Identification of Speech Source Coupling Between Sensors in Reverberant Noisy Environments

Israel Cohen, *Senior Member, IEEE*

Abstract—An important component of a multichannel hands-free communication system is the identification of the coupling between sensors in response to a desired speech signal. In this letter, a system identification approach adapted to speech signals is proposed. A weighted least-squares optimization criterion is introduced, which incorporates an indicator function for the presence of the desired signal in the observed signals. We show that compared to a competing nonstationarity-based method, a significantly smaller error variance is achievable.

Index Terms—Array signal processing, speech enhancement, system identification.

I. INTRODUCTION

AN important component of a multichannel hands-free communication system is the identification of the coupling between sensors in response to a desired speech signal [1]–[3]. This coupling, often referred to as the acoustical transfer function (ATF) ratio, represents the relation between the impulse responses of the sensors to the desired source. In reverberant and noisy environments, the coupling identification enables to construct an adaptive blocking channel, for an accurate derivation of a reference noise signal, and an adaptive noise canceller, for eliminating directional or coherent noise sources. Furthermore, it facilitates multichannel signal detection and postfiltering techniques, which employ the transient power ratio between the beamformer output and the reference signals [4].

We consider the following model:

$$x(t) = s(t) + u(t); \quad y(t) = a(t) * s(t) + w(t) \quad (1)$$

where $x(t)$ and $y(t)$ are signals measured by a primary and reference sensors, $s(t)$ is the desired signal received by the primary sensor, $u(t)$ and $w(t)$ are additive interfering signals that are uncorrelated with the desired signal, $a(t)$ represents the coupling of the desired signal to the reference sensor, and $*$ denotes convolution[1]. Our objective is to identify $a(t)$ in the general case where $u(t)$ is statistically correlated with $w(t)$. An equiva-

lent problem is to consider a linear time-invariant (LTI) system, whose input $x(t)$ and output $y(t)$ are related by

$$y(t) = a(t) * x(t) + v(t) \quad (2)$$

where $a(t)$ represents the impulse response of the system that we want to identify, $v(t)$ denotes additive noise that is generally correlated with $x(t)$, and $x(t)$ is the sum of a desired signal $s(t)$ and an interfering signal $u(t)$ as in (1). It is assumed that $u(t)$ and $v(t)$ are statistically uncorrelated with $s(t)$. Then, the two above-mentioned problems are equal, with

$$w(t) = v(t) + a(t) * u(t). \quad (3)$$

Shalvi and Weinstein [1] have proposed to use the nonstationarity of the desired signal. They assumed that the interfering signals are stationary, while the desired signal is nonstationary. Then, dividing the observation interval into a sequence of subintervals, and computing for each subinterval the cross power spectral density (PSD) of the sensors, they obtained an overdetermined set of equations for the two unknown quantities: the system's transfer function and the (presumably stationary) cross-PSD of $x(t)$ and $v(t)$. An asymptotically unbiased estimate for the system's transfer function was derived by using a weighted least-squares (WLS) approach for minimizing the error variance under certain assumptions.

A major limitation of the nonstationarity-based system identification is that both the system identification and noise estimation are carried out through the same WLS optimization criterion. The WLS optimization consists of two conflicting requirements: One is minimizing the error variance of the system's transfer function estimate, which pulls the weight up to higher values in higher SNR subintervals. The other requirement is minimizing the error variance of the noise estimate, which rather implies smaller weights in higher SNR subintervals. Furthermore, the observation interval is required to be sufficiently long, so that for all frequency bands it includes quite a few subintervals that contain the desired signal. This, together with the assumption that the interfering signals remain stationary during the entire observation interval, restrict the capability of this technique to track time-varying systems (e.g., tracking moving talkers in hands-free communication scenarios).

In this letter, a system identification approach adapted to speech signals is proposed. An indicator function for the speech presence in the time-frequency domain is incorporated into the optimization criterion, and a minimum variance WLS estimate for the system's transfer function is derived. We show that the error variance obtained by using the proposed method is generally smaller than that obtained by using the nonstationarity

[1]Note that $s(t)$ is generally a reverberated version of the source signal, i.e., $s(t) = a_0(t) * s_0(t)$, where $s_0(t)$ is the source signal and $a_0(t)$ is the impulse response of the primary sensor to the desired source. In that case, $a_1(t) \triangleq a(t) * a_0(t)$ represents the impulse response of the *reference* sensor to the desired source.

method. Experimental results under various noise conditions demonstrate the performance of the proposed method.

II. SYSTEM IDENTIFICATION

Using the short-time Fourier transform (STFT) and assuming the support of the window function is sufficiently large compared with the duration of $a(t)$, (2) can be written in the time-frequency domain as

$$Y(k, \ell) = A(k) X(k, \ell) + V(k, \ell) \tag{4}$$

where $A(k)$ is the transfer function of the system, k represents the frequency bin index ($k = 0, 1, \ldots, K - 1$), and ℓ is the frame index ($\ell = 0, 1, \ldots, L - 1$). The cross-PSD between y and x is therefore given by

$$\phi_{yx}(k, \ell) = A(k) \phi_{xx}(k, \ell) + \phi_{vx}(k, \ell). \tag{5}$$

Since the desired signal $s(t)$ is uncorrelated with the interfering signals $u(t)$ and $w(t)$, (1) and (3) imply

$$\phi_{yx}(k, \ell) = A(k) \phi_{ss}(k, \ell) + \phi_{wu}(k, \ell). \tag{6}$$

Writing this equation in terms of the PSD estimates, we have

$$\hat{\phi}_{yx}(k, \ell) - \hat{\phi}_{wu}(k, \ell) = A(k) \hat{\phi}_{ss}(k, \ell) + \varepsilon(k, \ell) \tag{7}$$

where $\varepsilon(k, \ell)$ denotes an estimation error. This gives us L equations, which may be written in a matrix form as

$$\hat{\psi} = \hat{\phi}_{ss} A + \varepsilon \tag{8}$$

where $\hat{\phi}_{ss} \triangleq [\hat{\phi}_{ss}(k, 1), \ldots, \hat{\phi}_{ss}(k, L)]^T, \hat{\psi} \triangleq \hat{\phi}_{yx} - \hat{\phi}_{wu}, \varepsilon \triangleq [\varepsilon(k, 1), \ldots, \varepsilon(k, L)]^T$, and the argument k has been omitted for notational simplicity. Since A represents the coupling associated with the desired source signal, the optimization criterion for its identification has to take into account only short-time frames which contain desired signal components. Specifically, let $I(k, \ell)$ denote an indicator function for the signal presence (i.e., $I(k, \ell) = 1$ if $\phi_{ss}(k, \ell) \neq 0$, and $I(k, \ell) = 0$ otherwise), and let $\mathbf{I}$ represent a diagonal matrix with the elements $[I(k, 1), I(k, 2), \ldots, I(k, L)]$ on its diagonal. Then the WLS estimate of A is obtained by

$$
\begin{aligned}
\hat{A} &= \arg \min_A \{ [\mathbf{I} \varepsilon]^H \mathbf{W} [\mathbf{I} \varepsilon] \} \\
&= \arg \min_A \{ [\hat{\psi} - \hat{\phi}_{ss} A]^H \mathbf{I} \mathbf{W} \mathbf{I} [\hat{\psi} - \hat{\phi}_{ss} A] \} \\
&= \left[\hat{\phi}_{ss}^T \mathbf{I} \mathbf{W} \mathbf{I} \hat{\phi}_{ss} \right]^{-1} \hat{\phi}_{ss}^T \mathbf{I} \mathbf{W} \mathbf{I} \hat{\psi}.
\end{aligned} \tag{9}
$$

Recognizing the product $\mathbf{I} \mathbf{W} \mathbf{I}$ as the equivalent weight matrix, the variance of $\hat{A}$ is given by [5, p. 405]

$$
\begin{aligned}
\text{var}(\hat{A}) = \left(\phi_{ss}^T \mathbf{I} \mathbf{W} \mathbf{I} \phi_{ss} \right)^{-1} \phi_{ss}^T \mathbf{I} \mathbf{W} \mathbf{I} \text{cov}(\varepsilon) \\
\times \mathbf{I} \mathbf{W} \mathbf{I} \phi_{ss} \left(\phi_{ss}^T \mathbf{I} \mathbf{W} \mathbf{I} \phi_{ss} \right)^{-1} \tag{10}
\end{aligned}
$$

where $\text{cov}(\varepsilon)$ is the covariance matrix of ε. The matrix $\mathbf{W}$ that minimizes the variance of $\hat{A}$ therefore satisfies [5, prop. 8.2.4]

$$\mathbf{I} \mathbf{W} \mathbf{I} = \mathbf{I} [\text{cov}(\varepsilon)]^{-1} \mathbf{I}. \tag{11}$$

The resulting estimator

$$\hat{A} = \left(\hat{\phi}_{ss}^T \mathbf{I} [\text{cov}(\varepsilon)]^{-1} \mathbf{I} \hat{\phi}_{ss} \right)^{-1} \hat{\phi}_{ss}^T \mathbf{I} [\text{cov}(\varepsilon)]^{-1} \mathbf{I} \hat{\psi} \tag{12}$$

is known as the *minimum variance* or *Markov estimator*, and its variance is given by

$$\text{var}(\hat{A}) = \left(\phi_{ss}^T \mathbf{I} [\text{cov}(\varepsilon)]^{-1} \mathbf{I} \phi_{ss} \right)^{-1}. \tag{13}$$

The elements of $\text{cov}(\varepsilon)$ are asymptotically given by (Appendix I)

$$\text{cov}(\varepsilon(k, \ell), \varepsilon(k, \ell')) = \begin{cases} \phi_{ss}(k, \ell) \phi_{vv}(k, \ell), & \text{if } \ell = \ell', \\ 0, & \text{otherwise.} \end{cases} \tag{14}$$

Assuming that the interfering signals $w(t)$ and $u(t)$ are stationary, (3) implies that $\phi_{vv}(k, \ell)$ is independent of the frame index ℓ (in practice, as demonstrated in the next section, it suffices that the statistics of the interfering signals is slowly changing compared with the statistics of the desired signal). Denoting by $\langle \cdot \rangle$ an average operation over the frame index ℓ

$$\langle \varphi(k, \ell) \rangle \triangleq \frac{1}{L} \sum_{\ell=1}^{L} \varphi(k, \ell) \tag{15}$$

and substituting (14) into (12) and (13) we obtain

$$\hat{A} = \frac{\langle I(k, \ell) [\hat{\phi}_{yx}(k, \ell) - \hat{\phi}_{wu}(k, \ell)] \rangle}{\langle I(k, \ell) \hat{\phi}_{ss}(k, \ell) \rangle} \tag{16}$$

$$\text{var}(\hat{A}) = \frac{\phi_{vv}(k)}{L \langle I(k, \ell) \phi_{ss}(k, \ell) \rangle}. \tag{17}$$

The proposed identification approach requires estimates for $\phi_{yx}(k, \ell), \phi_{ss}(k, \ell), \phi_{wu}(k, \ell)$ and $I(k, \ell)$. An estimate for the cross-PSD of the observed signals is obtained by $\hat{\phi}_{yx}(k, \ell) = Y(k, \ell) X^*(k, \ell)$. To obtain an estimate for the PSD of the desired signal, we apply the *Optimally Modified Log-Spectral Amplitude* (OM-LSA) estimator [6] to $X(k, \ell)$, and compute the resulting periodogram

$$\hat{\phi}_{ss}(k, \ell) = \left[G^{p(k, \ell)}(k, \ell) G_{\min}^{1-p(k, \ell)} |X(k, \ell)| \right]^2 \tag{18}$$

where $G(k, \ell)$ denotes the log-spectral amplitude gain function [7], $p(k, \ell)$ is the speech presence probability [6], and $G_{\min}$ is the minimal spectral gain. The cross-PSD of the interfering signals, $w(t)$ and $u(t)$, is estimated by using the *Minima Controlled Recursive Averaging* (MCRA) approach [8]. Specifically, past spectral cross-power values of the noisy observed signals are recursively averaged using a time-varying frequency-dependent smoothing parameter. An estimator for the indicator function is obtained by

$$\hat{I}(k, \ell) = \begin{cases} 1, & \text{if } p(k, \ell) \geq p_0, \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

where p_0 ($0 \leq p_0 < 1$) is a threshold for the speech presence probability. The parameter p_0 controls the trade-off between the detection and false alarm probabilities, which are defined by $P_D \triangleq \mathcal{P}\{p(k, \ell) \geq p_0 \mid I(k, \ell) = 1\}$ and $P_{FA} \triangleq \mathcal{P}\{p(k, \ell) \geq p_0 \mid I(k, \ell) = 0\}$. A smaller value of p_0 increases the detection probability and allows for more short-time frames to be involved in the estimation of A. However, a smaller value of p_0 also increases the false alarm probability, which may cause a mis-modification of $\hat{A}$ due to frames that do not contain desired speech components.

For the comparison with the nonstationarity method, we replace the subinterval index m in [1, Eq. 28] with the frame index

ℓ, and normalize the window function so that $\sum_n w^2(n) = N$, where N is the frame's length. Accordingly, the variance of $\hat{A}$ obtained using the nonstationarity method is

$$\text{var}(\hat{A})\,|_{\text{NS}} = \frac{1}{L} \cdot \frac{\phi_{vv}(k)\,\langle\phi_{xx}^{-1}(k,\ell)\rangle}{\langle\phi_{xx}(k,\ell)\rangle\,\langle\phi_{xx}^{-1}(k,\ell)\rangle - 1}. \quad (20)$$

The ratio between the variances obtained by the proposed and the nonstationarity methods is given by

$$\rho \triangleq \frac{\text{var}(\hat{A})}{\text{var}(\hat{A})\,|_{\text{NS}}} = \frac{\langle\phi_{xx}(k,\ell)\rangle\,\langle\phi_{xx}^{-1}(k,\ell)\rangle - 1}{\langle\phi_{xx}^{-1}(k,\ell)\rangle\,\langle\hat{I}(k,\ell)\,\phi_{ss}(k,\ell)\rangle}. \quad (21)$$

Let $\xi \triangleq \phi_{ss}(k,\ell)/\phi_{uu}(k)$ denote the *a priori* SNR at the primary sensor. Then approximating

$$\langle\hat{I}(k,\ell)\,\phi_{ss}(k,\ell)\rangle \approx \langle I(k,\ell)\,\phi_{ss}(k,\ell)\rangle = \langle\phi_{ss}(k,\ell)\rangle \quad (22)$$

and substituting $\phi_{xx}(k,\ell) = \phi_{ss}(k,\ell) + \phi_{uu}(k)$ into (21), we obtain (Appendix II)

$$\rho \approx \frac{\langle\xi+1\rangle\langle(\xi+1)^{-1}\rangle - 1}{\langle(\xi+1)^{-1}\rangle\langle\xi\rangle_\ell} < 1. \quad (23)$$

Thus, the variance of $\hat{A}$ obtained by using the proposed method is generally smaller than that obtained by using the nonstationarity method, as long as the approximation (22) is valid. Additionally, the nonstationarity method requires a sufficiently long observation interval, to account for the nonstationarity of $\phi_{xx}(k,\ell)$, and the interfering signals must remain stationary during the entire observation interval. On the other hand in the proposed method, not only a shorter observation interval suffices, but also the interfering signals are not required to be stationary during time-frequency windows that do not contain desired signal components. Accordingly, in case of a time-varying system, a faster convergence and higher reliability of the system identification is achieved by using the proposed method.

III. EXPERIMENTAL RESULTS

A quantitative comparison between the proposed and the nonstationarity system identification methods is obtained by evaluating the signal blocking factor (SBF), defined by

$$\text{SBF} = 10\log_{10}\frac{E\{s^2(t)\}}{E\{r^2(t)\}} \quad [\text{dB}] \quad (24)$$

where $E\{s^2(t)\}$ is the energy contained in the clean speech signal, and $E\{r^2(t)\}$ is the energy contained in the leakage signal

$$r(t) = a(t) * s(t) - \hat{a}(t) * s(t). \quad (25)$$

The leakage signal represents the difference between the reverberated clean signal at the reference sensor and its estimate $\hat{a}(t) * s(t)$ given the desired signal at the primary sensor. It has a major affect on the amount of distortion introduced by the *Transfer Function GSC* [2]. The SBF measure is associated with the capability to block the desired signal and produce a noise-only signal by computing $\hat{v}(t) = y(t) - \hat{a}(t) * x(t)$.

The first experiment was performed on a speech signal (female speaker) sampled at 8 kHz. Similar to the experiment in [1], the noise $u(t)$ is a stationary zero-mean Gaussian process
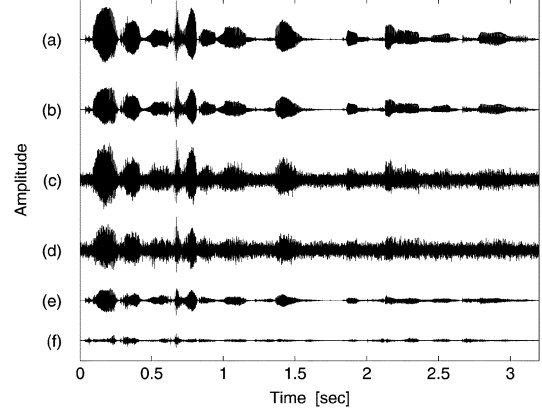

Fig. 1. Speech waveforms. (a) $s(t)$. (b) $a(t) * s(t)$. (c) Noisy signal at the primary sensor ($\text{SNR} = 3.0$ dB). (d) Noisy signal at the reference sensor ($\text{SNR} = -4.3$ dB). (e) Signal leakage using the nonstationarity method ($\text{SBF} = 8.5$ dB). (f) Signal leakage using the proposed method ($\text{SBF} = 18.3$ dB).

whose average power is half the average power of the speech ($\text{SNR} = 3$ dB). The impulse response of the reference sensor to the desired signal is $a(t) = \delta(t - 6T) - 0.5\,\delta(t - 7T) + 0.25\,\delta(t - 8T)$, where $T = 12.5$ ms is the sampling period. In addition, the reference sensor noise $w(t)$ is generated by $w(t) = g(t) * u(t)$, where $g(t) = -\delta(t) - 0.5\,\delta(t-T) + 0.1\,\delta(t-2T)$. Fig. 1(a)–(d) show the clean speech signals at the primary and reference sensors, and the observed noisy signals. We have applied the nonstationarity-based system identification algorithm [1] to a 4-s observation interval (32 000 samples) that was arbitrarily divided into disjoint subintervals of 128 samples length. As is suggested in [2], only subintervals in which speech is active (SNR in the subinterval is greater than 0 dB) were taken into account. The leakage signal $r(t)$ is plotted in Fig. 1(e). The resultant SBF is 8.5 dB.

In the proposed method, the STFT is implemented with Hamming windows of 256 samples length (32 ms) and 128 framing step (50% overlap between frames). We restricted the spectral gain in (18) to a minimum $G_{\min} = -20$ dB, and used a speech presence probability threshold $p_0 = 0.7$ in (19). The leakage signal $r(t)$ is plotted in Fig. 1(f). The resultant SBF is 18.3 dB, which is significantly higher than that obtained by using the nonstationarity method.

In the second experiment, two microphones with 10 cm spacing are mounted in a car on the visor. Clean (reverberated) speech signals are recorded at a sampling rate of 8 kHz in the absence of background noise (standing car, silent environment). Car noise signals are recorded while the car speed is about 60 km/h, and the window next to the driver is either closed or slightly open (about 5 cm; the other windows remain closed). The noise PSD is pseudo-stationary in the former case, while varies substantially in the latter case due to wind blows and passing cars. The input microphone signals are generated by mixing the speech and noise signals at various SNR levels in the range $[-10, 10]$ dB. Fig. 2 shows experimental results of the average SBF obtained under various car noise conditions using the competing system identification algorithms. Clearly, the proposed system identification method is considerably more efficient than the nonstationarity-based method. The
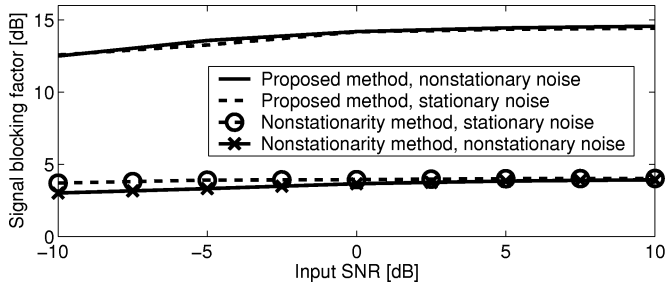
Fig. 2.   Average signal blocking factor under various car noise conditions.

rationale is that subintervals with low SNR are more useful for noise estimation, whereas subintervals with high SNR are more useful for system identification. Therefore, by weighting the subintervals for noise estimation differently than the weighting for system identification, improved performance is achieved. Moreover, the proposed algorithm is less sensitive to variations in the noise statistics in case the noise is nonstationary. For a given input SNR, the performance of the proposed algorithm in a *nonstationary* noise environment might be even slightly better than that obtained in a stationary noise environment. This is related to the fact that for a given input SNR and nonstationary noise, there are necessarily subintervals where the instantaneous noise power is lower than its average, and these subintervals are given higher weights in the system identification process. On the contrary, the performance of the nonstationarity-based algorithm, which is based on the nonstationarity of the desired signal alone, is essentially impaired in nonstationary noise environments.

IV. CONCLUSION

We have proposed a system identification approach for the coupling between sensors in response to speech signals. The optimization criterion takes into account only short-time frames which contain desired speech components. The proposed estimator incorporates estimators for the auto-PSD of the desired signal, cross-PSD of the interfering signals, and indicator function of speech presence. These estimates are derived by using the OM-LSA [6] and MCRA [8] estimation techniques, which also provide the speech presence probability and are particularly useful for enhancing speech signals acquired in nonstationary noisy environments. We showed that the proposed minimum variance WLS estimate for the system's transfer function generally yields a smaller error variance than that obtained by the nonstationarity method. Using the proposed method for the identification of the acoustical transfer function ratios, as part of the *Transfer-Function GSC* [2], may improve the adaptation of the blocking matrix and noise canceller, and as a result may improve the performance of the beamformer.

APPENDIX I
ASYMPTOTIC COVARIANCE OF ε

From (6) and (7), we have

$$\varepsilon = [\hat{\phi}_{yx} - \phi_{yx}] - [\hat{\phi}_{wu} - \phi_{wu}] - A[\hat{\phi}_{ss} - \phi_{ss}]. \qquad (26)$$

Using (2), (3) and the relation $\phi_{xx} = \phi_{ss} + \phi_{uu}$ we obtain

$$\varepsilon = [\hat{\phi}_{vx} - \phi_{vx}] - [\hat{\phi}_{vu} - \phi_{vu}] = \hat{\phi}_{vs} - \phi_{vs}. \qquad (27)$$

Since we use the cross-periodogram as an estimate for the cross-PDS, under the assumption that observations in the time-frequency domain associated with different frames are statistically independent, we have (e.g., [5, ch. 5])

$$\text{cov}(\varepsilon(k,\ell), \varepsilon(k,\ell')) = \begin{cases} \phi_{ss}(k,\ell)\,\phi_{vv}(k,\ell), & \text{if } \ell = \ell', \\ 0, & \text{otherwise.} \end{cases} \qquad (28)$$

APPENDIX II
DERIVATION OF (23)

By (21)

$$\rho = \frac{\langle \phi_{xx} \rangle \langle \phi_{xx}^{-1} \rangle - 1}{\langle \phi_{xx}^{-1} \rangle \langle \hat{I}\,\phi_{ss} \rangle} \qquad (29)$$

where, for notational simplicity, the arguments k and ℓ are omitted. Denoting by $\xi = \phi_{ss}/\phi_{uu}$ the *a priori* SNR at the primary sensor, and using $\phi_{xx} = \phi_{ss} + \phi_{uu}$ and $\langle \hat{I}\,\phi_{ss} \rangle \approx \langle I\,\phi_{ss} \rangle = \langle \phi_{ss} \rangle$, together with the assumption that $u(t)$ is stationary (ϕ_{uu} is independent of the frame index ℓ), we have

$$\rho \approx \frac{\langle \xi+1 \rangle \langle (\xi+1)^{-1} \rangle - 1}{\langle (\xi+1)^{-1} \rangle \langle \xi \rangle} = 1 - \frac{\langle \xi(\xi+1)^{-1} \rangle}{\langle (\xi+1)^{-1} \rangle \langle \xi \rangle} < 1. \qquad (30)$$

ACKNOWLEDGMENT

REFERENCES

[1] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Trans. Signal Processing*, vol. 44, pp. 2055–2063, 1996.
[2] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, pp. 1614–1626, Aug. 2001.
[3] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Processing*, vol. 47, pp. 2677–2684, Oct. 1999.
[4] I. Cohen and B. Berdugo, "Two-channel signal detection and speech enhancement based on the transient beam-to-reference ratio," in *Proc. ICASSP 2003*, Hong Kong, Apr. 6–10, 2003, pp. V 233–V 236.
[5] D. G. Manolakis, V. K. Ingle, and S. M. Kogan, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering and Array Processing*.   New York: McGraw-Hill, 2000.
[6] I. Cohen and B. Berdugo, "Speech enhancement for nonstationary noise environments," *Signal Process.*, vol. 81, no. 11, pp. 2403–2418, Oct. 2001.
[7] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443–445, Apr. 1985.
[8] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 466–475, Sept. 2003.