On Speech Enhancement Using Microphone Arrays in the Presence of Co-Directional Interference

Conference Paper · April 2018

DOI: 10.1109/ICASSP.2018.8462032

4 authors, including:


Jingdong Chen
Institute of Electrical and Electronics Engineers


Jacob Benesty
Institut National de la Recherche Scientifique


Israel Cohen
Technion - Israel Institute of Technology

ON SPEECH ENHANCEMENT USING MICROPHONE ARRAYS IN THE PRESENCE OF CO-DIRECTIONAL INTERFERENCE

Xin Leng[1], *Jingdong Chen*[1], *Jacob Benesty*[2], *and Israel Cohen*[3]

[1]Northwestern Polytechnical University, CIAIC and School of
Marine Science and Technology, Xi'an, China
[2]INRS-EMT, University of Quebec, Montreal, Canada
[3]Technion-Israel Institute of Technology, Department of Electrical Engineering, Haifa, Israel

ABSTRACT

Beamforming using microphone arrays has been widely used for enhancing speech signals of interest and suppressing noise and interference in a wide range of applications. In order to make it work, beamforming generally assumes that the speech source of interest and the interference source are incident to the array from different directions. In this paper, we study the case where both the speech and interference sources come from the same direction. A linearly constrained minimum variance (LCMV) beamformer is derived in this scenario based on the so-called widely linear (WL) estimation framework in the frequency domain. We analyze this beamformer and show how its performance depends on the second-order noncircularity of the desired speech and interference sources.

Index Terms— Beamforming, widely linear filtering, noncircularity, linearly constrained minimum variance (LCMV), speech enhancement.

1. INTRODUCTION

Microphone arrays have been widely used in various speech communication and human-machine interface systems, such as Amazon's Echo, Apple's HomePod, and iFlyTek's DingDong, to enhance speech signals from corruption of noise and interference. A core component of a microphone array system is the so-called beamforming, which is to estimate the signal of interest from microphone observations that consist of not only this signal but also noise, interference, and reverberation. Many different beamforming algorithms have been developed over the last few decades [1–3], which, by and large, can be classified into two categories: fixed beamformers (e.g., delay-and-sum, superdirective, and differential) and adaptive ones (e.g., linearly constrained minimum variance (LCMV) [4], generalized sidelobe canceller (GSC) [5–7], and minimum variance distortionless response (MVDR) [8]). While they differ in optimization principles, performance, and robustness with respect to noise, interference, and reverberation, those algorithms make a common assumption that the source and interference/noise are from different incidence angles. They suffer from significant performance degradation or may even fall apart if the incidence angles of the desired signal and interference/noise sources are close to each other.

In this paper, we investigate the scenario where the source of interest and the interference source are incident to the microphone array from the same direction. In other words, there is no spatial selectivity between the signal of interest and the interference. To deal with the problem of speech enhancement in this scenario, we develop a widely linear (WL) LCMV beamformer, which can be viewed as an extension of the work in [9–11]. Note that the developed algorithm also works or works better if the source and interference are from different directions.

This work was supported in part by the NSFC Distinguished Young Scientists Fund under Grant No. 61425005 and the NSFC-ISF joint research program under Grant No. 61761146001.

2. SIGNAL MODEL

In this paper, we explicitly focus on the problem of enhancing a speech signal of interest in the presence of co-directional interference. Let us consider the scenario where we use a uniform linear microphone array (ULMA) of M sensors for sound recording and there are a desired sound source and an interference source, both are in the far-field and incident to the array from the same direction, i.e., θ_d. If we choose the first microphone as the reference sensor, the signal received at the mth microphone can be written, in the frequency domain, as [12]

$$\begin{aligned} Y_m(\omega) &= X_m(\omega) + J_m(\omega) + V_m(\omega) \\ &= e^{-\jmath(m-1)\omega\tau_0\cos\theta_\mathrm{d}}\big[X(\omega)+J(\omega)\big] + V_m(\omega), \end{aligned} \tag{1}$$

where $X_m(\omega)$, $J_m(\omega)$, and $V_m(\omega)$ are, respectively, the speech signal component of interest, the interference component, and the background noise at the mth ($m = 1, 2, \ldots, M$) microphone, $\jmath$ is the imaginary unit with $\jmath^2 = -1$, ω is the angular frequency, $\tau_0 = \delta/c$ with δ being the spacing between two adjacent microphone sensors and c being the speed of sound in air, and $X(\omega)$ and $J(\omega)$ denote, respectively, the speech signal and interference components at the reference sensor. We further assume that all the signals are zero mean and broadband, and the desired signal, interference, and additive noise are statistically uncorrelated with each other.

Putting all the signals $Y_m(\omega)$'s into a vector form, we get

$$\begin{aligned} \mathbf{y}(\omega) &\triangleq \big[Y_1(\omega)\ Y_2(\omega)\ \cdots\ Y_M(\omega)\big]^T \\ &= \mathbf{x}(\omega) + \mathbf{j}(\omega) + \mathbf{v}(\omega) \\ &= \mathbf{d}(\omega, \cos\theta_\mathrm{d})\big[X(\omega)+J(\omega)\big] + \mathbf{v}(\omega), \end{aligned} \tag{2}$$

where superscript ^{T} denotes transpose of a vector or matrix,

$$\mathbf{d}(\omega, \cos\theta_\mathrm{d}) \triangleq \big[1\ e^{-\jmath\omega\tau_0\cos\theta_\mathrm{d}}\ \cdots\ e^{-\jmath(M-1)\omega\tau_0\cos\theta_\mathrm{d}}\big]^T \tag{3}$$

is the steering vector, $\mathbf{x}(\omega) \triangleq \mathbf{d}(\omega, \cos\theta_\mathrm{d})X(\omega)$, $\mathbf{j}(\omega) \triangleq \mathbf{d}(\omega, \cos\theta_\mathrm{d})J(\omega)$, and $\mathbf{v}(\omega)$ is defined analogously to $\mathbf{y}(\omega)$.

3. PROBLEM FORMULATION

With the signal model in (2), the objective of beamforming is to design a filter that can best recover $X(\omega)$ given the signal vector $\mathbf{y}(\omega)$.

Since we deal with speech and acoustic signals in the frequency (i.e., short-time Fourier transform) domain, which are nonstationary and noncircular, beamforming is achieved using the WL estimation theory [13–15], i.e.,

$$\begin{aligned}\widehat{X}(\omega) &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) + \mathbf{h}'^H(\omega)\mathbf{y}^*(\omega) \\ &= \widetilde{\mathbf{h}}^H(\omega)\widetilde{\mathbf{y}}(\omega) \\ &= \widetilde{\mathbf{h}}^H(\omega)\big[\widetilde{\mathbf{x}}(\omega) + \widetilde{\mathbf{j}}(\omega) + \widetilde{\mathbf{v}}(\omega)\big], \end{aligned} \tag{4}$$

where superscript ^{H} denotes the conjugate-transpose operator,

$$\mathbf{h}(\omega) \triangleq \big[H_1(\omega)\ H_2(\omega)\ \cdots\ H_M(\omega)\big]^T \tag{5}$$

$$\mathbf{h}'(\omega) \triangleq \big[H_1'(\omega)\ H_2'(\omega)\ \cdots\ H_M'(\omega)\big]^T \tag{6}$$

are two complex finite-impulse response (FIR) filters, both of length M,

$$\widetilde{\mathbf{h}}(\omega) \triangleq \begin{bmatrix}\mathbf{h}(\omega) \\ \mathbf{h}'(\omega)\end{bmatrix} \tag{7}$$

is called the augmented WL filter of length $2M$,

$$\widetilde{\mathbf{y}}(\omega) \triangleq \begin{bmatrix}\mathbf{y}(\omega) \\ \mathbf{y}^*(\omega)\end{bmatrix} \tag{8}$$

is the augmented noisy signal vector also of length $2M$, and $\widetilde{\mathbf{x}}(\omega)$, $\widetilde{\mathbf{j}}(\omega)$, and $\widetilde{\mathbf{v}}(\omega)$ are defined analogously to $\widetilde{\mathbf{y}}(\omega)$. If $\mathbf{h}'(\omega) = \mathbf{0}_M$ (where $\mathbf{0}_M$ is a zero vector of size $M \times 1$) for any frequency ω, the WL beamforming in (4) degenerates to the classical beamforming; but this is generally not true for speech signals [15].

For the noncircular speech signal, $X(\omega)$, we can decompose its conjugate as [16, 17]

$$X^*(\omega) = \gamma_X^*(\omega)X(\omega) + X'(\omega), \tag{9}$$

where

$$X'(\omega) = X^*(\omega) - \gamma_X^*(\omega)X(\omega), \tag{10}$$

$$E[X'(\omega)X^*(\omega)] = 0, \tag{11}$$

and

$$\gamma_X(\omega) = \frac{E[X^2(\omega)]}{E[|X(\omega)|^2]} \tag{12}$$

is the (second-order) circularity quotient of $X(\omega)$, which satisfies $0 \leqslant |\gamma_X(\omega)| \leqslant 1$ [16, 17], and $E[\cdot]$ denotes mathematical expectation. Using (9), we can write the vector $\widetilde{\mathbf{x}}(\omega)$ as

$$\widetilde{\mathbf{x}}(\omega) = \mathbf{d}_X(\omega, \cos\theta_\mathrm{d})X(\omega) + \widetilde{\mathbf{x}}'(\omega), \tag{13}$$

where

$$\begin{aligned}\mathbf{d}_X(\omega, \cos\theta_\mathrm{d}) &\triangleq \begin{bmatrix}\mathbf{d}(\omega, \cos\theta_\mathrm{d}) \\ \gamma_X^*(\omega)\mathbf{d}^*(\omega, \cos\theta_\mathrm{d})\end{bmatrix} \\ &= \frac{E[\widetilde{\mathbf{x}}(\omega)X^*(\omega)]}{E[|X(\omega)|^2]}, \end{aligned} \tag{14}$$

$$\widetilde{\mathbf{x}}'(\omega) \triangleq \begin{bmatrix}\mathbf{0}_M \\ X'(\omega)\mathbf{d}^*(\omega, \cos\theta_\mathrm{d})\end{bmatrix}. \tag{15}$$

Similarly, we have

$$J^*(\omega) = \gamma_J^*(\omega)J(\omega) + J'(\omega), \tag{16}$$

$$\widetilde{\mathbf{j}}(\omega) = \mathbf{d}_J(\omega, \cos\theta_\mathrm{d})J(\omega) + \widetilde{\mathbf{j}}'(\omega), \tag{17}$$

where $\gamma_J(\omega)$, $J'(\omega)$, $\mathbf{d}_J(\omega, \cos\theta_\mathrm{d})$, and $\widetilde{\mathbf{j}}'(\omega)$ are defined, respectively, analogously to $\gamma_X(\omega)$, $X'(\omega)$, $\mathbf{d}_X(\omega, \cos\theta_\mathrm{d})$, and $\widetilde{\mathbf{x}}'(\omega)$.

Combining (13) and (17), one can rewrite (4) as

$$\begin{aligned}\widehat{X}(\omega) = \widetilde{\mathbf{h}}^H(\omega)\big[&\mathbf{d}_X(\omega, \cos\theta_\mathrm{d})X(\omega) + \widetilde{\mathbf{x}}'(\omega) \\ &+ \mathbf{d}_J(\omega, \cos\theta_\mathrm{d})J(\omega) + \widetilde{\mathbf{j}}'(\omega) + \widetilde{\mathbf{v}}(\omega)\big]. \end{aligned} \tag{18}$$

The problem of beamforming now becomes one of finding an optimal filter $\widetilde{\mathbf{h}}(\omega)$ so that $\widehat{X}(\omega)$ is a good estimate of $X(\omega)$.

4. PERFORMANCE MEASURES

In this section, we describe four important measures: signal-to-noise ratio (SNR), signal-to-interference ratio (SIR), signal-to-interference-plus-noise ratio (SINR), and beampatten, which will be used in the subsequent sections to evaluate the performance of WL beamformers.

4.1. SNR, SIR, and SINR

With the signal model (18), the input SNR, SIR, and SINR are defined, respectively, as

$$\mathrm{iSNR}(\omega) \triangleq \frac{\phi_X(\omega)}{\phi_V(\omega)}, \tag{19}$$

$$\mathrm{iSIR}(\omega) \triangleq \frac{\phi_X(\omega)}{\phi_J(\omega)}, \tag{20}$$

$$\mathrm{iSINR}(\omega) \triangleq \frac{\phi_X(\omega)}{\phi_V(\omega) + \phi_J(\omega)}, \tag{21}$$

where $V(\omega)$ is the noise received at the reference microphone, $\phi_X(\omega) \triangleq E[|X(\omega)|^2]$, $\phi_V(\omega) \triangleq E[|V(\omega)|^2]$, and $\phi_J(\omega) \triangleq E[|J(\omega)|^2]$ are the variances of $X(\omega)$, $V(\omega)$, and $J(\omega)$, respectively.

The output SINR is defined according to (18) as

$$\begin{aligned}\mathrm{oSINR}[\widetilde{\mathbf{h}}(\omega)] &\triangleq \frac{\phi_X(\omega)\big|\widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_X(\omega, \cos\theta_\mathrm{d})\big|^2}{\widetilde{\mathbf{h}}^H(\omega)\big[\mathbf{\Phi}_{\widetilde{\mathbf{x}}'}(\omega) + \mathbf{\Phi}_{\widetilde{\mathbf{j}}}(\omega) + \mathbf{\Phi}_{\widetilde{\mathbf{v}}}(\omega)\big]\widetilde{\mathbf{h}}(\omega)} \\ &= \frac{\phi_X(\omega)\big|\widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_X(\omega, \cos\theta_\mathrm{d})\big|^2}{\widetilde{\mathbf{h}}^H(\omega)\mathbf{\Phi}_\mathrm{in}(\omega)\widetilde{\mathbf{h}}(\omega)}, \end{aligned} \tag{22}$$

where $\mathbf{\Phi}_{\widetilde{\mathbf{x}}'}(\omega) \triangleq E[\widetilde{\mathbf{x}}'(\omega)\widetilde{\mathbf{x}}'^H(\omega)]$, $\mathbf{\Phi}_{\widetilde{\mathbf{j}}}(\omega) \triangleq E[\widetilde{\mathbf{j}}(\omega)\widetilde{\mathbf{j}}^H(\omega)]$, and $\mathbf{\Phi}_{\widetilde{\mathbf{v}}}(\omega) \triangleq E[\widetilde{\mathbf{v}}(\omega)\widetilde{\mathbf{v}}^H(\omega)]$ are the covariance matrices of $\widetilde{\mathbf{x}}'(\omega)$, $\widetilde{\mathbf{j}}(\omega)$, and $\widetilde{\mathbf{v}}(\omega)$, respectively, and

$$\mathbf{\Phi}_\mathrm{in}(\omega) = \mathbf{\Phi}_{\widetilde{\mathbf{x}}'}(\omega) + \mathbf{\Phi}_{\widetilde{\mathbf{j}}}(\omega) + \mathbf{\Phi}_{\widetilde{\mathbf{v}}}(\omega) \tag{23}$$

is the interference-plus-noise covariance matrix.

Using (21) and (22), the array gain, which quantified the SINR improvement, is then defined as

$$\begin{aligned}\mathcal{A}[\widetilde{\mathbf{h}}(\omega)] &\triangleq \frac{\mathrm{oSINR}[\widetilde{\mathbf{h}}(\omega)]}{\mathrm{iSINR}(\omega)} \\ &= \frac{\big|\widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_X(\omega, \cos\theta_\mathrm{d})\big|^2\big[\phi_V(\omega) + \phi_J(\omega)\big]}{\widetilde{\mathbf{h}}^H(\omega)\mathbf{\Phi}_\mathrm{in}(\omega)\widetilde{\mathbf{h}}(\omega)}. \end{aligned} \tag{24}$$

4.2. Beampattern

The beampattern describes the array's response to a plane wave impinging from different directions. For a WL beamformer, we define its beampattern as

$$\begin{aligned}\mathcal{B}[\widetilde{\mathbf{h}}(\omega), \theta] &\triangleq \mathbf{d}_X^H(\omega, \cos\theta)\widetilde{\mathbf{h}}(\omega) \\ &= \mathbf{d}^H(\omega, \cos\theta)\mathbf{h}(\omega) + \gamma_X(\omega)\mathbf{d}^T(\omega, \cos\theta)\mathbf{h}'(\omega) \\ &= \sum_{n=-M+1}^{M-1} a_n(\omega) e^{\jmath n\omega\tau_0 \cos\theta},\end{aligned} \tag{25}$$

where

$$a_n(\omega) \triangleq \begin{cases} H_{n+1}(\omega), & n > 0 \\ H_1(\omega) + \gamma_X(\omega)H_1'(\omega), & n = 0 \\ \gamma_X(\omega)H_{-n+1}'(\omega), & n < 0 \end{cases} \tag{26}$$

are complex coefficients. Different values of these coefficients determine a different beampattern.

We can also rewrite (25) as

$$\mathcal{B}[\widetilde{\mathbf{h}}(\omega), \theta] = \mathcal{B}[\mathbf{h}(\omega), \theta] + \gamma_X(\omega)\mathcal{B}[\mathbf{h}'(\omega), \pi - \theta], \tag{27}$$

where

$$\begin{aligned}\mathcal{B}[\mathbf{h}(\omega), \theta] &\triangleq \mathbf{d}^H(\omega, \cos\theta)\mathbf{h}(\omega) \\ &= \sum_{m=1}^{M} H_m(\omega) e^{\jmath(m-1)\omega\tau_0\cos\theta}\end{aligned} \tag{28}$$

and

$$\begin{aligned}\mathcal{B}[\mathbf{h}'(\omega), \pi - \theta] &\triangleq \mathbf{d}^T(\omega, \cos\theta)\mathbf{h}'(\omega) \\ &= \sum_{m=1}^{M} H_m'(\omega) e^{\jmath(m-1)\omega\tau_0\cos(\pi-\theta)}\end{aligned} \tag{29}$$

are, respectively, the Mth-order traditional beampatterns with respect to the incident angle of θ and $\pi - \theta$. The beampattern $\mathcal{B}[\mathbf{h}'(\omega), \pi - \theta]$ quantifies the degree of the orthogonal component $\widetilde{\mathbf{x}}'(\omega)$ is rejected by $\widetilde{\mathbf{h}}(\omega)$. Ideally, if $\mathcal{B}[\mathbf{h}'(\omega), \pi - \theta] = 0$ as studied in [14], the $\widetilde{\mathbf{x}}'(\omega)$ will be rejected whatever the correlation between $X(\omega)$ and $X^*(\omega)$.

5. WIDELY LINEAR LCMV BEAMFORMER

Different WL beamformers can be derived according to (18). In this section, we consider deriving the WL LCMV beamformer by minimizing the variance of the filtered interference-plus-noise at the beamformer's output with the constraint that the desired signal is passing through without any distortion and the interference is suppressed to zero. Specifically,

$$\begin{aligned}&\arg\min_{\widetilde{\mathbf{h}}(\omega)} \widetilde{\mathbf{h}}^H(\omega)\mathbf{\Phi}_{\mathrm{in}}(\omega)\widetilde{\mathbf{h}}(\omega) \\ &\quad \text{s.t.} \quad \widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_X(\omega, \cos\theta_{\mathrm{d}}) = 1 \\ &\qquad\quad\;\; \widetilde{\mathbf{h}}^H(\omega)\mathbf{d}_J(\omega, \cos\theta_{\mathrm{d}}) = 0\end{aligned} \tag{30}$$

from which the solution is

$$\widetilde{\mathbf{h}}_{\mathrm{LCMV}}(\omega) = \mathbf{\Phi}_{\mathrm{in}}^{-1}(\omega)\mathbf{D}(\omega)\left[\mathbf{D}^H(\omega)\mathbf{\Phi}_{\mathrm{in}}^{-1}(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{i}_{2,1}, \tag{31}$$

where

$$\mathbf{D}(\omega, \cos\theta_{\mathrm{d}}) = \left[\mathbf{d}_X(\omega, \cos\theta_{\mathrm{d}}) \;\; \mathbf{d}_J(\omega, \cos\theta_{\mathrm{d}})\right] \tag{32}$$

is the constraint matrix of size $2M \times 2$ and

$$\mathbf{i}_{2,1} = \begin{bmatrix}1 & 0\end{bmatrix}^T. \tag{33}$$

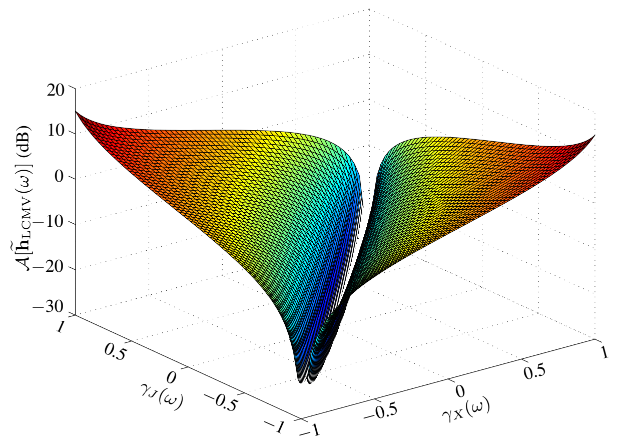

Fig. 1. Theoretical array gain of the proposed WL LCMV beamformer as a function of $\gamma_X(\omega)$ and $\gamma_J(\omega)$.

Obviously, we can rewrite the LCMV beamformer as

$$\widetilde{\mathbf{h}}_{\mathrm{LCMV}}(\omega) = \mathbf{\Phi}_{\widetilde{\mathbf{y}}}^{-1}(\omega)\mathbf{D}(\omega)\left[\mathbf{D}^H(\omega)\mathbf{\Phi}_{\widetilde{\mathbf{y}}}^{-1}(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{i}_{2,1}. \tag{34}$$

Substituting (31) into (22) and (24), we deduce that the output SINR and array gain of the proposed WL LCMV beamformer are

$$\mathrm{oSINR}[\widetilde{\mathbf{h}}_{\mathrm{LCMV}}(\omega)] = \frac{\phi_X(\omega)}{\mathbf{i}_{2,1}^T\left[\mathbf{D}^H(\omega)\mathbf{\Phi}_{\mathrm{in}}^{-1}(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{i}_{2,1}}, \tag{35}$$

$$\mathcal{A}[\widetilde{\mathbf{h}}_{\mathrm{LCMV}}(\omega)] = \frac{\phi_V(\omega) + \phi_J(\omega)}{\mathbf{i}_{2,1}^T\left[\mathbf{D}^H(\omega)\mathbf{\Phi}_{\mathrm{in}}^{-1}(\omega)\mathbf{D}(\omega)\right]^{-1}\mathbf{i}_{2,1}}. \tag{36}$$

With the WL LCMV beamformer, we can find

$$\begin{aligned}\mathcal{B}[\mathbf{h}_{\mathrm{LCMV}}(\omega), \theta_{\mathrm{d}}] &= 1 - \gamma_X(\omega)\mathcal{B}[\mathbf{h}_{\mathrm{LCMV}}'(\omega), \pi - \theta_{\mathrm{d}}] \\ &= -\gamma_J(\omega)\mathcal{B}[\mathbf{h}_{\mathrm{LCMV}}'(\omega), \pi - \theta_{\mathrm{d}}].\end{aligned} \tag{37}$$

As a result, we have

$$\mathcal{B}[\mathbf{h}_{\mathrm{LCMV}}(\omega), \theta_{\mathrm{d}}] = \frac{-\gamma_J(\omega)}{\gamma_X(\omega) - \gamma_J(\omega)}, \tag{38}$$

$$\mathcal{B}[\mathbf{h}_{\mathrm{LCMV}}'(\omega), \pi - \theta_{\mathrm{d}}] = \frac{1}{\gamma_X(\omega) - \gamma_J(\omega)}, \tag{39}$$

from which we can see that the reduction of $\widetilde{\mathbf{x}}'(\omega)$ and $\widetilde{\mathbf{j}}'(\omega)$ depends on $\gamma_X(\omega) - \gamma_J(\omega)$.

5.1. Case Study: Spatially White Noise

To illustrate the array gain performance, let us assume the $v_m(t)$ are spatially white, identically distributed, and circular Gaussian noise. In this case, we have $\mathbf{\Phi}_{\widetilde{\mathbf{v}}}(\omega) = \sigma^2\mathbf{I}_{2M}$, where $\sigma^2 \triangleq E[|V(\omega)|^2]$, and $\mathbf{I}_{2M}$ is the identity matrix of size $2M \times 2M$. With some mathematical manipulation, the array gain is deduced as in (40).

Figure 1 plots the array gain as a function of $\gamma_X(\omega)$ and $\gamma_J(\omega)$, where 8 microphones are used and the input SNR and SIR are both set to $0\,\mathrm{dB}$. For ease of illustration, $\gamma_X(\omega)$ and $\gamma_J(\omega)$ are assumed to be real numbers. It is seen from Fig. 1 that the larger the difference between the values of $\gamma_X(\omega)$ and $\gamma_J(\omega)$, the higher the obtained the array gain.

6. SIMULATIONS

In this simulation, we use the well-known image model [18] to simulate acoustic environments. We consider a room of size $4\,\mathrm{m} \times 4\,\mathrm{m} \times$

$$\mathcal{A}[\widetilde{\mathbf{h}}_{\mathrm{LCMV}}(\omega)] = \frac{M|\gamma_X(\omega) - \gamma_J(\omega)|^2 \left[\mathrm{iSNR}^{-1}(\omega) + \mathrm{iSIR}^{-1}(\omega)\right]}{(1 + |\gamma_J(\omega)|^2)\mathrm{iSNR}^{-1}(\omega) + M\left[1 - |\gamma_X(\omega)|^2 + \mathrm{iSIR}^{-1}(\omega)(1 - |\gamma_J(\omega)|^2)\right]} \tag{40}$$

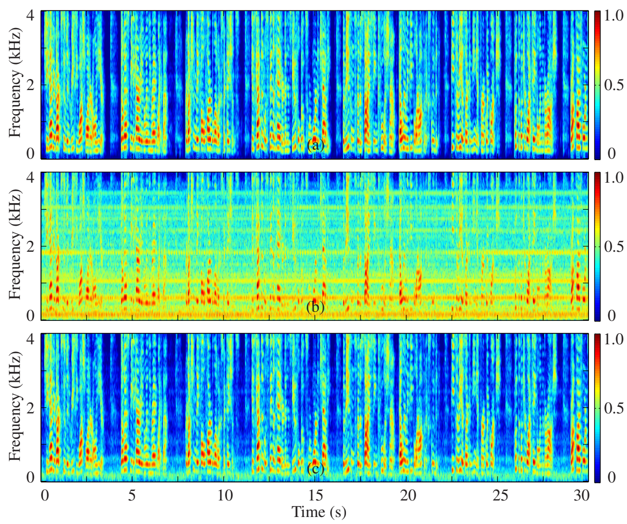

Fig. 2. Spectrograms of the (a) clean, (b) noisy, and (c) enhanced speech in car noise. The input SIR is $5\,\mathrm{dB}$.

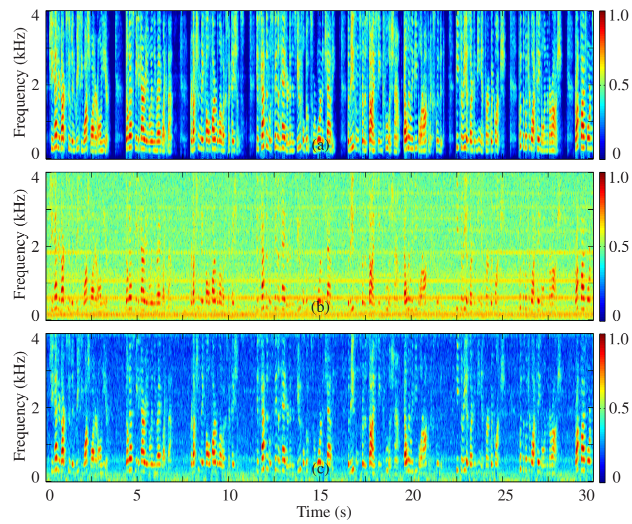

Fig. 3. Spectrograms of the (a) clean, (b) noisy, and (c) enhanced speech in car-plus-Gaussian noise. The input SIR is $5\,\mathrm{dB}$, and input SNR is $10\,\mathrm{dB}$.

$3\,\mathrm{m}$, where reflection coefficients of the six walls are set to 0.8. An ULMA with 7 omnidirectional sensors is configured. The positions of the 7 microphones are placed from (1.7, 1.0, 1.5) (in meters) to (2.3, 1.0, 1.5) with a spacing of $0.1\,\mathrm{m}$. The desired source, which is taken from the TIMIT database [19], is from (2.0, 2.0, 1.5). To simulate an interference source, a car noise is played back from (2.0, 2.5, 1.5). The microphone signals are generated by convolving the speech and car noise with the corresponding impulse responses and white Gaussian noise is then added to the convolution results to control the input SINR level. All the signals are sampled to $8\,\mathrm{kHz}$. The total length of the signal is $30\,\mathrm{s}$. Note that we compute the covariance matrices directly from the noise and noisy signals using a recursive method. Besides, the regularization method [20] is used while computing the inverse matrix.

In the first experiment, we consider a simple scenario where only the car noise is present. To implement the proposed WL LCMV beamformer, we use the overlap-add technique with a Kaiser window applied in both the analysis and synthesis steps. The short-time frame length is $16\,\mathrm{ms}$ and the overlap between neighboring frames is 75%. The beamformer in each subband is designed according to (34) and then applied to the noisy signal to reconstruct the enhanced speech in the time domain. Figure 2 plots the spectrograms of the clean, noisy, and enhanced speech at the reference microphone, where the input SIR is set to $5\,\mathrm{dB}$. One can clearly see that the speech signal has been recovered and a significant amount of interference is rejected. The corresponding output SINR is $12.43\,\mathrm{dB}$. The gain in perceptual evaluation of speech quality (PESQ) is 1.36, which is dramatic.

Figure 3 plots the speech spectrograms of a more generic case where the car and Gaussian noise are both present. The input SIR is $5\,\mathrm{dB}$, and input SNR is $10\,\mathrm{dB}$. Therefore, the overall input SINR is $3.8\,\mathrm{dB}$. It is clear that the proposed beamformer (34) has enhanced the speech spectrogram by rejecting both the car and the white noise. The corresponding output SINR is $8.29\,\mathrm{dB}$, and the PESQ gain is 0.82.

7. CONCLUSIONS

In this paper, we investigated the problem of beamforming with microphone arrays in the presence of noise and co-directional interference. In such a scenario, traditional beamformers would suffer from significant performance degradation as there is no spatial selectivity between the source of interest and the interference. To deal with the problem, we developed a WL LCMV beamformer, which can preserve the signal of interest while reducing interference that propagates to the array from the same direction as the desired source signal. An illustration with spatially white noise and two simulations were presented to demonstrate the effectiveness of the developed beamformer.

8. RELATION TO PRIOR WORK

In this paper, we study the problem of speech enhancement using microphone arrays in the presence of co-directional interference. The approach taken here is based on the WL estimation theory. This theory was first introduced by [13] to deal with complex random variables (CRVs). Since then, it has received a great amount of research interest in many respects [21, 22], such as noise reduction [15–17, 23, 24], interference cancellation [25], and echo cancelation [26–28]. In terms of beamforming, a WL MVDR beamformer was developed in [14] for dealing with second-order noncircular interference. This method was subsequently extended to an improved version by considering the second-order noncircularities of both the desired signal and interferences [9–11]. Some effort was also devoted to this WL beamforming to improve beamforming robustness [29–31]. The work in this paper can be viewed as an extension of the work in [9–11]. The major focus is placed on studying the case where the desired and interference sources are incident from the same direction while it is obvious that the method in this paper should work if the desired and interference sources are incident from different directions as assumed in the traditional as well as in WL beamforming.

9. REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2008, ch. 47, pp. 945–978.

[3] G. Huang, J. Benesty, and J. Chen, "Superdirective beamforming based on the Krylov matrix," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 12, pp. 2531–2543, Oct. 2016.

[4] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[5] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[6] C. W. Jim, "A comparison of two LMS constrained optimal array structures," *Proc. IEEE*, vol. 65, no. 12, pp. 1730–1731, Dec. 1977.

[7] B. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.

[8] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[9] P. Chevalier, J.-P. Delmas, and A. Oukaci, "Optimal widely linear MVDR beamforming for noncircular signals," in *Proc. IEEE ICASSP*, 2009, pp. 3573–3576.

[10] ——, "Performance analysis of the optimal widely linear MVDR beamformer," in *Proc. EUSIPCO*, 2009, pp. 587–591.

[11] ——, "Properties, performance and practical interest of the widely linear MMSE beamformer for nonrectilinear signals," *Signal Process.*, vol. 97, pp. 269–281, Apr. 2014.

[12] C. Pan, J. Chen, and J. Benesty, "Performance study of the MVDR beamformer as a function of the source incidence angle," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 67–79, Jan. 2014.

[13] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, no. 8, pp. 2030–2033, Aug. 1995.

[14] P. Chevalier and A. Blin, "Widely linear MVDR beamformers for the reception of an unknown signal corrupted by noncircular interferences," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5323–5336, Nov. 2007.

[15] J. Benesty, J. Chen, and Y. Huang, "Study of the widely linear Wiener filter for noise reduction," in *Proc. IEEE ICASSP*, 2010, pp. 205–208.

[16] ——, "On widely linear Wiener and tradeoff filters for noise reduction," *Speech Commun.*, vol. 52, no. 5, pp. 427–439, May 2010.

[17] ——, "A widely linear distortionless filter for single-channel noise reduction," *IEEE Signal Process. Lett.*, vol. 17, no. 5, pp. 469–472, May 2010.

[18] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[19] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, *TIMIT acoustic-phonetic continuous speech corpus*. Linguistic Data Consort., 1993.

[20] Y. Huang and J. Benesty, "A multi-frame approach to the frequency-domain single-channel noise reduction problem," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 4, pp. 1256–1269, May 2012.

[21] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[22] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[23] J. Benesty, J. Chen, and Y. Huang, "Binaural noise reduction in the time domain with a stereo setup," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 8, pp. 2260–2272, Nov. 2011.

[24] J. Chen and J. Benesty, "On the time-domain widely linear LCMV filter for noise reduction with a stereo system," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 21, no. 7, pp. 1343–1354, Jul. 2013.

[25] P. Chevalier and F. Pipon, "New insights into optimal widely linear array receivers for the demodulation of BPSK, MSK, and GMSK signals corrupted by noncircular interferences-application to SAIC," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 870–883, Mar. 2006.

[26] D. P. Mandic and V. S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. New York: Wiley, 2009, vol. 59.

[27] J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, *A Perspective on Stereophonic Acoustic Echo Cancellation*. Berlin, Germany: Springer-Verlag, 2011, vol. 4.

[28] C. Stanciu, J. Benesty, C. Paleologu, T. Gänsler, and S. Ciochină, "A widely linear model for stereophonic acoustic echo cancellation," *Signal Process.*, vol. 93, no. 2, pp. 511–516, 2013.

[29] D. Xu, L. Huang, X. Xu, and Z. Ye, "Widely linear MVDR beamformers for noncircular signals based on time-averaged second-order noncircularity coefficient estimation," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3219–3227, Sept. 2013.

[30] F. Wen, Q. Wan, H. Wei, R. Fan, and Y. Luo, "Robust Capon beamforming exploiting the second-order noncircularity of signals," *Signal Process.*, vol. 102, pp. 100–111, Sept. 2014.

[31] D. Xu, C. Gong, S. Cao, X. Xu, and Z. Ye, "Robust widely linear beamforming based on spatial spectrum of noncircularity coefficient," *Signal Process.*, vol. 104, pp. 167–173, Nov. 2014.