Multichannel Signal Detection Based on the Transient Beam-to-Reference Ratio

Israel Cohen and Baruch Berdugo

***Abstract*—In this letter, we present a multichannel signal detection approach that is particularly advantageous in nonstationary noise environments. A beamformer is realistically assumed to have a steering error, a blocking matrix that is unable to block all of the desired signal components, and a noise canceler that is adapted to the pseudostationary noise, but not modified during transient interferences. Signal components are detected at the beamformer output based on a measure of their local nonstationarity, and discriminated from transient noise components based on the transient beam-to-reference ratio.**

***Index Terms*—Adaptive signal processing, array signal processing, interference suppression, signal detection, spectral analysis.**

I. Introduction

MULTICHANNEL systems advance high-quality hands-free communication in reverberant and noisy environments [1]. Compared to single-channel systems, a substantial gain in performance is obtainable due to the spatial filtering capability to suppress interfering signals coming from undesired directions. However, in case of incoherent or diffuse noise fields, beamforming alone does not provide sufficient noise reduction, and postfiltering is normally required [2]–[4]. Postfiltering includes signal detection, noise estimation, and spectral enhancement. A major drawback of existing multichannel postfiltering techniques is that highly nonstationary noise components are not dealt with. The time variation of the interfering signals is assumed to be sufficiently slow, such that the postfilter can track and adapt to the changes in the noise statistics. Transient interferences, on the other hand, are not differentiated from the desired signal components.

In this letter, we present a multichannel signal detection approach based on the transient beam-to-reference ratio. A desired signal component is presumably stronger at the beamformer output than at any reference noise signal, and a noise component is strongest at one of the reference signals. Hence, the ratio between the transient power at beamformer output and the transient power at the reference signals indicates whether such a transient is desired or interfering. In Section II, we review the multichannel generalized sidelobe canceler, and derive linear relations in the power-spectral domain between the beamformer output, the reference noise signals, the desired source signal, and the input transient interferences. Then, in Section III, signal components are detected based on a measure of their local nonstationarity, and discriminated from transient noise components based on the transient beam-to-reference ratio.

Manuscript received August 3, 2002; revised October 23, 2002. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Raffaele Parisi.

I. Cohen is with the Department of Electrical Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 32000, Israel (e-mail: icohen@ee.technion.ac.il).

B. Berdugo is with Lamar Signal Processing Ltd., Andrea Electronics Corp.—Israel, Yokneam Ilit 20692, Israel (e-mail: bberdugo@lamar.co.il).

Digital Object Identifier 10.1109/LSP.2003.815612

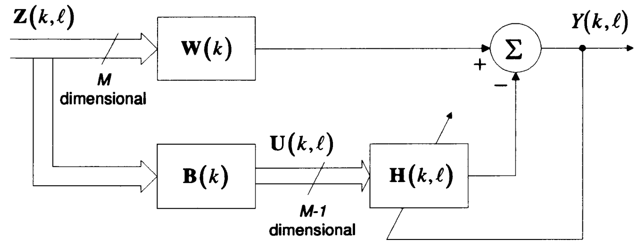

Fig. 1. Multichannel generalized sidelobe canceler.

II. Multichannel Generalized Sidelobe Canceling

Let $x(t)$ denote a desired source signal, and let signal vectors $\mathbf{d}_s(t)$ and $\mathbf{d}_t(t)$ denote multichannel uncorrelated interfering signals at the output of M sensors. The vector $\mathbf{d}_s(t)$ represents pseudostationary interferences, and $\mathbf{d}_t(t)$ represents undesired transient components. The observed signal at the ith sensor is given by

$$z_i(t) = a_i(t) * x(t) + d_{is}(t) + d_{it}(t), \qquad i = 1, \ldots, M \tag{1}$$

where $a_i(t)$ is the impulse response of the ith sensor to the desired source, $*$ denotes convolution, and d_{is} and d_{it} are the interference signals corresponding to the ith sensor. Using the short-time Fourier transform (STFT), and assuming time-invariant impulse responses [5], we have in the time-frequency domain

$$\mathbf{Z}(k,\ell) = \mathbf{A}(k)X(k,\ell) + \mathbf{D}_s(k,\ell) + \mathbf{D}_t(k,\ell) \tag{2}$$

where k represents the frequency bin index, ℓ the frame index, and

$$\begin{aligned}
\mathbf{Z}(k,\ell) &\triangleq [\,Z_1(k,\ell) \quad Z_2(k,\ell) \quad \cdots \quad Z_M(k,\ell)\,]^T \\
\mathbf{A}(k) &\triangleq [\,A_1(k) \quad A_2(k) \quad \cdots \quad A_M(k)\,]^T \\
\mathbf{D}_s(k,\ell) &\triangleq [\,D_{1s}(k,\ell) \quad D_{2s}(k,\ell) \quad \cdots \quad D_{Ms}(k,\ell)\,]^T \\
\mathbf{D}_t(k,\ell) &\triangleq [\,D_{1t}(k,\ell) \quad D_{2t}(k,\ell) \quad \cdots \quad D_{Mt}(k,\ell)\,]^T.
\end{aligned}$$

Fig. 1 shows a generalized sidelobe canceler structure for a linearly constrained adaptive beamformer [6], [7]. This structure is also utilizable in case $\mathbf{A}$, the transfer function from the desired source to the sensor array, is arbitrary [5]. The beamformer comprises a fixed beamformer $\mathbf{W}(k)$, a blocking matrix $\mathbf{B}(k)$ which yields the reference noise signals $\mathbf{U}(k,\ell)$, and a multichannel adaptive noise canceler $\mathbf{H}(k,\ell)$ which eliminates the stationary noise that pass through the fixed beamformer. We assume that the noise canceler is adapted only to the stationary noise, and not modified during transient interferences. Furthermore, we expect that some desired signal components may leak through the blocking matrix due to steering error.

Assuming homogeneous noise fields, the power-spectral density (PSD) matrices of the input noise signals are related to the corresponding spatial coherence matrices, $\mathbf{\Gamma}_s(k,\ell)$ and $\mathbf{\Gamma}_t(k,\ell)$, by

$$\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell) = \lambda_s(k,\ell)\mathbf{\Gamma}_s(k,\ell)$$
$$\mathbf{\Phi}_{\mathbf{D}_t\mathbf{D}_t}(k,\ell) = \lambda_t(k,\ell)\mathbf{\Gamma}_t(k,\ell)$$

where $\lambda_s(k,\ell)$ and $\lambda_t(k,\ell)$ represent the input noise power at a single sensor. Since the source signal, the stationary noise and transient noise are uncorrelated, the input PSD matrix is given by

$$\mathbf{\Phi}_{\mathbf{ZZ}}(k,\ell) = \lambda_x(k,\ell)\mathbf{A}(k)\mathbf{A}^H(k) + \lambda_s(k,\ell)\mathbf{\Gamma}_s(k,\ell) + \lambda_t(k,\ell)\mathbf{\Gamma}_t(k,\ell) \quad (3)$$

where $\lambda_x(k,\ell) \triangleq E\left\{|X(k,\ell)|^2\right\}$ is the PSD of the desired source signal. The PSD of the beamformer output and the PSD matrix of the reference signals are obtained by

$$\phi_{YY}(k,\ell) = [\mathbf{W}(k) - \mathbf{B}(k)\mathbf{H}(k,\ell)]^H \mathbf{\Phi}_{\mathbf{ZZ}}(k,\ell) \times [\mathbf{W}(k) - \mathbf{B}(k)\mathbf{H}(k,\ell)] \quad (4)$$
$$\mathbf{\Phi}_{\mathbf{UU}}(k,\ell) = \mathbf{B}^H(k)\mathbf{\Phi}_{\mathbf{ZZ}}(k,\ell)\mathbf{B}(k). \quad (5)$$

Substituting (3) into (4) and (5), we have the following linear relation between the PSDs of the beamformer output, the reference signals, the desired source signal, and the input interferences:

$$\begin{bmatrix} \phi_{YY} \\ \phi_{U_2U_2} \\ \vdots \\ \phi_{U_MU_M} \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} \\ \vdots & \vdots & \vdots \\ C_{M1} & C_{M2} & C_{M3} \end{bmatrix} \begin{bmatrix} \lambda_x \\ \lambda_s \\ \lambda_t \end{bmatrix} \quad (6)$$

where

$$[\,C_{11} \quad C_{12} \quad C_{13}\,] = [\mathbf{W} - \mathbf{B}\mathbf{H}]^H [\,\mathbf{A}\mathbf{A}^H \quad \mathbf{\Gamma}_s \quad \mathbf{\Gamma}_t\,] \times (\mathbf{I}_3 \otimes [\mathbf{W} - \mathbf{B}\mathbf{H}]) \quad (7)$$
$$[\,C_{21} \quad \cdots \quad C_{M1}\,] = \mathrm{diag}\left\{\mathbf{B}^H\mathbf{A}\mathbf{A}^H\mathbf{B}\right\} \quad (8)$$
$$[\,C_{22} \quad \cdots \quad C_{M2}\,] = \mathrm{diag}\left\{\mathbf{B}^H\mathbf{\Gamma}_s\mathbf{B}\right\} \quad (9)$$
$$[\,C_{23} \quad \cdots \quad C_{M3}\,] = \mathrm{diag}\left\{\mathbf{B}^H\mathbf{\Gamma}_t\mathbf{B}\right\} \quad (10)$$

$\mathbf{I}_3$ is a 3×3 identity matrix; $\otimes$ denotes Kronecker product; and $\mathrm{diag}\{\cdot\}$ represents a row vector constructed from the diagonal of a square matrix.

III. Signal Detection

Transient *signal* components are relatively strong at the beamformer output, whereas transient *noise* components are relatively strong at one of the reference signals. Hence, the transient power ratio between the beamformer output and the reference signals is expected to be large for desired transients, and small for noise components. Let $\mathcal{S}$ be a smoothing operator in the power spectral domain

$$\mathcal{S}Y(k,\ell) = \alpha_s \cdot \mathcal{S}Y(k,\ell-1) + (1-\alpha_s)\sum_{i=-w}^{w} b_i |Y(k-i,\ell)|^2 \quad (11)$$

where α_s ($0 \leq \alpha_s \leq 1$) is a parameter for the smoothing in time, and b is a normalized window function ($\sum_{i=-w}^{w} b_i = 1$) that determines the smoothing in frequency. Let $\mathcal{M}$ denote an estimator for the PSD of the background pseudostationary noise, derived using the MCRA approach [8]. We define the *transient beam-to-reference ratio* (TBRR) by the ratio between the transient power of the beamformer output and the transient power of the strongest reference signal

$$\Omega(k,\ell) = \frac{\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)}{\max\limits_{2\leq i\leq M}\{\mathcal{S}U_i(k,\ell) - \mathcal{M}U_i(k,\ell)\}}. \quad (12)$$

Let three hypotheses H_{0s}, H_{0t}, and H_1 indicate respectively absence of transients, presence of an interfering transient, and presence of a desired transient at the beamformer output. Then, given that H_1 or H_{0t} is true, we have

$$\begin{aligned}\Omega(k,\ell)\big|_{H_1\cup H_{0t}} &\approx \frac{\phi_{YY}(k,\ell) - C_{12}(k,\ell)\lambda_s(k,\ell)}{\max\limits_{2\leq i\leq M}\{\phi_{U_iU_i}(k,\ell) - C_{i2}(k,\ell)\lambda_s(k,\ell)\}} \\ &= \frac{C_{11}(k,\ell)\lambda_x(k,\ell) + C_{13}(k,\ell)\lambda_t(k,\ell)}{\max\limits_{2\leq i\leq M}\{C_{i1}(k)\lambda_x(k,\ell) + C_{i3}(k,\ell)\lambda_t(k,\ell)\}}. \end{aligned} \quad (13)$$

Assuming there exist thresholds $\Omega_{\mathrm{high}}(k)$ and $\Omega_{\mathrm{low}}(k)$ such that

$$\begin{aligned}\Omega(k,\ell)\big|_{H_{0t}} &\approx \frac{C_{13}(k,\ell)}{\max\limits_{2\leq i\leq M}\{C_{i3}(k,\ell)\}} \leq \Omega_{\mathrm{low}}(k) \leq \Omega_{\mathrm{high}}(k) \\ &\leq \frac{C_{11}(k,\ell)}{\max\limits_{2\leq i\leq M}\{C_{i1}(k)\}} \approx \Omega(k,\ell)\big|_{H_1} \end{aligned} \quad (14)$$

we determine the likelihood of signal presence proportionally to $\Omega(k,\ell)$ by

$$\psi(k,\ell) = \begin{cases} 0, & \text{if } \Omega(k,\ell) \leq \Omega_{\mathrm{low}}(k) \\ 1, & \text{if } \Omega(k,\ell) > \Omega_{\mathrm{high}}(k) \\ \dfrac{\Omega(k,\ell) - \Omega_{\mathrm{low}}(k)}{\Omega_{\mathrm{high}}(k) - \Omega_{\mathrm{low}}(k)}, & \text{otherwise.} \end{cases} \quad (15)$$

The decision rules for detecting transients at the beamformer output and reference signal are

$$\Lambda_Y(k,\ell) \triangleq \mathcal{S}Y(k,\ell)/\mathcal{M}Y(k,\ell) > \Lambda_0 \quad (16)$$
$$\Lambda_U(k,\ell) \triangleq \max_{2\leq i\leq M}\{\mathcal{S}U_i(k,\ell)/\mathcal{M}U_i(k,\ell)\} > \Lambda_1 \quad (17)$$

respectively, where Λ_Y and Λ_U denote measures of the local nonstationarities (LNS), and Λ_0 and Λ_1 are the corresponding threshold values for detecting transients. For a given signal, the LNS fluctuates about one in the absence of transients, and increases well above one in the neighborhood of time-frequency bins that contain transients. The false-alarm and detection probabilities are defined by

$$P_{f,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \,|\, H_{0s}\right) \quad (18)$$

$$P_{d,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \,|\, H_1 \cup H_{0t}\right) \tag{19}$$
$$P_{f,U}(k,\ell) = \mathcal{P}\left(\Lambda_U(k,\ell) > \Lambda_1 \,|\, H_{0s}\right) \tag{20}$$
$$P_{d,U}(k,\ell) = \mathcal{P}\left(\Lambda_U(k,\ell) > \Lambda_1 \,|\, H_1 \cup H_{0t}\right). \tag{21}$$

Then, for specified $P_{f,Y}$ and $P_{f,U}$, the required threshold values and the detection probabilities are given by [9]

$$\Lambda_0 = \frac{1}{\mu} F^{-1}_{\chi^2;\,\mu}\left(1 - P_{f,Y}\right) \tag{22}$$
$$P_{d,Y}(k,\ell) = 1 - F_{\chi^2;\,\mu}\left[\frac{1}{1+\xi_Y(k,\ell)} F^{-1}_{\chi^2;\,\mu}\left(1 - P_{f,Y}\right)\right] \tag{23}$$
$$\Lambda_1 = \frac{1}{\mu} F^{-1}_{\chi^2;\,\mu}\left[\left(1 - P_{f,U}\right)^{1/(M-1)}\right] \tag{24}$$
$$\begin{aligned} P_{d,U}(k,\ell) = 1 - &\left(1 - P_{f,U}\right)^{(M-2)/(M-1)} \\ &\cdot F_{\chi^2;\,\mu}\left(\frac{1}{1+\xi_U(k,\ell)} \right. \\ &\qquad \left. \cdot F^{-1}_{\chi^2;\,\mu}\left[\left(1 - P_{f,U}\right)^{1/(M-1)}\right]\right) \end{aligned} \tag{25}$$

where

$$\xi_Y(k,\ell) \triangleq \frac{C_{11}(k,\ell)\lambda_x(k,\ell) + C_{13}(k,\ell)\lambda_t(k,\ell)}{C_{12}(k,\ell)\lambda_s(k,\ell)}$$

and

$$\xi_U(k,\ell) \triangleq \max_{2\le i\le M}\left\{\frac{C_{i1}(k,\ell)\lambda_x(k,\ell) + C_{i3}(k,\ell)\lambda_t(k,\ell)}{C_{i2}(k,\ell)\lambda_s(k,\ell)}\right\}$$

represent the ratio between the transient and pseudostationary power at the beamformer output and reference signals, and $F_{\chi^2;\,\mu}(x)$ denotes the standard chi-square distribution function with μ degrees of freedom. Fig. 2 shows the receiver operating characteristic (ROC) curves for detection of transients at the beamformer output and reference signals, with the false-alarm probability as parameter. Four sensors are used, and μ is set to 32.2 (this value of μ is obtained for a smoothing $\mathcal{S}$ of the form (11), with $\alpha_s = 0.9$, and $b = [0.25\ 0.5\ 0.25]$).

Fig. 3 summarizes a block diagram for the detection of desired source components at the beamformer output. The detection is carried out in the time-frequency plane for each frame and frequency bin. First, we detect transients at the beamformer output. If there are no simultaneous transients at the reference noise signals, the transient is presumably a desirable source component. However, if transients are simultaneously detected at the beamformer output and at one of the reference signals, the likelihood of signal presence is determined by the TBRR. The larger the TBRR is, the higher the likelihood that a transient comes from a desired source.

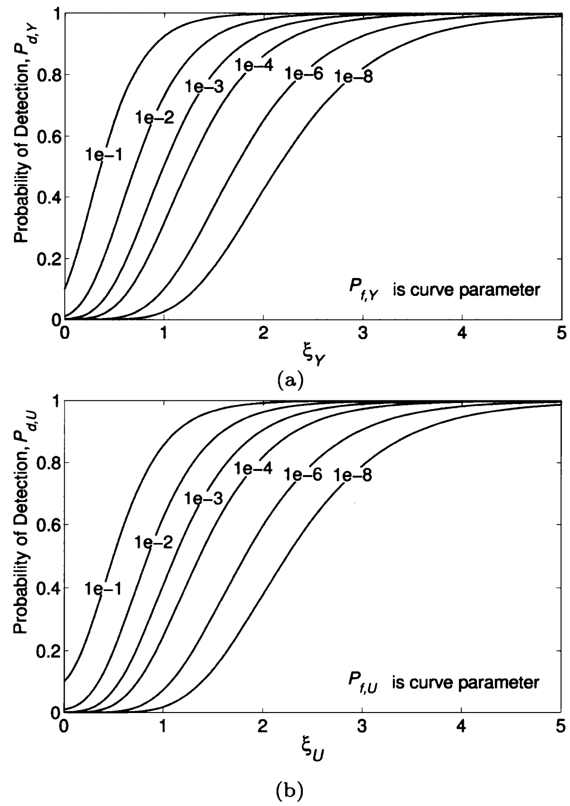

Fig. 2. Receiver operating characteristic curves for detection of transients at (a) the beamformer output, and at (b) the reference noise signals, using $M = 4$ sensors ($\mu = 32.2$).

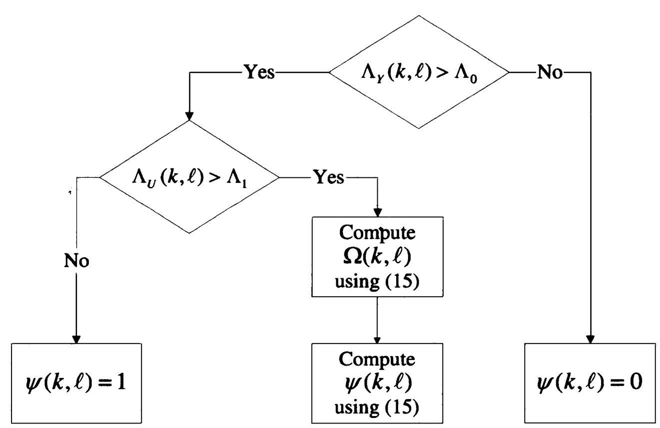

Fig. 3. Block diagram for detection of desired source components at the beamformer output.

IV. Conclusion

We have described a multichannel signal detection approach, that is particularly advantageous in nonstationary noise environments. The beamformer is realistically assumed to have a steering error, a blocking matrix that is unable to block all of the desired signal components, and a noise canceler that is adapted to the pseudostationary noise, but not modified during transient interferences. Accordingly, the reference noise signals may include some desired signal components. Furthermore, transient noise components that leak through the sidelobes of the fixed beamformer may proceed to the beamformer primary output. A mild assumption is made with regard to the beamformer, that a desired signal component is stronger at the beamformer output than at any reference noise signal, and a noise component is strongest at one of the reference signals. Consequently, signal components are detected at the beamformer output based on a measure of their local nonstationarity, and discriminated from transient noise components based on the transient beam-to-reference ratio.

REFERENCES

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

[2] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with post-filtering," *IEEE Trans. Speech Audio Processing*, vol. 6, pp. 240–259, May 1998.

[3] K. U. Simmer, J. Bitzer, and C. Marro, "Post-filtering techniques," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. S. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001, ch. 3, pp. 39–60.

[4] I. Cohen and B. Berdugo, "Microphone array post-filtering for nonstationary noise suppression," in *Proc. ICASSP*, Orlando, FL, May 13–17, 2002, pp. 901–904.

[5] 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.

[6] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. AP-30, pp. 27–34, Jan. 1982.

[7] C. W. Jim, "A comparison of two LMS constrained optimal array structures," *Proc. IEEE*, vol. 65, pp. 1730–1731, Dec. 1977.

[8] I. Cohen and B. Berdugo, "Noise estimation by minima controlled recursive averaging for robust speech enhancement," *IEEE Signal Processing Lett.*, vol. 9, pp. 12–15, Jan. 2002.

[9] I. Cohen, "Multi-channel post-filtering in nonstationary noise environments," Technion—Israel Institute of Technology, Haifa, Israel, Tech. Rep., CCIT Rep. 376, EE Pub. 1314, Apr. 2002.