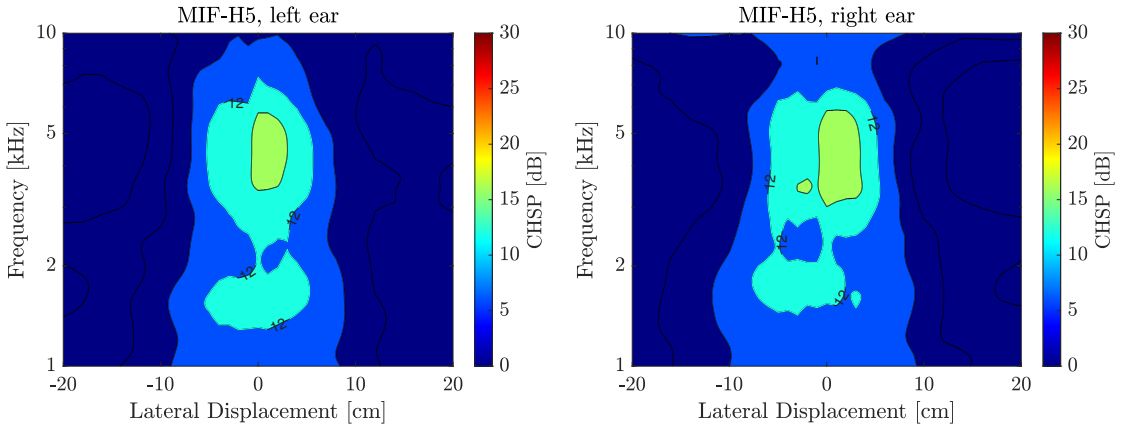


Fig. 3: Contour plots of channel separation versus lateral displacement for different methods: (a) Base, (b) MIF-H3, (c) WCC-H3, (d) MIF-H5, and (e) WCC-H5. Each sub-figure contains two plots representing the measurements for the left and right ears, respectively.

Through simulations and experiments, it is found that when $\alpha = 1$, although the filters designed by the WCC and Base methods differ, the binaural CHSP at the target head position is nearly identical for both methods. Therefore, only the frequency range between 1 kHz and 10 kHz is considered in following comparative evaluation of different methods. Below 1 kHz, there is minimal performance difference between the methods, while above 10 kHz, the sweet spot size is primarily influenced by natural separation rather than by the algorithmic contribution.

After applying the optimal CTC filters, the dummy head is incrementally shifted laterally from -20 cm to $+20$ cm to measure the CHSP at various offsets. The contour plots of the measurement results are presented in Fig. 3, and one-third octave band smoothing is applied along the frequency axis.

To quantitatively evaluate the sweet spot size, either the interaural time difference (ITD) or the interaural level difference (ILD) can be utilized [19, 20]. Since ITD primarily influences auditory localization below 1.5 kHz [21], This study employs the CHSP metric (equivalent to ILD) and evaluates two measures: absolute sweet spot and relative sweet spot [20]. The size of absolute sweet spot is defined as the range between the maximum leftward and rightward displacements at

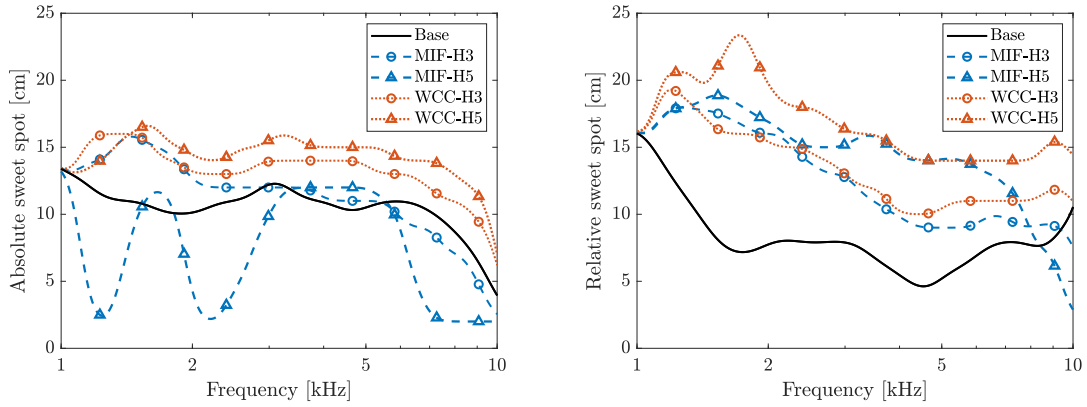


Fig. 4: Measured sizes of (a) the absolute sweet spot and (b) the relative sweet spot.

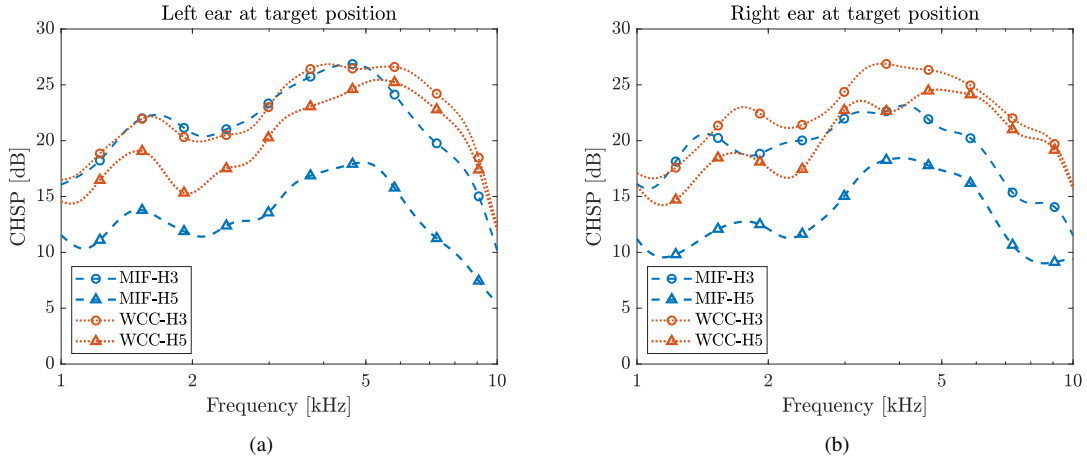


Fig. 5: Measured CHSP at (a) the left ear and (b) the right ear of the target head position.

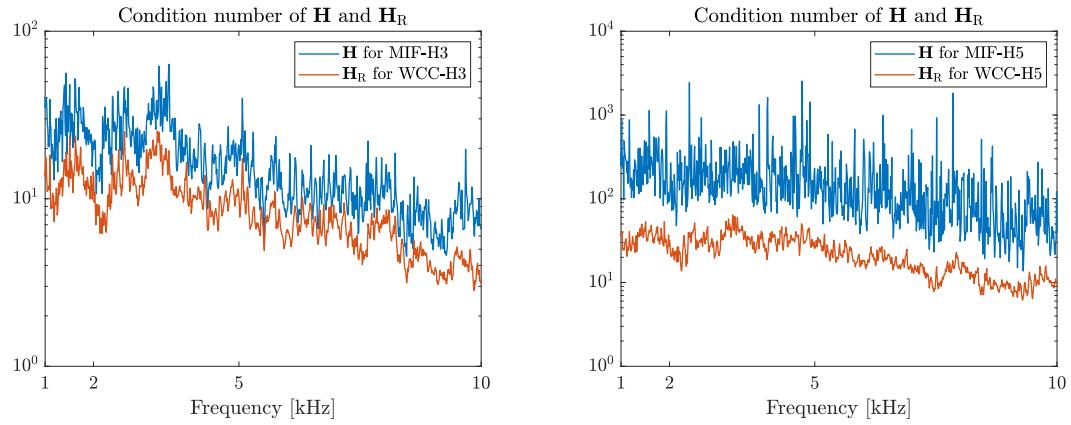
which the average CHSP (measured at the left and right ears) exceeds 12 dB. The size of relative sweet spot is defined as the range between the maximum leftward and rightward displacements where the average CHSP decreases by 12 dB compared to the nominal center position.

Fig. 4 illustrates the absolute and relative sweet spots achieved using different methods. Under the same control point layout, the WCC method outperforms the MIF method. Additionally, as the number of control points increases, the absolute and relative sweet spots of WCC-H5 expand further compared to WCC-H3. However, the absolute sweet spot of MIF-H5 is smaller than that of MIF-H3, indicating that the WCC method exhibits greater robustness in expanding the sweet spot size compared to the MIF method.

On the other hand, as the sweet spot size increases, the binaural CHSP at the target head position becomes equally important. For both control point layouts, the WCC method achieves a higher CHSP at the target position than the MIF method, as shown in Fig. 5. This demonstrates another advantage of the proposed approach, suggesting better crosstalk cancellation performance.

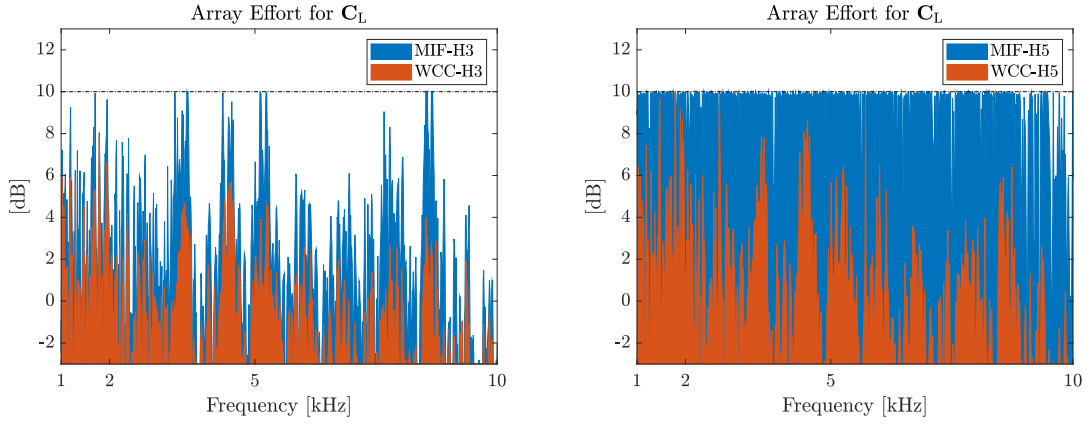
4 Discussion

In the experimental system, the performance differences between the MIF and the WCC methods primarily arise from the ill-conditioned nature of matrix



(a) (b)

Fig. 6: Condition numbers of the matrices $\mathbf{H}$ and $\mathbf{H}_\mathrm{R}$ used in designing the CTC filter for the left ear under (a) 3-pair and (b) 5-pair control points.



(a) (b)

Fig. 7: AE curves for the left ear CTC filter under (a) 3-pair and (b) 5-pair control points.

inversion. Notably, the Hermitian matrices $\mathbf{S}_\mathrm{L}$ and $\mathbf{S}_\mathrm{R}$ in Eq. (12) and (13) can be further represented as $\mathbf{S}_\mathrm{L} = \mathbf{H}_\mathrm{L}^* \mathbf{H}_\mathrm{L}$ and $\mathbf{S}_\mathrm{R} = \mathbf{H}_\mathrm{R}^* \mathbf{H}_\mathrm{R}$, where

$$\mathbf{H}_\mathrm{L} = \begin{bmatrix} \sqrt{\alpha} \cdot \mathbf{H}_{\mathrm{L},0} \\ \sqrt{1-\alpha} \cdot \mathbf{H}_{\mathrm{L},1} \\ \vdots \\ \sqrt{1-\alpha} \cdot \mathbf{H}_{\mathrm{L},Q} \end{bmatrix}, \quad \mathbf{H}_\mathrm{R} = \begin{bmatrix} \sqrt{\alpha} \cdot \mathbf{H}_{\mathrm{R},0} \\ \sqrt{1-\alpha} \cdot \mathbf{H}_{\mathrm{R},1} \\ \vdots \\ \sqrt{1-\alpha} \cdot \mathbf{H}_{\mathrm{R},Q} \end{bmatrix}. \tag{21}$$

Therefore, the WCC method requires only the pseudo-inverse of $\mathbf{H}_\mathrm{L}$ or $\mathbf{H}_\mathrm{R}$ when calculating the CTC filter for left or right ear. In contrast, the MIF method involves the pseudo-inverse of $\mathbf{H}$, which is the concatenation of $\mathbf{H}_\mathrm{L}$ and $\mathbf{H}_\mathrm{R}$, see Eq. (6). The condition numbers of the respective matrices differ between these two methods. For instance, when solving the CTC filter for left ear, the condition number of the matrix $\mathbf{H}$ in the MIF method is consistently higher than that of the matrix $\mathbf{H}_\mathrm{R}$ in the WCC method, as illustrated in Fig. 6.

The larger the condition number of a matrix, the more ill-conditioned its inversion becomes, necessitating a larger regularization factor. Compared to the MIF method, the WCC method requires smaller regularization constraints, resulting in lower residual crosstalk and improved CHSP performance. Moreover, as shown in Fig. 7, even with a larger regularization factor, the AE values of the MIF method remain higher than those of the WCC method across most discrete frequency

bins, indicating reduced output efficiency for the MIF method.

It should be noted, however, that the choice of regularization is not only a numerical issue but also has perceptual implications. Excessive regularization may degrade performance, while insufficient regularization can introduce audible artifacts due to over-optimization. In practice, striking a balance between numerical stability and reproduced audio quality is critical. As part of future work, systematic listening tests with diverse source material will be conducted to evaluate the perceptual effects of different regularization settings and to refine the optimization strategy accordingly.

5 Conclusion

In multi-channel crosstalk cancellation systems, adding control points near the listener's head is a common strategy for expanding the sweet spot. However, as the number of control points increases, the condition number of the transfer function matrix between the loudspeakers and the control points also rises, particularly in the low- and mid-frequency ranges. This can lead to an ill-conditioned matrix and degraded crosstalk cancellation performance when using deconvolution methods. To mitigate this limitation, this study proposes an enhanced algorithm that reduces reliance on full matrix inversion by controlling the weighted acoustic energy between the ears. More precisely, in solving for the optimal crosstalk cancellation filter, the proposed method involves an inverse matrix with a consistently lower condition number than that of multi-channel inverse filtering, thereby improving robustness and enabling more effective sweet spot expansion.

Experimental results show that the proposed approach outperforms multi-channel inverse filtering in both absolute and relative sweet spot sizes and effectively utilizes more adjacent control points to further expand the sweet spot. Moreover, the proposed approach offers higher binaural channel separation at the target position and improved array output efficiency. These benefits emphasize its potential for a wide range of spatial audio applications.

References

[1] Kirkeby, O., Nelson, P. A., Hamada, H., and Orduna-Bustamante, F., "Fast deconvolution of multichannel systems using regularization," *IEEE Transactions on speech and audio processing*, 6(2), pp. 189–194, 1998.

[2] Ward, D. B., "Joint least squares optimization for robust acoustic crosstalk cancellation," *IEEE Transactions on Speech and Audio Processing*, 8(2), pp. 211–215, 2000.

[3] Bai, M. R. and Lee, C.-C., "Development and implementation of cross-talk cancellation system in spatial audio reproduction based on subband filtering," *Journal of Sound and Vibration*, 290(3-5), pp. 1269–1289, 2006.

[4] Rao, H. I., Mathews, V. J., and Park, Y.-C., "A minimax approach for the joint design of acoustic crosstalk cancellation filters," *IEEE transactions on audio, speech, and language processing*, 15(8), pp. 2287–2298, 2007.

[5] Masiero, B. and Vorländer, M., "A framework for the calculation of dynamic crosstalk cancellation filters," *IEEE/ACM transactions on audio, speech, and language processing*, 22(9), pp. 1345–1354, 2014.

[6] Kirkeby, O., Nelson, P. A., and Hamada, H., "The'stereo dipole': A virtual source imaging system using two closely spaced loudspeakers," *Journal of the Audio Engineering Society*, 46(5), pp. 387–395, 1998.

[7] Ward, D. B. and Elko, G. W., "Effect of loudspeaker position on the robustness of acoustic crosstalk cancellation," *IEEE Signal Processing Letters*, 6(5), pp. 106–108, 1999.

[8] Takeuchi, T. and Nelson, P. A., "Optimal source distribution for binaural synthesis over loudspeakers," *The Journal of the Acoustical Society of America*, 112(6), pp. 2786–2797, 2002.

[9] Parodi, Y. L. and Rubak, P., "Objective evaluation of the sweet spot size in spatial sound reproduction using elevated loudspeakers," *The Journal of the Acoustical Society of America*, 128(3), pp. 1045–1055, 2010.

[10] Lee, K.-S., "Position-dependent crosstalk cancellation using space partitioning," *IEEE transactions on audio, speech, and language processing*, 21(6), pp. 1228–1239, 2013.

[11] Song, M.-S., Zhang, C., Florencio, D., and Kang, H.-G., "An interactive 3-D audio system with loudspeakers," *IEEE Transactions on Multimedia*, 13(5), pp. 844–855, 2011.

[12] Bai, M. R., Tung, C.-W., and Lee, C.-C., "Optimal design of loudspeaker arrays for robust cross-talk cancellation using the Taguchi method and the genetic algorithm," *The Journal of the Acoustical Society of America*, 117(5), pp. 2802–2813, 2005.

[13] Bai, M. R., Chen, Y. W., Hsu, Y.-C., and Wu, T. Y., "Robust binaural rendering with the time-domain underdetermined multichannel inverse prefilters," *The Journal of the Acoustical Society of America*, 146(2), pp. 1302–1313, 2019.

[14] Choi, J.-W. and Kim, Y.-H., "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, 111(4), pp. 1695–1700, 2002.

[15] Samejima, T., Kobayashi, K., and Sekiguchi, S., "Transaural system using acoustic contrast as its objective function," *Acoustical Science and Technology*, 42(4), pp. 206–209, 2021.

[16] Parodi, Y. L. and Rubak, P., "Analysis of design parameters for crosstalk cancellation filters applied to different loudspeaker configurations," *Journal of the Audio Engineering Society*, 59(5), pp. 304–320, 2011.

[17] Gálvez, M. F. S., Menzies, D., and Fazi, F. M., "Dynamic audio reproduction with linear loudspeaker arrays," *Journal of the Audio Engineering Society*, 67(4), pp. 190–200, 2019.

[18] Liao, X., Cheer, J., Elliott, S., and Zheng, S., "Design of a loudspeaker array for personal audio in a car cabin," *Journal of the Audio Engineering Society*, 65(3), pp. 226–238, 2017.

[19] Rose, J., Nelson, P., Rafaely, B., and Takeuchi, T., "Sweet spot size of virtual acoustic imaging systems at asymmetric listener locations," *The Journal of the Acoustical Society of America*, 112(5), pp. 1992–2002, 2002.

[20] Bai, M. R. and Lee, C.-C., "Objective and subjective analysis of effects of listening angle on crosstalk cancellation in spatial sound reproduction," *The Journal of the Acoustical Society of America*, 120(4), pp. 1976–1989, 2006.

[21] Wightman, F. L. and Kistler, D. J., "The dominant role of low-frequency interaural time differences in sound localization," *The Journal of the Acoustical Society of America*, 91(3), pp. 1648–1661, 1992.