BLIND ESTIMATION OF ACOUSTIC TRANSFER FUNCTIONS WITH APPLICATION TO DEREVERBERATION USING CONVOLUTIVE TRANSFER FUNCTIONS

Anchi Yuan, You-Siang Chen, and Mingsian R. Bai

Department of Power Mechanical Engineering, National Tsing Hua University, No. 101, Section 2, Kuang-Fu Road, Hsinchu, Taiwan 30013.
Corresponding author: Mingsian R. Bai, Email: msbai@pme.nthu.edu.tw

ABSTRACT

While Acoustic Transfer Functions (ATFs) provide better performance than Relative Transfer Functions (RTFs) in many array signal processing tasks, the source signal is often inaccessible to obtain a reliable ATF estimate. To address this problem, we propose a novel blind ATF estimation approach based on Convolutive Transfer Functions (CTFs). First, the Weighted Prediction Error (WPE) algorithm and the Delay and Sum (DAS) beamformer are used to obtain a crude estimate of the target source signal at the source. Next, the CTF coefficients are computed using either Wiener filter or the Recursive Least Squares (RLS) algorithm. To recover the impulse responses of ATFs, the short-time Fourier transform (STFT) of a unit pulse sequence is convolved with the CTF coefficients before the inverse STFT is used. To demonstrate the effectiveness of the proposed ATF estimation technique, we take the dereverberation using the Multiple Input/Output Inverse Theorem (MINT), which requires accurate ATF estimates, as an application example. The simulation results using a 30-element linear array show that the proposed method yields ATF estimates in close agreement with the ground truth room impulse responses. The MINT-dereverberated signals closely match the dry source signals in terms of two objective metrics.

Index Terms **—** convolutive transfer functions, weighted prediction error, delay and sum beamformer, wiener filter, adaptive filter, multiple input/output inverse theorem

1. INTRODUCTION

Blind estimation refers to the identification of systems for which only the output signals are known, while minimal information about the input signals is available. This problem is of significant importance, as the estimated ATFs find applications in various acoustic scenarios, including acoustic echo cancellation [1] , dereverberation [2] , blind source separation [3], and beamforming in reverberant environments [4]. The majority of methods for identifying systems have been developed within the framework of the Short Time Fourier Transform (STFT) domain [5] [6] [7], where convolution in the time domain is approximated by a product of the source STFT and the room impulse response (RIR) STFT. This approximation is known as the multiplicative transfer function (MTF) approximation [8], or the narrowband approximation. Nonetheless, the MTF approximation is theoretically valid only when the length of the RIR is shorter than that of the STFT window. However, in practical settings, this condition is rarely met, even for moderate reverberant environments. This is due to the STFT window's limitation in assuming local stationarity of audio signals. Moreover, the use of a long STFT window can result in increased estimation variance and computational complexity.

In an attempt to accurately represent convolution in the STFT domain, particularly in situations with prolonged RIRs, cross-band filters (CBFs) were proposed in [9] for linear system identification. These CBFs offer an alternative to the MTF. In this approach, an STFT coefficient output is represented as the sum across frequency bins of multiple convolutions between the input source signal's STFT coefficients and the RIR in the time-frequency (TF) domain along the frame coordinate. To ensure analytical tractability, an approximation of the convolutive transfer function (CTF) was proposed in [10]. This approximation suggests that for each frequency, the output STFT coefficient can be modelled as a unique convolution between the input source signal's STFT coefficients and the CTF along the frame axis.

This paper presents a technique for estimating blind ATF based on the CTF approximation. The first step involves the dereverberation and extraction of the source signal required for computing the CTF coefficients through Weighted Prediction Error (WPE) [11] and Delay and Sum (DAS) beamforming [12]. Subsequently, the CTF coefficients are calculated using Wiener filters [13] or adaptive filters like Recursive Least Squares (RLS) [14]. To obtain the ATFs in the time domain (i.e. RIRs) from the CTF coefficients, the estimated CTF coefficients are convolved with the STFT of a time-shifted unit pulse sequence. The resulting convolved sequence is then processed using inverse STFT.

The validation of the proposed method for estimating the ATF involves a comparison between two versions of processed source signals. One of these signals undergoes

dereverberation by applying the Multiple Input/Output Inverse Theorem (MINT) [15], utilizing the estimated RIRs mentioned above. Meanwhile, the other variant is dereverberated solely by using WPE. The effectiveness of these signals is assessed through metrics such as the Perceptual Evaluation of Speech Quality (PESQ) [16] and Signal-to-Distortion Ratio (SDR) [17]. The simulations cover different reverberation times and utilize a thirty-microphone Uniform Linear Array (ULA). The results confirm that the proposed method outperforms WPE in various reverberation scenarios.

2. CTF SIGNAL MODEL

In a noise-free and echoless environment, the signal received by the microphone is presented in the time domain, as specified by

$$y(n) = a(n) * s(n), . \tag{1}$$

where the $s(n)$ and $a(n)$ denote the source signal and the RIR, respectively, while * indicates the linear convolution. In (1), the RIR is commonly estimated using the MTF in the STFT domain, as demonstrated by

$$y_{p,k} = a_k s_{p,k}, . \tag{2}$$

where $y_{p,k}$ and $s_{p,k}$ represent the STFTs of their respective signals, while a_k denotes the Fourier transformation of the RIR $a(n)$. In addition, $p \in [1, P]$ refers to the frame index, N indicates the STFT window size and $k \in [0, N–1]$ represents the frequency index. However, it should be emphasized that this approximation is only valid if the length of the RIR $a(n)$ is shorter than the STFT window [9]. Therefore, we utilize the cross-band filter model in this study. The STFT coefficient $y_{p,k}$ is presented as the sum of multiple convolutions between the STFT-domain source signal and the filter over the frequency bins in the following manner:

$$y_{p,k} = \sum_{k'=0}^{N-1} \sum_{p'} s_{p-p',k'}\ a_{p',k,k'} \cdot \tag{3}$$

Let D denote the STFT frame step. If $D < N$, then $a_{p',k,k'}$ is non-causal, with $\lceil N/D \rceil - 1$ non-causal coefficients [9]. The number of causal filter coefficients is related to the reverberation time. For the sake of notation simplicity, we assume that the filter index p' is in [0, L –1], with L being the filter length. This involves relocating the non-causal coefficients to the causal component, leading to a fixed delay shift of the frame index for the received microphone signal [9]. The STFT analysis and synthesis windows are represented as $\tilde{w}(n)$ and $w(n)$, respectively. The STFT domain impulse response $a_{p',k,k'}$ is related to the time domain impulse response $a(n)$ by

$$a_{p',k,k'} = (a(n) * \zeta_{k,k'}(n))\big|_{n=p'D}, \tag{4}$$

which indicates the convolution with respect to the time index n evaluated at frame steps using

$$\zeta_{k,k'}(n) = e^{j\frac{2\pi}{N}k'n} \sum_{m=-\infty}^{+\infty} \tilde{w}(m) w(m) e^{j\frac{2\pi}{N}m(k-k')} \cdot \tag{5}$$

To simplify the analysis, we utilize the CTF approximation, which solely considers the band-to-band filters with $k = k'$ as outlined in

$$y_{p,k} \approx \sum_{p'=0}^{L-1} s_{p-p',k}\ a_{p',k} = s_{p,k} * a_{p,k} \cdot \tag{6}$$

Based on this, we consider a multi-channel version with M microphones

$$y^i_{p,k} = \sum_{p'=0}^{L-1} s_{p-p',k}\ a^i_{p',k}, \tag{7}$$

where $y^i_{p,k}$ and $a^i_{p,k}$ signify the i-th microphone signal and the corresponding CTF, respectively. Hence, the source signals can be rewritten as the matrix form:

$$\underbrace{\begin{bmatrix} y^1_{p,k} \\ y^2_{p,k} \\ \vdots \\ y^M_{p,k} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} a^1_{0,k} & a^1_{1,k} & \cdots & a^1_{L-1,k} \\ a^2_{0,k} & a^2_{1,k} & \cdots & a^2_{L-1,k} \\ \vdots & \vdots & \ddots & \vdots \\ a^M_{0,k} & a^M_{1,k} & \cdots & a^M_{L-1,k} \end{bmatrix}}_{\mathbf{A}} \underbrace{\begin{bmatrix} s_{p,k} \\ s_{p-1,k} \\ \vdots \\ s_{p-(L-1),k} \end{bmatrix}}_{\mathbf{s}}. \tag{8}$$

As the proposed algorithm operates on a frequency basis, the frequency index is omitted hereafter for brevity.

3. PROPOSED METHOD

In this section, we present a method to estimate the ATF in a blind manner. It is important to note that we only have access to the microphone signal $\mathbf{y}$ (delayed by $\lceil N/D \rceil$ – 1 frames [9]) and source signal $\mathbf{s}_{DAS}$ obtained via DAS beamformer. Notably we offer two techniques for the estimation of the CTF coefficients.

3.1. Wiener filtering approach

For the first technique, we estimate the CTF coefficients matrix via the Wiener-based derivations, which minimizes the expectation of the mean squared error as expressed by

$$\min_{\mathbf{A} \in \mathbb{C}^{M \times L}} \mathbb{E}\left[\left\| \mathbf{y} - \mathbf{A}\mathbf{s}_{DAS} \right\|_2^2 \right], \tag{9}$$

where $\mathbb{E}[\cdot]$ denotes the expectation with respect to the frames. Therefore, (9) can be rewritten as

$$\min_{\mathbf{A} \in \mathbb{C}^{M \times L}} tr\left\{ \mathbf{R}_{\mathbf{yy}} - \mathbf{A}\mathbf{R}_{\mathbf{sy}} - \mathbf{R}_{\mathbf{sy}}^H \mathbf{A}^H + \mathbf{A}\mathbf{R}_{\mathbf{ss}} \mathbf{A}^H \right\}, \tag{10}$$

where H denotes the Hermitian transpose and $tr\{\cdot\}$ denotes the matrix trace, and the associated covariance matrices,

$$\begin{aligned} \mathbf{R}_{\mathbf{yy}} &= E\left[\mathbf{y}\mathbf{y}^H \right] \in \mathbb{C}^{M \times M}, \\ \mathbf{R}_{\mathbf{ss}} &= E\left[\mathbf{s}_{DAS} \mathbf{s}_{DAS}{}^H \right] \in \mathbb{C}^{L \times L}, \cdot \\ \mathbf{R}_{\mathbf{sy}} &= E\left[\mathbf{s}_{DAS} \mathbf{y}^H \right] \in \mathbb{C}^{L \times M}. \end{aligned} \tag{11}$$

By taking the derivative of (10) with respect to $\mathbf{A}^H$, we obtain

$$\nabla_{\mathbf{A}^H} J = -2\mathbf{R}_{\mathbf{sy}} + 2\mathbf{R}_{\mathbf{ss}} \mathbf{A}^H = 0, \cdot \tag{12}$$

The optimal solution can be obtained as:

$$\mathbf{A} = \mathbf{R}_{\mathbf{sy}}^H \mathbf{R}_{\mathbf{ss}}^{-1} \cdot \tag{13}$$

In practical implementation, the recursive averaging is adopted to obtain $\mathbf{R}_{\mathbf{sy}}$ and $\mathbf{R}_{\mathbf{ss}}$, as given by

$$\begin{aligned}\mathbf{R}_{\mathbf{ss}}(p) &= \alpha\mathbf{R}_{\mathbf{ss}}(p-1)+(1-\alpha)\mathbf{s}_{DAS}(p-1)\mathbf{s}_{DAS}{}^{H}(p-1),\\ \mathbf{R}_{\mathbf{sy}}(p) &= \alpha\mathbf{R}_{\mathbf{sy}}(p-1)+(1-\alpha)\mathbf{s}_{DAS}(p-1)\mathbf{y}^{H}(p-1)\end{aligned} \quad (14)$$

where $\mathbf{s}_{DAS}$ denotes the relatively clean source signals attained by utilizing DAS beamformer, p and α denotes frame index and the forgetting factor for the expectation process, respectively. The Wiener filtering method can be summarized as follows.

Algorithm 1 CTF estimation using Wiener filtering

Input: $\mathbf{y}(n)$, $\mathbf{s}_{DAS}(n)$

Initialize Forgetting factor α

Covariance matrices $\mathbf{R}_{\mathbf{ss}}$, $\mathbf{R}_{\mathbf{sy}}$

For each instant of frame, $n = 1, 2, \ldots$

Compute

$$\mathbf{R}_{ss}(n) = \alpha\mathbf{R}_{ss}(n-1)+(1-\alpha)\hat{\mathbf{s}}(n-1)\hat{\mathbf{s}}^{H}(n-1)$$

$$\mathbf{R}_{sy}(n) = \alpha\mathbf{R}_{sy}(n-1)+(1-\alpha)\hat{\mathbf{s}}(n-1)\mathbf{y}^{H}(n-1)$$

$$\mathbf{A}(n) = \mathbf{R}_{sy}^{H}(n)\mathbf{R}_{ss}^{-1}(n)$$

3.2. RLS approach

For the second technique, we estimate the CTF coefficients matrix using the adaptive filter algorithm. Notably, the RLS algorithm optimization process in [14] is being carried out in the complex domain now. The RLS algorithm minimizes the sum of the weighted error square as:

$$\min_{\mathbf{A}\in\mathbb{C}^{M\times L}} \sum_{i=1}^{n} \lambda^{n-i} \left\|\mathbf{y}(n)-\mathbf{A}(n)\mathbf{s}_{DAS}(n)\right\|_2^2, \quad (16)$$

where n represents both the adaptation iteration and the frame index, and λ represents the forgetting factor that weights the error with respect to the iteration. Based on the objective function outlined in (16), the RLS algorithm can be applied for the estimation of the CTF coefficient $\mathbf{A}$. The RLS approach can be summarized in following routine algorithm.

Algorithm 2 CTF estimation using RLS

Input: $\mathbf{y}(n)$, $\mathbf{s}_{DAS}(n)$, λ

Initialize RLS weight and inverse of correlation matrix:

$\mathbf{w}(0) = \mathbf{0} \in \mathbb{C}^{L\times M}$, $\mathbf{P}(0) = \varepsilon^{-1}\mathbf{I} \in \mathbb{C}^{L\times L}$,

where ε is a positive constant

For each instant of frame, $n = 1, 2, \ldots$

Compute

$$\mathbf{e}(n) = \mathbf{y}(n)^{H} - \mathbf{s}_{DAS}(n)\mathbf{w}(n-1)$$

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{s}_{DAS}(n)}{1+\lambda^{-1}\mathbf{s}_{DAS}(n)^{H}\mathbf{P}(n-1)\mathbf{s}_{DAS}(n)}$$

$$\mathbf{w}(n) = \mathbf{w}(n-1)+\mathbf{k}(n)\mathbf{e}(n)$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1)-\lambda^{-1}\mathbf{k}(n)\mathbf{s}_{DAS}(n)^{H}\mathbf{P}(n-1)$$

$$\mathbf{A}(n) = \mathbf{w}(n)^{H}$$

3.3. ATF reconstruction

Once we have estimated the CTF coefficients from each of two filters described above, we can move on to produce the ATFs using these coefficients. First, we generate a unit pulse sequence $\delta(n)$ which is delayed by (L – 1)D points as

$$\delta(n) = \begin{cases}1, & n=(L-1)D\\ 0, & \text{else } n\end{cases}. \quad (17)$$

Then, transform it to the STFT domain, obtaining δ_p. Finally, the estimated CTF coefficients are convolved with the transformed unit pulse sequence δ_p to obtain the signal below:

$$G_p^i = \sum_{p'=0}^{L-1} \delta_{p+L-p'}\, a_{p'}^i, \quad (18)$$

where $p \in [0, P_{ATF}]$, a indicates CTF coefficient obtained above and i indicates ordinal number of microphones. The estimated RIRs can be obtained through the inverse STFT with respect to G_p^i, and the estimated ATFs $\hat{\mathbf{a}}^i$ in vector form with each element representing different frequency bin are produced simply by performing the fast Fourier transform (FFT) with respect to the estimated RIRs. In summary, the flowchart of our proposed method can be presented as follows.

Flow chart

Step 1: Acquire microphone signal $\mathbf{y}(n)$

Step 2: $\mathbf{s}_{DAS}(n)$ obtained by WPE followed by DAS

Step 3: Initialize the parameters for algorithm 1 or 2

Step 4: Estimate the CTF coefficients $\mathbf{A}$

a. Wiener filtering algorithm

b. RLS algorithm

Step 5: Do (18) and reconstruct RIR via inverse STFT

Step 6: Applications: MINT, etc.

4. SIMULATIONS

4.1. Simulation setting and parameters

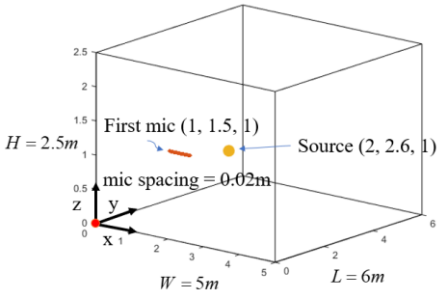

Fig. 1 configuration of the room for simulations

To assess the performance of the proposed method for estimating ATFs, we use MINT [15] to conduct the dereverberation by employing the estimated RIRs. For the comparative study, we also perform the dereverberation utilizing the state-of-the-art technique, WPE [11]. The RIR Generator [18] was employed to create the ground-truth RIRs. The room dimensions were 5 m × 6 m × 2.5 m. The first sensor of a 30-microphone ULA was positioned at (1 m, 1.5 m, 1 m) with a 0.02 m spacing along the x-axis. Figure 1 shows the configuration of the whole room. To generate microphone signals, speech signals sampled at 16 kHz were

used as sources convolved with the ground-truth RIRs. The speech sources were located at (2 m, 2.6 m, 1 m). Seven different reverberation times were validated in the simulation, i.e., T_{60} = 0.4s, 0.6s, 0.8s, 1.0s, 1.2s, 1.4s and 1.6s. To perform the STFT, we used a 1024-sample Hamming window (64 ms) with 75% overlap. The free parameters α, λ, and ε were fixed at a consistent value of 0.999, 0.99, and 0.01 throughout the simulations, as it proved to be suitable for all conditions.

4.2. Results and discussions

The absolute error of the estimated ATF across all frequency bins and the magnitude of the estimated RIR over its ground truth values with T_{60} = 0.6 s are plotted in Figure 2. The plot shows a low absolute error across all frequency bins and a remarkable correspondence between the magnitude of the estimated RIR and its ground-truth counterparts, thus satisfying our requirements. Figure 3 presents the Matching Error (ME) of the estimated ATFs for all reverberation times. The ME is formulated as follow:

$$\mathrm{ME} = \frac{1}{M}\sum_{i=1}^{M}\left\|\mathbf{a}^i - \hat{\mathbf{a}}^i\right\|_2, \qquad (19)$$

where $\mathbf{a}^i$ denotes the ground-truth ATF of the ith microphone. The data shows that the ME is consistently low across all reverberation times, indicating a positive outcome.

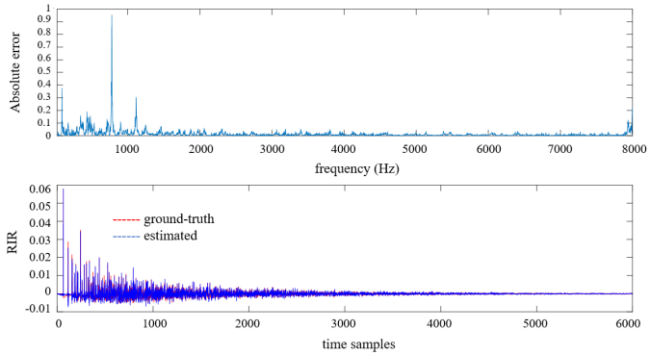

Fig. 2 Error of the estimated ATF and magnitude of the estimated RIR when T_{60} = 0.6s

Four types of signals in abbreviation, namely proposed method with RLS algorithm, proposed method with Wiener filter, WPE and unprocessed, were evaluated for their dereverberation performance metrics using the Perceptual Evaluation of Speech Quality (PESQ) [16] and Signal-to-Distortion Ratio (SDR) [17]. Figure 4 illustrates the obtained PESQ for the four reverberation times. It can be observed from the plot that the two proposed methods achieve a higher PESQ compared to WPE, except for T_{60} = 1.6 s, which is the limit of our proposed method. In addition, the PESQs of four signals decrease as the reverberation time increases. Figure 5 illustrates the SDR obtained for four reverberation times. It is evident in this graph that the proposed two methods still achieve higher SDR than WPE, which is a convincing result, and that the SDRs of all four signals decrease with increasing reverberation time.

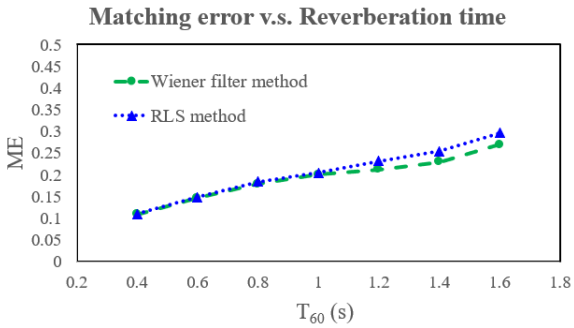

Fig. 3 ME of the estimated ATFs based on the proposed methods in various reverberation times

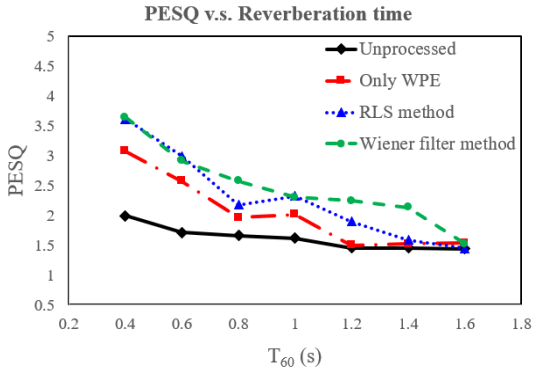

Fig. 4 The PESQ values of the processed and unprocessed signals at different reverberation time

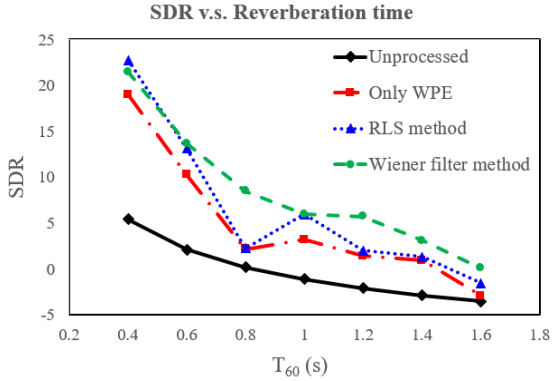

Fig. 5 The SDR values of the processed and unprocessed signals at different reverberation time

5. CONCLUSIONS

In this paper, a blind estimation method for ATF based on the CTF model is proposed. Two techniques for estimating CTF coefficient matrices are developed using the Wiener filter and RLS algorithm respectively. After comparing the magnitude of the estimated ATF with the ground-truth ATF, as well as the ME under varying reverberation times, we conclude that the proposed method achieves accurate ATF estimation. By using PESQ and SDR, we compared the scores obtained from the dereverberation signal produced by WPE and MINT which relies on the estimated RIR of our proposed method. The results indicate a significant advantage of our proposed method over WPE.

REFERENCES

[1] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "Advances in Network and Acoustic Echo Cancellation," *New York: Springer.,* 2001.

[2] Y. Huang, J. Benesty, and J. Chen, "A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.,* vol. 13, no. 5, pp. 882–895, 2005.

[3] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 14, no. 3, pp. 774–784, May 2006.

[4] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[5] Y. Inouye, and K. Hirano, "Cumulant-Based Blind Identification of Linear Multi-Input Multi-Output Systems Driven by Colored Inputs," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 45, no. 6, pp. 1543-1552, June 1997.

[6] J. K. Tugnait, "Adaptive blind separation of convolutive mixtures of independent linear signals," *Signal Process.,* vol. 73, pp. 139-152, 1999.

[7] B. Chen and A. Petropulu, "Frequency domain blind MIMO system identification based on second and higher order statistics," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 49, pp. 1677-1688, 2001.

[8] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *IEEE Signal Processing Letters.,* vol. 14, no. 5, pp. 337–340, 2007.

[9] Y. Avargel and I. Cohen, "System identification in the short-time Fourier transform domain with crossband filtering," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 15, no. 4, pp. 1305–1319, 2007.

[10] R. Talmon, I. Cohen, and S. Gannot, "Relative transfer function identification using convolutive transfer function approximation," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 17, no. 4, pp. 546–555, 2009.

[11] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi and B. -H. Juang, "Speech Dereverberation Based on Variance-Normalized Delayed Linear Prediction," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 18, no. 7, pp. 1717–1731, 2010.

[12] Harry L. Van Trees, "Optimum array processing: Part Iv of detection, estimation, and modulation theory," *New York: Wiley*, 2002.

[13] J. Benesty, S. Makino, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," *Speech enhancement*, pp. 9-41, 2005.

[14] S. A. U. Islam and D. S. Bernstein, "Recursive Least Squares for Real-Time Implementation," *IEEE Control Systems Magazine*, vol. 39, no. 3, pp. 82-85, June 2019, doi: 10.1109/MCS.2019.2900788.

[15] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 36, no. 2, pp. 145-152, Feb. 1988, doi: 10.1109/29.1509.

[16] A. W. Rix, J. G. Beerends, M. P. Hollier and A. P. Hekstra, "Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs," *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings*, pp. 749-752 vol.2, 200, doi: 10.1109/ICASSP.2001.941023.

[17] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. Audio, Speech, Lang. Process.,* vol. 14, no. 4, pp. 1462–1469, 2006

[18] Emanuël Habets, "Room Impulse Response Generator," *Internal Report*, pp. 1-17, 2006.