THE OPTIMAL WIDELY LINEAR MVDR BEAMFORMER IN ROOM ACOUSTICS

Jacob Benesty[1], *Jingdong Chen*[2], *and Yiteng (Arden) Huang*[2]

[1]: INRS-EMT, University of Quebec
800 de la Gauchetiere Ouest, Suite 6900
Montreal, QC H5A 1K6, Canada

[2]: WeVoice, Inc.
1065 Route 22 West, Suite 2E
Bridgewater, NJ 08807, USA

ABSTRACT

This paper develops, in a rigorous way, the optimal frequency-domain widely linear minimum variance distortionless response (MVDR) beamformer in the context of room acoustics. Therefore, all second-order statistics of complex random signals will be taken into account. We also develop the most useful performance measures needed to evaluate the derived MVDR filter.

Index Terms— Minimum variance distortionless response (MVDR), beamforming, speech enhancement, frequency domain, noncircularity, widely linear estimation.

1. INTRODUCTION

The minimum variance distortionless response (MVDR) beamformer for speech applications is usually formulated in the frequency domain [1]. Frequency-domain signals are generally complex random variables, even though the original time-domain signals are real in the context of room acoustics. The main concern is then, how to design the optimal frequency-domain MVDR filter that can fully exploit the different statistics of the involved complex components.

A very important statistical characteristic of a complex random variable (CRV) is the so-called circularity property or lack of it (noncircularity) [2], [3]. A zero-mean CRV, A, is circular if and only if the nonnull moments and cumulants are the only moments and cumulants constructed with the same power in A and A^* [4], [5], where the superscript * denotes complex conjugation. In particular, A is said to be a second-order circular CRV (CCRV) if its so-called pseudo-variance [2] is equal to zero, i.e., $E\left(A^2\right) = 0$, where $E(\cdot)$ denotes mathematical expectation, while its variance is nonnull, i.e., $E\left(|A|^2\right) \neq 0$. A good measure of the second-order circularity is the circularity quotient [2] defined as the ratio between the pseudo-variance and the variance, i.e.,

$$\gamma_A = \frac{E\left(A^2\right)}{E\left(|A|^2\right)}. \tag{1}$$

It is easy to show that $0 \leq |\gamma_A| \leq 1$. We see from the above that the second-order behavior of a CCRV is well described by its variance. Note that the Fourier components of stationary signals are CCRVs [6]. Another powerful aspect of the second-order circularity is that the classical linear estimation theory for real random variables straightforwardly applies to CCRVs. However, the frequency-domain components of some nonstationary signals (e.g., speech) are not circular variables [7]. Then, many natural questions arise: is the noncircularity useful in beamforming for speech applications? If so, how do we use the noncircularity? How much this information can improve the beamforming performance? What is the optimal frequency-domain MVDR filter for noncircular CRVs? In this paper, we don't attempt to give answers to all these questions but we only focus on the rigorous derivation of the MVDR in the frequency domain. For that, we will investigate the use of the so-called widely linear (WL) estimation theory [8] where two complex weight filters are applied, one to the frequency-domain observation signals and the other one to the complex conjugate of the same signals[1]. In particular, we will re-derive in the context of speech enhancement in room acoustics, the optimal WL MVDR beamformer proposed in [9] in the context of communications.

2. SIGNAL MODEL

We consider the conventional signal model in which an N-element microphone array captures a convolved source signal in some noise field. The received signals, at the discrete time-index k, are expressed as [1], [10], [11]

$$\begin{aligned} y_n(k) &= g_n(k) * s(k) + v_n(k) \\ &= x_n(k) + v_n(k),\ n = 1, 2, \ldots, N, \end{aligned} \tag{2}$$

where $g_n(k)$ is the impulse response from the unknown source $s(k)$ to the nth microphone, $*$ stands for linear convolution, and $v_n(k)$ is the additive noise at microphone n. We assume that the signals $x_n(k)$ and $v_n(k)$ are uncorrelated and zero mean. By definition, $x_n(k)$ is coherent across the array. The noise signals $v_n(k)$ are typically only partially coherent across the array. All previous signals are considered to be real, nonstationary, and broadband.

In this paper, our desired signal is designated by the clean (but convolved) signal received at microphone 1, namely $x_1(k)$. The problem statement may be posed as follows [1]: given N mixtures of two uncorrelated signals $x_n(k)$ and $v_n(k)$, our aim is to preserve $x_1(k)$ while minimizing the contribution of the noise terms $v_n(k)$ in the array output.

In the frequency domain, (2) can be rewritten as

$$\begin{aligned} Y_n(f) &= G_n(f)S(f) + V_n(f) \\ &= X_n(f) + V_n(f),\ n = 1, 2, \ldots, N, \end{aligned} \tag{3}$$

where $Y_n(f)$, $G_n(f)$, $S(f)$, $X_n(f) = G_n(f)S(f)$, and $V_n(f)$ are the frequency-domain representations of $y_n(k)$, $g_n(k)$, $s(k)$, $x_n(k)$, and $v_n(k)$, respectively, at temporal frequency f.

The N microphone signals in the frequency domain are better summarized in a vector notation as

$$\begin{aligned} \mathbf{y}(f) &= \mathbf{g}(f)S(f) + \mathbf{v}(f) \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}(f)X_1(f) + \mathbf{v}(f), \end{aligned} \tag{4}$$

[1]We recall that only one complex weight filter is used in the classical linear estimation theory

where

$$\begin{aligned}
\mathbf{y}(f) &= \begin{bmatrix} Y_1(f) & Y_2(f) & \cdots & Y_N(f) \end{bmatrix}^T, \\
\mathbf{x}(f) &= \begin{bmatrix} X_1(f) & X_2(f) & \cdots & X_N(f) \end{bmatrix}^T, \\
&= S(f) \begin{bmatrix} G_1(f) & G_2(f) & \cdots & G_N(f) \end{bmatrix}^T \\
&= S(f)\mathbf{g}(f), \\
\mathbf{v}(f) &= \begin{bmatrix} V_1(f) & V_2(f) & \cdots & V_N(f) \end{bmatrix}^T, \\
\mathbf{d}(f) &= \begin{bmatrix} 1 & \frac{G_2(f)}{G_1(f)} & \cdots & \frac{G_N(f)}{G_1(f)} \end{bmatrix}^T \\
&= \frac{\mathbf{g}(f)}{G_1(f)},
\end{aligned}$$

and superscript ^{T} denotes transpose of a vector or a matrix. The vector $\mathbf{d}(f)$ is termed the steering vector or direction vector since it determines the direction of the desired signal $X_1(f)$. This definition is a generalization of the classical steering vector [12], [13] to a reverberant (multipath) environment. Indeed, the acoustic impulse response ratios from a broadband source to the aperture convey information about the position of the source.

From (4), we easily deduce the covariance matrix of $\mathbf{y}(f)$, which is

$$\begin{aligned}
\mathbf{\Phi}_\mathbf{y}(f) &= E\left[\mathbf{y}(f)\mathbf{y}^H(f)\right] \\
&= \phi_{X_1}(f)\mathbf{d}(f)\mathbf{d}^H(f) + \mathbf{\Phi}_\mathbf{v}(f), \qquad (5)
\end{aligned}$$

where $\phi_{X_1}(f) = E\left[|X_1(f)|^2\right]$ is the variance of $X_1(f)$ and $\mathbf{\Phi}_\mathbf{v}(f)$ is the covariance matrix of $\mathbf{v}(f)$. The $N \times N$ matrix $\mathbf{\Phi}_\mathbf{y}(f)$ is the sum of two matrices: one is of rank equal to 1 and the other one (covariance matrix of the noise) is assumed to be full rank.

3. WIDELY LINEAR ARRAY MODEL

The classical linear frequency-domain beamforming is performed by applying a complex weight to the output of each sensor and summing across the aperture. Since the frequency-domain components are complex variables and may be noncircular due to the nonstationarity nature of the speech signal, the classical linear estimation theory is not able to exploit all second-order statistics of the complex components. Therefore, we need to resort to the WL estimation theory [8] in order to consider all second-order statistics (including the pseudo-variance) of the CRVs. With this in mind, the output of the WL beamformer is now

$$Z(f) = \mathbf{h}^H(f)\mathbf{y}(f) + \mathbf{h}'^H(f)\mathbf{y}^*(f), \qquad (6)$$

where

$$\begin{aligned}
\mathbf{h}(f) &= \begin{bmatrix} H_1(f) & H_2(f) & \cdots & H_N(f) \end{bmatrix}^T, \\
\mathbf{h}'(f) &= \begin{bmatrix} H'_1(f) & H'_2(f) & \cdots & H'_N(f) \end{bmatrix}^T,
\end{aligned}$$

are the two WL filters of length N which are suitable for performing spatial filtering at frequency f and superscript ^{H} denotes transpose conjugation of a vector or a matrix. The particular case $\mathbf{h}'(f) = \mathbf{0}_{N\times 1}$ leads us obviously to the classical linear beamforming.

Let us define the extended WL filter of length $2N$:

$$\tilde{\mathbf{h}}(f) = \begin{bmatrix} \mathbf{h}(f) \\ \mathbf{h}'(f) \end{bmatrix}, \qquad (7)$$

and the extended observation vector of length $2N$:

$$\begin{aligned}
\tilde{\mathbf{y}}(f) &= \begin{bmatrix} \mathbf{y}(f) \\ \mathbf{y}^*(f) \end{bmatrix} \qquad (8) \\
&= X_1(f)\tilde{\mathbf{d}}_1(f) + X_1^*(f)\tilde{\mathbf{d}}_2(f) + \tilde{\mathbf{v}}(f) \\
&= \tilde{\mathbf{x}}(f) + \tilde{\mathbf{v}}(f),
\end{aligned}$$

where

$$\begin{aligned}
\tilde{\mathbf{d}}_1(f) &= \begin{bmatrix} \mathbf{d}(f) \\ \mathbf{0}_{N\times 1} \end{bmatrix}, \\
\tilde{\mathbf{d}}_2(f) &= \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{d}^*(f) \end{bmatrix},
\end{aligned}$$

and $\tilde{\mathbf{x}}(f)$ and $\tilde{\mathbf{v}}(f)$ are defined in a similar way to $\tilde{\mathbf{y}}(f)$. The output signal of the WL beamformer can be rewritten as

$$\begin{aligned}
Z(f) &= \tilde{\mathbf{h}}^H(f)\tilde{\mathbf{y}}(f) \\
&= X_1(f)\tilde{\mathbf{h}}^H(f)\tilde{\mathbf{d}}_1(f) + X_1^*(f)\tilde{\mathbf{h}}^H(f)\tilde{\mathbf{d}}_2(f) \\
&\quad + \tilde{\mathbf{h}}^H(f)\tilde{\mathbf{v}}(f). \qquad (9)
\end{aligned}$$

Unlike the classical linear beamforming theory where the beamformer output signal consists of only the filtered desired signal and residual noise, the signal estimate in (9) consists of an additional term called interference [9]. Indeed if $\gamma_{X_1}(f) \neq 0$, the signal $X_1^*(f)$ is correlated with $X_1(f)$ and contains both the desired signal and an interference component. Therefore, we need to decompose $X_1^*(f)$ into two orthogonal components as suggested in [9]:

$$X_1^*(f) = \gamma_{X_1}^*(f)X_1(f) + X'_1(f), \qquad (10)$$

where

$$X'_1(f) = X_1^*(f) - \gamma_{X_1}^*(f)X_1(f), \qquad (11)$$

$$E\left[X_1(f)X_1'^*(f)\right] = 0. \qquad (12)$$

The variance of $X'_1(f)$ is

$$\begin{aligned}
\phi_{X'_1}(f) &= E\left[\left|X'_1(f)\right|^2\right] \\
&= \phi_{X_1}(f)\left[1 - |\gamma_{X_1}(f)|^2\right]. \qquad (13)
\end{aligned}$$

Substituting (10) into (9), we find that

$$\begin{aligned}
Z(f) &= X_1(f)\tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma_{X_1}^*(f)\tilde{\mathbf{d}}_2(f)\right] \\
&\quad + X'_1(f)\tilde{\mathbf{h}}^H(f)\tilde{\mathbf{d}}_2(f) + \tilde{\mathbf{h}}^H(f)\tilde{\mathbf{v}}(f) \qquad (14) \\
&= X_{1,\mathrm{fd}}(f) + X'_{1,\mathrm{ri}}(f) + V_{\mathrm{rn}}(f).
\end{aligned}$$

Now, we see that the right-hand side of (14) is the sum of three mutually uncorrelated terms: the overall filtered desired signal, the residual interference, and the residual additive noise. We then deduce the variance of $Z(f)$:

$$\begin{aligned}
\phi_Z(f) &= \tilde{\mathbf{h}}^H(f)\mathbf{\Phi}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{h}}(f) \\
&= \phi_{X_{1,\mathrm{fd}}}(f) + \phi_{X'_{1,\mathrm{ri}}}(f) + \phi_{V_{\mathrm{rn}}}(f), \qquad (15)
\end{aligned}$$

where

$$\begin{aligned}
\phi_{X_{1,\mathrm{fd}}}(f) &= E\left[|X_{1,\mathrm{fd}}(f)|^2\right] \qquad (16) \\
&= \phi_{X_1}(f)\left|\tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma_{X_1}^*(f)\tilde{\mathbf{d}}_2(f)\right]\right|^2, \\
\phi_{X'_{1,\mathrm{ri}}}(f) &= E\left[\left|X'_{1,\mathrm{ri}}(f)\right|^2\right] \qquad (17) \\
&= \phi_{X_1}(f)\left[1 - |\gamma_{X_1}(f)|^2\right]\left|\tilde{\mathbf{h}}^H(f)\tilde{\mathbf{d}}_2(f)\right|^2, \\
\phi_{V_{\mathrm{rn}}}(f) &= \tilde{\mathbf{h}}^H(f)\mathbf{\Phi}_{\tilde{\mathbf{v}}}(f)\tilde{\mathbf{h}}(f). \qquad (18)
\end{aligned}$$

It is of interest to develop the covariance matrix of $\tilde{\mathbf{y}}(f)$. We have

$$\begin{aligned}
\mathbf{\Phi}_{\tilde{\mathbf{y}}}(f) &= E\left[\tilde{\mathbf{y}}(f)\tilde{\mathbf{y}}^H(f)\right] \qquad (19) \\
&= \phi_{X_1}(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right] \times \\
&\quad \left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]^H + \\
&\quad \phi_{X_1}(f)\left[1 - |\gamma_{X_1}(f)|^2\right]\tilde{\mathbf{d}}_2(f)\tilde{\mathbf{d}}_2^H(f) + \mathbf{\Phi}_{\tilde{\mathbf{v}}}(f) \\
&= \mathbf{\Phi}_{\tilde{\mathbf{x}}}(f) + \mathbf{\Phi}_{\tilde{\mathbf{v}}}(f) \\
&= \left[\begin{array}{cc} \mathbf{\Phi}_{\mathbf{y}}(f) & \mathbf{\Psi}_{\mathbf{y}}(f) \\ \mathbf{\Psi}^*_{\mathbf{y}}(f) & \mathbf{\Phi}^*_{\mathbf{y}}(f) \end{array}\right],
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{\Psi}_{\mathbf{y}}(f) &= E\left[\mathbf{y}(f)\mathbf{y}^T(f)\right] \\
&= \psi_{X_1}(f)\mathbf{d}(f)\mathbf{d}^T(f) + \mathbf{\Psi}_{\mathbf{v}}(f) \qquad (20)
\end{aligned}$$

is the pseudo-covariance matrix of $\mathbf{y}(f)$, $\psi_{X_1}(f) = E\left[X_1^2(f)\right]$ is the pseudo-variance of $X_1(f)$, and $\mathbf{\Psi}_{\mathbf{v}}(f)$ is the pseudo-covariance matrix of $\mathbf{v}(f)$.

4. PERFORMANCE MEASURES

In this section, we are going to define some narrowband performance measures useful to study WL beamformers. Since the signal we want to recover is the clean (but convolved) signal received at microphone 1, i.e., $x_1(k)$, microphone 1 is serving as the reference sensor.

We define the narrowband input SNR as

$$\mathrm{iSNR}(f) = \frac{\phi_{X_1}(f)}{\phi_{V_1}(f)}, \qquad (21)$$

where $\phi_{V_1}(f) = E\left[|V_1(f)|^2\right]$ is the variance of the additive noise at microphone 1.

To quantify the level of noise remaining in the output signal of the WL beamformer, $Z(f)$, we define the narrowband output SNR as the ratio of the power of the filtered desired signal over the power of the residual interference and noise[2], i.e.,

$$\mathrm{oSNR}\left[\tilde{\mathbf{h}}(f)\right] = \frac{\phi_{X_{1,\mathrm{fd}}}(f)}{\phi_{X'_{1,\mathrm{ri}}}(f) + \phi_{V_{\mathrm{rn}}}(f)}. \qquad (22)$$

The role of the WL beamformer is to produce a signal whose SNR is higher than that which was received. To that end, the array gain is defined as the ratio of the output SNR (after beamforming) over the input SNR (at the reference microphone) [14]. This leads to the following definition:

$$\mathcal{A}\left[\tilde{\mathbf{h}}(f)\right] = \frac{\mathrm{oSNR}\left[\tilde{\mathbf{h}}(f)\right]}{\mathrm{iSNR}(f)}. \qquad (23)$$

The narrowband noise-reduction factor [15], [16] or narrowband noise-rejection factor [17] quantifies the amount of noise being rejected by the beamformer. This quantity is defined as the ratio of the power of the noise at the reference microphone over the power of the interference and noise remaining at the beamformer output, i.e.,

$$\xi_{\mathrm{nr}}\left[\tilde{\mathbf{h}}(f)\right] = \frac{\phi_{V_1}(f)}{\phi_{X'_{1,\mathrm{ri}}}(f) + \phi_{V_{\mathrm{rn}}}(f)}. \qquad (24)$$

[2] In this paper, we consider the interference as part of the noise in the definitions of the performance measures.

The noise-rejection factor is expected to be lower bounded by 1; otherwise, the beamformer amplifies the noise received at the microphones. The higher the value of the noise-rejection factor, the more the noise is rejected.

In practice, most beamforming algorithms distort the desired signal. In order to quantify the level of this distortion, we define the narrowband desired-signal-reduction factor [1] or narrowband desired-signal-cancellation factor [17] as the ratio of the variance of the desired signal at the reference microphone over the variance of the filtered desired signal at the beamformer output, i.e.,

$$\begin{aligned}
\xi_{\mathrm{dsc}}\left[\tilde{\mathbf{h}}(f)\right] &= \frac{\phi_{X_1}(f)}{\phi_{X_{1,\mathrm{fd}}}(f)} \\
&= \frac{1}{\left|\tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]\right|^2}. \qquad (25)
\end{aligned}$$

A key observation is that the design of broadband beamformers that do not cancel the broadband desired signal requires the constraint

$$\tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right] = 1, \forall f. \qquad (26)$$

Thus, the desired-signal-cancellation factor is equal to 1 if there is no cancellation and expected to be greater than 1 when cancellation happens.

By making the appropriate substitutions, one can derive the following relationship between the array gain, noise-rejection factor, and desired-signal-cancellation factor:

$$\mathcal{A}\left[\tilde{\mathbf{h}}(f)\right] = \frac{\xi_{\mathrm{nr}}\left[\tilde{\mathbf{h}}(f)\right]}{\xi_{\mathrm{dsc}}\left[\tilde{\mathbf{h}}(f)\right]}. \qquad (27)$$

When no distortion occurs, the array gain coincides with the noise-reduction factor.

The narrowband beampattern is a convenient way to represent the response of the beamformer to the signal $x_1(k)$ as a function of the steering vector $\mathbf{d}(f)$ (or equivalently, the location of the source), assuming the absence of any noise or interference. Therefore, the narrowband beampattern is

$$\mathcal{B}\left[\mathbf{d}(f)\right] = \left|\tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]\right|^2. \qquad (28)$$

It is interesting to observe how this beampattern depends on the circularity quotient of the desired signal. For a second-order CCRV [i.e., $\gamma_{X_1}(f) = 0$], (28) simplifies to the classical narrowband beampattern:

$$\mathcal{B}\left[\mathbf{d}(f)\right] = \left|\mathbf{h}^H(f)\mathbf{d}(f)\right|^2. \qquad (29)$$

5. OPTIMAL WIDELY LINEAR MVDR

The optimal WL MVDR filter is found by minimizing the variance of the output signal of the WL beamformer with the constraint that the desired signal is not distorted [9]. Mathematically this is equivalent to

$$\begin{aligned}
&\min_{\tilde{\mathbf{h}}(f)} \tilde{\mathbf{h}}^H(f)\mathbf{\Phi}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{h}}(f) \\
&\quad \text{subject to } \tilde{\mathbf{h}}^H(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right] = 1. \qquad (30)
\end{aligned}$$

We easily deduce the optimal WL MVDR beamformer:

$$\tilde{\mathbf{h}}_{\mathrm{WL-MVDR}}(f) = \frac{\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]}{C(f)}, \qquad (31)$$

where

$$C(f) = \left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]^H \mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f) \times \left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]. \tag{32}$$

Expression (31) depends on the acoustic impulse response ratios. It is more convenient to write it as a function of the signal statistics only. Let us first define two matrices and two vectors:

$$\mathbf{U}_{10} = \begin{bmatrix} \mathbf{I}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \end{bmatrix}, \tag{33}$$

$$\mathbf{U}_{01} = \begin{bmatrix} \mathbf{0}_{N\times N} & \mathbf{0}_{N\times N} \\ \mathbf{0}_{N\times N} & \mathbf{I}_{N\times N} \end{bmatrix}, \tag{34}$$

$$\mathbf{u}_1 = \begin{bmatrix} \mathbf{i} \\ \mathbf{0}_{N\times 1} \end{bmatrix}, \tag{35}$$

$$\mathbf{u}_2 = \begin{bmatrix} \mathbf{0}_{N\times 1} \\ \mathbf{i} \end{bmatrix}, \tag{36}$$

where $\mathbf{i} = \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix}^T$ is a vector of length N. We start by rewriting the elements that appear in the numerator of (31). The first one is

$$\begin{aligned} \mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{d}}_1(f) &= \mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{d}}_1(f)\tilde{\mathbf{d}}_1^H(f)\mathbf{u}_1 \\ &= \phi^{-1}_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{10}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{U}_{10}\mathbf{u}_1 \\ &= \phi^{-1}_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{10}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{u}_1. \end{aligned} \tag{37}$$

The second element is

$$\begin{aligned} &\gamma^*_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{d}}_2(f) \\ &= \gamma^*_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{d}}_2(f)\tilde{\mathbf{d}}_2^H(f)\mathbf{u}_2 \\ &= \gamma^*_{X_1}(f)\phi^{-1}_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{01}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{U}_{01}\mathbf{u}_2 \\ &= \phi^{-1}_{X_1}(f)\gamma^*_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{01}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{u}_2. \end{aligned} \tag{38}$$

Using (19), the denominator of (31) can be rewritten as

$$\begin{aligned} &\phi^{-1}_{X_1}(f)\mathrm{tr}\left\{\phi_{X_1}(f)\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right] \times \right. \\ &\quad \left. \left[\tilde{\mathbf{d}}_1(f) + \gamma^*_{X_1}(f)\tilde{\mathbf{d}}_2(f)\right]^H \right\} \\ &= \phi^{-1}_{X_1}(f)\mathrm{tr}\left\{\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f) - \phi_{X_1}(f)\left[1 - |\gamma_{X_1}(f)|^2\right] \times \right. \\ &\quad \left. \mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\tilde{\mathbf{d}}_2(f)\tilde{\mathbf{d}}_2^H(f) \right\} \\ &= \phi^{-1}_{X_1}(f)\mathrm{tr}\left\{\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f) - \left[1 - |\gamma_{X_1}(f)|^2\right] \times \right. \\ &\quad \left. \mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{01}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{U}_{01} \right\}, \end{aligned} \tag{39}$$

where $\mathrm{tr}\{\cdot\}$ denotes the trace of a square matrix.

Using (37), (38), and (39), we deduce a new form of the optimal WL MVDR filter:

$$\tilde{\mathbf{h}}_{\mathrm{WL-MVDR}}(f) = \frac{\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\left[\mathbf{U}_{10}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{u}_1 + \gamma^*_{X_1}(f)\mathbf{U}_{01}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{u}_2\right]}{C(f)}, \tag{40}$$

where now

$$C(f) = \mathrm{tr}\left[\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\right] - \left[1 - |\gamma_{X_1}(f)|^2\right] \times \mathrm{tr}\left[\mathbf{\Phi}^{-1}_{\tilde{\mathbf{y}}}(f)\mathbf{U}_{01}\mathbf{\Phi}_{\tilde{\mathbf{x}}}(f)\mathbf{U}_{01}\right]. \tag{41}$$

6. CONCLUSIONS

When we work in the frequency domain, we generally deal with complex random variables even though the original time-domain signals are real in the context of speech applications. A complex random variable can be either (second-order) circular or noncircular depending on whether its pseudo-variance is zero or not. Traditionally, the frequency-domain components of speech are assumed to be circular and most beamforming approaches consider only the variance of these components (or power spectra). Because speech signals are highly nonstationary, they are noncircular in the frequency domain and this noncircularity should be taken into account. In this paper, we have first derived all important beamforming performance measures for a noncircular CRV and then derived the optimal frequency-domain WL MVDR beamformer in the context of room acoustics.

7. REFERENCES

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

[2] E. Ollila, "On the circularity of a complex random variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008.

[3] D. P. Mandic and S. L. Goh, *Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear and Neural Models*. Wiley, 2009.

[4] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals – Part I: variables," *Elsevier Signal Process.*, vol. 53, pp. 1–13, 1996.

[5] P. O. Amblard, M. Gaeta, and J. L. Lacoume, "Statistics for complex variables and signals – Part II: signals," *Elsevier Signal Process.*, vol. 53, pp. 15–25, 1996.

[6] B. Picinbono, "On circularity," *IEEE Trans. Signal Process.*, vol. 42, pp. 3473–3482, Dec. 1994.

[7] 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.

[8] B. Picinbono and P. Chevalier, "Widely linear estimation with complex data," *IEEE Trans. Signal Process.*, vol. 43, pp. 2030–2033, Aug. 1995.

[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] M. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[11] 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, Chapter 47, pp. 945–978.

[12] B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE Acoust., Speech, Signal Process. Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[13] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds., Berlin, Germany: Springer-Verlag, 2003, Chapter 6, pp. 155–194.

[14] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing–Concepts and Techniques*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[15] J. Benesty, J. Chen, Y. Huang, and S. Doclo, "Study of the Wiener filter for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., Berlin, Germany: Springer-Verlag, 2005, Chapter 2, pp. 9–41.

[16] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, pp. 1218–1234, July 2006.

[17] W. Herbordt, *Combination of Robust Adaptive Beamforming with Acoustic Echo Cancellation for Acoustic Human/Machine Interfaces*. PhD Thesis, Erlangen-Nuremberg University, Germany, 2004.