CCIT Report #400

October 2002

An Integrated Real-Time Beamforming and Postfiltering System for Non-Stationary Noise Environments

Israel Cohen, Sharon Gannot and Baruch Berdugo

Abstract

In this paper, we present a novel approach for real-time multichannel speech enhancement in environments of non-stationary noise and time-varying acoustical transfer functions (ATFs). The proposed system integrates adaptive beamforming, ATF identification, soft signal detection, and multichannel postfiltering. The noise canceller branch of the beamformer and the ATF identification are adaptively updated on-line based on hypothesis test results. The noise canceller is updated only during stationary noise frames, and the ATF identification is carried out only when desired source components have been detected. The hypothesis testing is based on the non-stationarity of the signals and the transient power ratio between the beamformer primary output and its reference noise signals. Following the beamforming and the hypothesis testing, estimates for the signal presence probability and for the noise power spectral density are derived. Subsequently, an optimal spectral gain function is applied that minimizes the mean-square error of the log-spectral amplitude. Experimental results demonstrate the usefulness of the proposed system in non-stationary noise environments.

Keywords

Array signal processing, signal detection, acoustic noise measurement, speech enhancement, spectral analysis, adaptive signal processing.

I. Cohen and S. Gannot are with the Department of Electrical Engineering, Technion - Israel Institute of Technology, Technion City, Haifa 32000, Israel (email: icohen@ee.technion.ac.il; gannot@siglab.technion.ac.il).

B. Berdugo is with Lamar Signal Processing Ltd., Andrea Electronics Corp. - Israel, P.O.Box 573, Yokneam Ilit 20692, Israel (email: bberdugo@lamar.co.il).

I. Introduction

Postfiltering methods for multi-microphone speech enhancement algorithms have recently attracted an increased interest. It is well known that an improvement in speech quality is encountered by beamforming methods [1]. However, when the noise field is spatially incoherent (*e.g.*, diffuse noise field) significant performance degradation entails additional postfiltering [2]. Most multi-microphone speech enhancement methods comprise a multichannel part (either *delay and sum* beamformer or *generalized sidelobe canceller* (GSC) [3]) followed by a postfilter, which is based on Wiener filtering (sometimes in conjunction with spectral subtarction). Numerous articles have been published on the subject, e.g. [4], [5], [6], [7], [8], [9], [10], [11], [12] to mention just a few. A major drawback of these multichannel postfiltering techniques is that highly non-stationary 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. Unfortunately, transient interferences are often much too brief and abrupt for the conventional tracking methods.

Recently, a multichannel postfilter was incorporated into the GSC beamformer [13], [14]. The use of both the beamformer primary output and the reference noise signals (resulting from the blocking branch of the GSC) for distinguishing between speech transients and noise transients, enables the algorithm to work in non-stationary noise environments. In [15], the multichannel postfilter is combined with the *transfer function GSC* (TF-GSC) [16], and compared with single-microphone postfilters, namely the Mixture-Maximum (MIXMAX) [17] and the *optimally modified log spectral amplitude estimator* (OM-LSA) [18]. The multichannel postfilter, combined with the TF-GSC, proved the best for handling abrupt noise spectral variations. However, in all past contributions the beamformer stage feeds the postfilter, but the adverse is not true. The decisions made by the postfilter, distinguishing between speech, stationary noise and transient noise, might be fed back to the beamformer to enable the use of the method in real-time applications. Exploiting this information will also enable tracking of the acoustical transfer functions (ATFs), caused by talker movements.

In this paper, we present a real-time multichannel speech enhancement system, which integrates

adaptive beamforming and multichannel postfiltering. The beamformer is based on the TF-GSC. However, the requirement for the stationarity of the noise is relaxed. Furthermore, we allow the ATFs to vary in time, which entails an on-line system identification procedure. We define hypotheses that indicate absence of transients, presence of an interfering transient, and presence of desired source components. The noise canceller branch of the beamformer is updated only during the stationary noise frames, and the ATF identification is carried out only when desired source components are present. Following the beamforming and the hypothesis testing, estimates for the signal presence probability and for the noise power spectral density (PSD) are derived. Subsequently, an optimal spectral gain function is applied that minimizes the mean-square error of the log-spectral amplitude.

The performance of the proposed system is evaluated under non-stationary noise conditions, and compared to that obtained with a single-channel postfiltering approach. We show that single-channel postfiltering is inefficient at attenuating highly non-stationary noise components, since it lacks the ability to differentiate such components from the desired source components. By contrast, the proposed system achieves a significantly reduced level of background noise, whether stationary or not, without further distorting the signal components.

The paper is organized as follows. In Section II, we introduce a novel approach for real-time beamforming in non-stationary noise environments, under the circumstances of time-varying ATFs. The noise canceller branch of the beamformer and the ATF identification are adaptively updated on-line based on hypothesis test results. In Section III, the problem of hypothesis testing in the time-frequency plane is addressed. Signal components are detected and discriminated from the transient noise components based on the transient power ratio between the beamformer primary output and its reference noise signals. In Section IV, we introduce the multichannel postfilter, and outline the implementation steps of the integrated TF-GSC and multichannel postfiltering algorithm. Finally, in Section V, we evaluate the proposed system, and present experimental results, which validate its usefulness.

II. Transfer Function Generalized Sidelobe Cancelling

Let $x(t)$ denote a desired speech source signal that, subject to some acoustic propagation, is received by M microphones along with additive uncorrelated interfering signals. The interference at the ith sensor comprises a pseudo-stationary noise signal, $d_{is}(t)$, and a transient noise component, $d_{it}(t)$. The observed signals are given by

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

where $a_i(t)$ is the acoustical transfer function from the desired source to the ith sensor, and $*$ denotes convolution. Using the short-time Fourier transform (STFT), we have in the time-frequency domain

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

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

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

The observed noisy signals are processed by the system shown in Fig. 1. This structure is a modification to the recently proposed TF-GSC [16], which is an extension of the linearly constrained adaptive beamformer [3], [19] for the arbitrary transfer function case. In [16], transient interferences are not dealt with, as signal enhancement is based on the non-stationarity of the desired source signal, contrasted with the noise signal stationarity. As such, the ATF estimation was conducted in an off-line manner. Here, the requirement for the stationarity of the noise is relaxed. So a mechanism for discriminating interfering transients from desired signal components must be included. Furthermore, in contrast to the assumption of time-invariant ATFs in [16], we allow time-varying ATFs provided that their change rate is slow in comparison to that of the speech statistics. This entails on-line adaptive estimates for the ATFs.

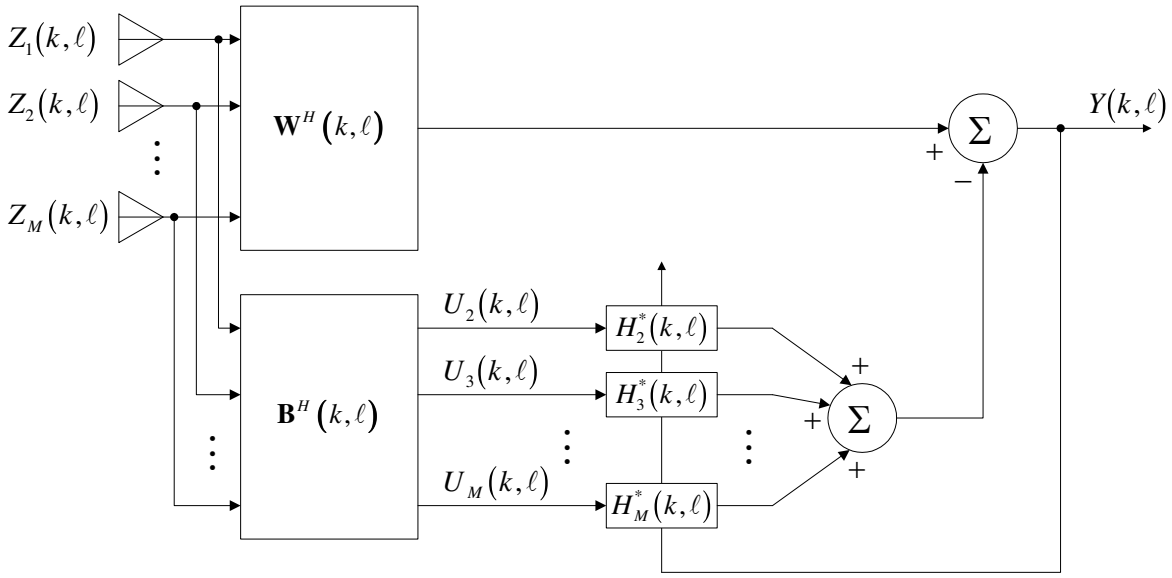


Fig. 1. Block diagram of the transfer function generalized sidelobe canceller (TF-GSC).

The beamformer comprises three parts: a fixed beamformer $\mathbf{W}$, which aligns the desired signal components; a blocking matrix $\mathbf{B}$, which blocks the desired components thus yielding the reference noise signals $\{U_i \ : \ 2 \leq i \leq M\}$; and a multichannel adaptive noise canceller $\{H_i \ : \ 2 \leq i \leq M\}$, which eliminates the stationary noise that leaks through the sidelobes of the fixed beamformer. The reference noise signals $\mathbf{U}(k,\ell) = \begin{bmatrix} U_2(k,\ell) & U_3(k,\ell) & \cdots & U_M(k,\ell) \end{bmatrix}^T$ are generated by applying the blocking matrix to the observed signal vector:

$$\begin{aligned} \mathbf{U}(k,\ell) &= \mathbf{B}^H(k,\ell)\mathbf{Z}(k,\ell) \\ &= \mathbf{B}^H(k,\ell)\left[\mathbf{A}(k,\ell)X(k,\ell) + \mathbf{D}_s(k,\ell) + \mathbf{D}_t(k,\ell)\right] . \end{aligned} \tag{3}$$

The reference noise signals are emphasized by the adaptive noise canceller and subtracted from the output of the fixed beamformer, yielding

$$Y(k,\ell) = \left[\mathbf{W}^H(k,\ell) - \mathbf{H}^H(k,\ell)\mathbf{B}^H(k,\ell)\right]\mathbf{Z}(k,\ell)\,, \tag{4}$$

where $\mathbf{H}(k,\ell) = \begin{bmatrix} H_2(k,\ell) & H_3(k,\ell) & \cdots & H_M(k,\ell) \end{bmatrix}^T$. It is worth mentioning that a perfect blocking matrix implies $\mathbf{B}^H(k,\ell)\mathbf{A}(k,\ell) = 0$. In that case, $\mathbf{U}(k,\ell)$ indeed contains only noise components,

$$\mathbf{U}(k,\ell) = \mathbf{B}^H(k,\ell)\left[\mathbf{D}_s(k,\ell) + \mathbf{D}_t(k,\ell)\right] .$$

In general, however, $\mathbf{B}^H(k,\ell)\mathbf{A}(k,\ell) \neq 0$, thus desired signal components may leak into the noise reference signals.

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 source transient at the beamformer output. The optimal solution for the filters $\mathbf{H}(k,\ell)$ is obtained by minimizing the power of the beamformer output during the stationary noise frames (*i.e.*, when H_{0s} is true) [20]. Let $\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell) = E\left\{\mathbf{D}_s(k,\ell)\mathbf{D}_s^H(k,\ell)\right\}$ denote the power-spectral density (PSD) matrix of the input stationary noise. Then, the power of the stationary noise at the beamformer output is minimized by solving the unconstrained optimization problem:

$$\min_{\mathbf{H}} \left\{ \left[\mathbf{W}(k,\ell) - \mathbf{B}(k,\ell)\mathbf{H}(k,\ell)\right]^H \mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell) \left[\mathbf{W}(k,\ell) - \mathbf{B}(k,\ell)\mathbf{H}(k,\ell)\right] \right\} . \tag{5}$$

A multichannel Wiener solution is given by [21]

$$\mathbf{H}(k,\ell) = \left[\mathbf{B}^H(k,\ell)\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell)\mathbf{B}(k)\right]^{-1} \mathbf{B}^H(k,\ell)\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell)\mathbf{W}(k,\ell)\,. \tag{6}$$

In practice, this optimization problem is solved by using the normalized LMS algorithm [20]:

$$\mathbf{H}(k,\ell+1) = \begin{cases} \mathbf{H}(k,\ell) + \frac{\mu_h}{P_{\text{est}}(k,\ell)}\mathbf{U}(k,\ell)Y^*(k,\ell) & \text{if } H_{0s} \text{ is true,} \\ \mathbf{H}(k,\ell)\,, & \text{otherwise,} \end{cases} \tag{7}$$

where

$$P_{\text{est}}(k,\ell) = \alpha_p P_{\text{est}}(k,\ell-1) + (1-\alpha_p)\|\mathbf{U}(k,\ell)\|^2 \tag{8}$$

represents the power of the noise reference signals, μ_h is a step factor that regulates the convergence rate, and α_p is a smoothing parameter.

The fixed beamformer implements the alignment of the desired signal by applying a matched filter to the ATF ratios [16]:

$$\mathbf{W}(k,\ell) \triangleq \frac{\tilde{\mathbf{A}}(k,\ell)}{\|\tilde{\mathbf{A}}(k,\ell)\|^2}$$

where

$$\tilde{\mathbf{A}}(k,\ell) \triangleq \frac{\mathbf{A}(k,\ell)}{A_1(k,\ell)} = \left[1 \quad \tfrac{A_2(k,\ell)}{A_1(k,\ell)} \quad \cdots \quad \tfrac{A_M(k,\ell)}{A_1(k)}\right]^T \triangleq \left[1 \quad \tilde{A}_2(k,\ell) \quad \cdots \quad \tilde{A}_M(k,\ell)\right]^T$$

denotes ATF ratios, with $A_1(k,\ell)$ chosen arbitrarily as the reference ATF. The blocking matrix $\mathbf{B}$ is aimed at eliminating the desired signal and constructing reference noise signals. A proper (but not unique) choice of the blocking matrix is given by [16],

$$\mathbf{B}(k,\ell) = \begin{bmatrix} -\tilde{A}_2^*(k,\ell) & -\tilde{A}_3^*(k,\ell) & \dots & -\tilde{A}_M^*(k,\ell) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \dots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{9}$$

Hence, for implementing both the fixed beamformer and the blocking matrix, we need to estimate the ATF ratios. In contrast to previous works [16], [15], [14], the system identification should be incorporated into the adaptive procedure, since the ATFs are time-varying. In [16], the system identification procedure is based on the nonstationarity of the desired signal. Here, a modified version is introduced, employing the already available time-frequency analysis of the beamformer and the decisions made by hypothesis testing.

From (3) and (9) we have the following input-output relation between $Z_1(k,\ell)$ and $Z_i(k,\ell)$:

$$Z_i(k,\ell) = \tilde{A}_i(k,\ell) Z_1(k,\ell) + U_i(k,\ell)\,, \quad i = 2,\dots,M\,. \tag{10}$$

Accordingly,

$$\phi_{Z_i Z_1}(k,\ell) = \tilde{A}_i(k,\ell)\phi_{Z_1 Z_1}(k,\ell) + \phi_{U_i Z_1}(k,\ell), \quad i = 2,\dots,M \tag{11}$$

where $\phi_{Z_i Z_1}(k,\ell) = E\left\{Z_i(k,\ell) Z_1^*(k,\ell)\right\}$ is the cross-PSD between $z_i(t)$ and $z_1(t)$, and $\phi_{U_i Z_1}(k,\ell)$ is the cross-PSD between $u_i(t)$ and $z_1(t)$. The use of standard system identification methods is inapplicable, since the interference signal $u_i(t)$ is strongly correlated to the system input $z_1(t)$. However, when the hypothesis H_1 is true, *i.e.* when transient noise is absent, the cross-PSD $\phi_{U_i Z_1}(k,\ell)$ becomes stationary. Therefore, $\phi_{U_i Z_1}(k,\ell)$ may be replaced with $\phi_{U_i Z_1}(k)$.

For estimating the ATF ratios $\tilde{\mathbf{A}}(k,\ell)$, we need to collect several estimates of the PSD $\boldsymbol{\phi}_{\mathbf{Z}Z_1}(k,\ell)$, each of which is based on averaging several frames. Let a segment define a concatenation of N frames for which the hypothesis H_1 is true, and let an interval contain R such segments. Then, the PSD estimation in each segment r ($r = 1,\ldots,R$) is obtained by averaging the periodograms over N frames:

$$\hat{\boldsymbol{\phi}}^{(r)}_{\mathbf{Z}Z_1}(k,\ell) = \frac{1}{N}\sum_{\ell\in\mathcal{L}_r}\mathbf{Z}(k,\ell)Z_1^*(k,\ell)$$

where $\mathcal{L}_r$ represents the set of frames that belong to the rth segment. Denoting by $\varepsilon_i^{(r)}(k,\ell) = \hat{\phi}^{(r)}_{U_iZ_1}(k,\ell) - \phi_{U_iZ_1}(k)$ the estimation error of the cross-PSD between $u_i(t)$ and $z_1(t)$ in the rth segment, eq. (11) implies

$$\hat{\phi}^{(r)}_{Z_iZ_1}(k,\ell) = \tilde{A}_i(k,\ell)\hat{\phi}^{(r)}_{Z_1Z_1}(k,\ell) + \phi_{U_iZ_1}(k) + \varepsilon_i^{(r)}(k,\ell),\ i = 2,\ldots,M,\ r = 1,2,\ldots,R. \tag{12}$$

The *least-squares* (LS) solution to this over-determined set of equation is given by [16]

$$\tilde{\mathbf{A}}(k,\ell) = \frac{\left\langle\hat{\phi}_{Z_1Z_1}(k,\ell)\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}(k,\ell)\right\rangle - \left\langle\hat{\phi}_{Z_1Z_1}(k,\ell)\right\rangle\left\langle\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}(k,\ell)\right\rangle}{\left\langle\hat{\phi}^2_{Z_1Z_1}(k,\ell)\right\rangle - \left\langle\hat{\phi}_{Z_1Z_1}(k,\ell)\right\rangle^2} \tag{13}$$

where the average operation on $\beta(k,\ell)$ is defined by

$$\langle\beta(k,\ell)\rangle \triangleq \frac{1}{R}\sum_{r=1}^{R}\beta^{(r)}(k,\ell)\,.$$

Practically, the estimates for $\hat{\boldsymbol{\phi}}^{(r)}_{\mathbf{Z}Z_1}(k,\ell)$ ($r = 1,\ldots,R$) are recursively obtained as follows. In each time-frequency bin (k,ℓ) we assume that R PSD estimates are already available (excluding initial conditions). Values of $\tilde{\mathbf{A}}(k,\ell)$ are thus ready for use in the next frame $(k,\ell+1)$. Frames, for which the hypothesis H_1 is true, are collected for obtaining a new PSD estimate $\hat{\boldsymbol{\phi}}^{(R+1)}_{\mathbf{Z}Z_1}(k,\ell)$,

$$\hat{\boldsymbol{\phi}}^{(R+1)}_{\mathbf{Z}Z_1}(k,\ell+1) = \hat{\boldsymbol{\phi}}^{(R+1)}_{\mathbf{Z}Z_1}(k,\ell) + \frac{1}{N}\mathbf{Z}(k,\ell)Z_1^*(k,\ell)\,. \tag{14}$$

A counter n_k is used for counting the times eq. (14) is processed (counting the number of H_1 frames in frequency bin k). Whenever n_k reaches N, the estimate in segment $R+1$ is stacked into the

previous estimates, the oldest estimate ($r = 1$) is discarded, and n_k is initialized. The new R estimates are then used for obtaining a new estimate for the ATF ratios $\tilde{\mathbf{A}}(k, \ell+1)$ for the next bin $(k, \ell+1)$. This procedure is active for all frames ℓ enabling a real-time tracking of the beamformer.

Altogether, an interval containing $N \times R$ frames, for which H_1 is true, is used for obtaining an estimate for $\tilde{\mathbf{A}}(k, \ell)$. Special attention should be given for choosing this quantity. On the one hand, it should be long enough for stabilizing the solution. On the other hand, it should be short enough for the ATF quasi-stationarity assumption to hold during the interval. We note that for frequency bins with low speech content, the interval (observation time) required for obtaining an estimate for $\tilde{\mathbf{A}}(k, \ell)$ might be very long, since only frames for which H_1 is true are collected.

III. Hypothesis Testing

Generally, the TF-GSC output comprises three components: a non-stationary desired source component, a pseudo-stationary noise component, and a transient interference. Our objective is to determine which category a given time-frequency bin belongs to, based on the beamformer output and the reference signals. Clearly, if transients have not been detected at the beamformer output and the reference signals, we can accept the H_{0s} hypothesis. In case a transient is detected at the beamformer output, but not at the reference signals, the transient is likely a source component and therefore we determine that H_1 is true. On the contrary, a transient that is detected at one of the reference signals but not at the beamformer output is likely an interfering component, which implies that H_{0t} is true. In case a transient is simultaneously detected at the beamformer output and at one of the reference signals, a further test is required, which involves the ratio between the transient power at beamformer output and the transient power at the reference signals.

From (3) and (4), the PSD-matrix of the reference signals and the PSD of the beamformer output are obtained by

$$\mathbf{\Phi_{UU}}(k, \ell) = \mathbf{B}^H(k)\mathbf{\Phi_{ZZ}}(k, \ell)\mathbf{B}(k) \tag{15}$$

$$\phi_{YY}(k, \ell) = \left[\mathbf{W}(k) - \mathbf{B}(k)\mathbf{H}(k, \ell)\right]^H \mathbf{\Phi_{ZZ}}(k, \ell) \left[\mathbf{W}(k) - \mathbf{B}(k)\mathbf{H}(k, \ell)\right] . \tag{16}$$

If we assume that the stationary, as well as transient, noise fields are homogeneous, then the 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

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

where $\lambda_s(k,\ell)$ and $\lambda_t(k,\ell)$ represent the input noise power at a single sensor. The input PSD-matrix is therefore given by

$$\mathbf{\Phi}_{\mathbf{ZZ}}(k,\ell) = \lambda_x(k,\ell)\mathbf{A}(k,\ell)\mathbf{A}^H(k,\ell) + \lambda_s(k,\ell)\mathbf{\Gamma}_s(k,\ell) + \lambda_t(k,\ell)\mathbf{\Gamma}_t(k,\ell) \tag{17}$$

where $\lambda_x(k,\ell) \triangleq E\left\{|X(k,\ell)|^2\right\}$ is the PSD of the desired source signal. Substituting (17) into (15) and (16), we have the following linear relation between the PSD's of the beamformer output, the reference signals, the desired source signal, and the input interferences [14]:

$$\begin{bmatrix} \phi_{YY}(k,\ell) \\ \phi_{U_2U_2}(k,\ell) \\ \vdots \\ \phi_{U_MU_M}(k,\ell) \end{bmatrix} = \begin{bmatrix} C_{11}(k,\ell) & C_{12}(k,\ell) & C_{13}(k,\ell) \\ \vdots & \vdots & \vdots \\ C_{M1}(k,\ell) & C_{M2}(k,\ell) & C_{M3}(k,\ell) \end{bmatrix} \begin{bmatrix} \lambda_x(k,\ell) \\ \lambda_s(k,\ell) \\ \lambda_t(k,\ell) \end{bmatrix} \tag{18}$$

where

$$\begin{aligned}\begin{bmatrix} C_{11} & C_{12} & C_{13} \end{bmatrix} &= [\mathbf{W}-\mathbf{BH}]^H \begin{bmatrix} \mathbf{AA}^H & \mathbf{\Gamma}_s & \mathbf{\Gamma}_t \end{bmatrix} (\mathbf{I}_3 \otimes [\mathbf{W}-\mathbf{BH}]) && (19)\\ \begin{bmatrix} C_{21} & \cdots & C_{M1} \end{bmatrix} &= \mathrm{diag}\left\{\mathbf{B}^H\mathbf{AA}^H\mathbf{B}\right\} && (20)\\ \begin{bmatrix} C_{22} & \cdots & C_{M2} \end{bmatrix} &= \mathrm{diag}\left\{\mathbf{B}^H\mathbf{\Gamma}_s\mathbf{B}\right\} && (21)\\ \begin{bmatrix} C_{23} & \cdots & C_{M3} \end{bmatrix} &= \mathrm{diag}\left\{\mathbf{B}^H\mathbf{\Gamma}_t\mathbf{B}\right\}, && (22)\end{aligned}$$

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

Let $\mathcal{S}$ be a smoothing operator in the power spectral domain,

$$\mathcal{S}Y(k,\ell) = \alpha_s \cdot \mathcal{S}Y(k,\ell-1) + (1-\alpha_s)\sum_{i=-w}^{w} b_i|Y(k-i,\ell)|^2 \tag{23}$$

where α_s ($0 \le \alpha_s \le 1$) is a forgetting factor for the smoothing in time, and b is a normalized window function ($\sum_{i=-w}^{w} b_i = 1$) that determines the order of smoothing in frequency. Let $\mathcal{M}$ denote

an estimator for the PSD of the background pseudo-stationary noise, derived using the *Minima Controlled Recursive Averaging* approach [18], [22]. The decision rules for detecting transients at the TF-GSC output and reference signals are

$$\Lambda_Y(k,\ell) \triangleq \mathcal{S}Y(k,\ell)/\,\mathcal{M}Y(k,\ell) > \Lambda_0 \tag{24}$$

$$\Lambda_{\mathbf{U}}(k,\ell) \triangleq \max_{2\le i\le M}\left\{\frac{\mathcal{S}U_i(k,\ell)}{\mathcal{M}U_i(k,\ell)}\right\} > \Lambda_1\,, \tag{25}$$

respectively, where Λ_Y and Λ_U denote measures of the local non-stationarities (LNS) [14], and Λ_0 and Λ_1 are the corresponding threshold values for detecting transients. For a given signal, the LNS fluctuates about 1 in the absence of transients, and increases well above 1 in the neighborhood of time-frequency bins that contain transients. The false alarm and detection probabilities are defined by

$$P_{f,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \mid H_{0s}\right) \tag{26}$$

$$P_{d,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \mid H_1 \cup H_{0t}\right) \tag{27}$$

$$P_{f,U}(k,\ell) = \mathcal{P}\left(\Lambda_U(k,\ell) > \Lambda_1 \mid H_{0s}\right) \tag{28}$$

$$P_{d,U}(k,\ell) = \mathcal{P}\left(\Lambda_U(k,\ell) > \Lambda_1 \mid H_1 \cup H_{0t}\right)\,. \tag{29}$$

Then for specified $P_{f,Y}$ and $P_{f,U}$, the required threshold values and the detection probabilities are given by [14]

$$\Lambda_0 = \frac{1}{\mu}F^{-1}_{\chi^2;\mu}\left(1-P_{f,Y}\right) \tag{30}$$

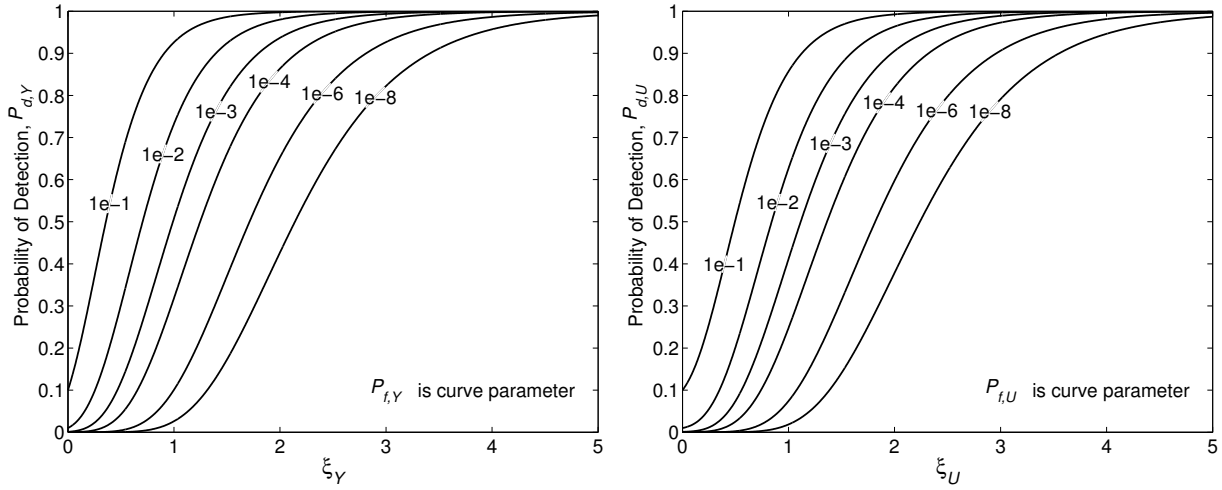
$$P_{d,Y}(k,\ell) = 1 - F_{\chi^2;\mu}\left[\frac{1}{1+\xi_Y(k,\ell)}F^{-1}_{\chi^2;\mu}\left(1-P_{f,Y}\right)\right] \tag{31}$$

$$\Lambda_1 = \frac{1}{\mu}F^{-1}_{\chi^2;\mu}\left[(1-P_{f,U})^{\frac{1}{M-1}}\right] \tag{32}$$

$$P_{d,U}(k,\ell) = 1-(1-P_{f,U})^{\frac{M-2}{M-1}}F_{\chi^2;\mu}\left(\frac{1}{1+\xi_U(k,\ell)}F^{-1}_{\chi^2;\mu}\left[(1-P_{f,U})^{\frac{1}{M-1}}\right]\right) \tag{33}$$

where

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



(a) (b)

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 ($\mu = 32.2$).

and

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

represent the ratios between the transient and pseudo-stationary 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 (23), with $\alpha_s = 0.9$, and $b = \begin{bmatrix} 0.25 & 0.5 & 0.25 \end{bmatrix}$).

The *transient beam-to-reference ratio* (TBRR) is defined by the ratio between the transient power of the beamformer output and the transient power of the strongest reference signal:

$$\Omega(k,\ell) = \frac{\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)}{\max\limits_{2\le i\le M} \{\mathcal{S}U_i(k,\ell) - \mathcal{M}U_i(k,\ell)\}} \,. \tag{34}$$

Given that H_1 or H_{0t} is true, we have

$$\begin{aligned}\Omega(k,\ell)|_{H_1\cup H_{0t}} &\approx \frac{\phi_{YY}(k,\ell)-C_{12}(k,\ell)\lambda_s(k,\ell)}{\max\limits_{2\le i\le M}\{\phi_{U_iU_i}(k,\ell)-C_{i2}(k,\ell)\lambda_s(k,\ell)\}}\\ &= \frac{C_{11}(k,\ell)\lambda_x(k,\ell)+C_{13}(k,\ell)\lambda_t(k,\ell)}{\max\limits_{2\le i\le M}\{C_{i1}(k)\lambda_x(k,\ell)+C_{i3}(k,\ell)\lambda_t(k,\ell)\}}\,. \end{aligned} \tag{35}$$

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

$$\begin{aligned}\Omega(k,\ell)|_{H_{0t}} &\approx \frac{C_{13}(k,\ell)}{\max\limits_{2\le i\le M}\{C_{i3}(k,\ell)\}} \le \Omega_{\text{low}}(k) \le \Omega_{\text{high}}(k)\\ &\le \frac{C_{11}(k,\ell)}{\max\limits_{2\le i\le M}\{C_{i1}(k)\}} \approx \Omega(k,\ell)|_{H_1} \end{aligned} \tag{36}$$

the decision rule for differentiating desired signal components from the transient interference components is

$$\begin{aligned} H_{0t} &: \gamma_s(k,\ell)\le 1 \text{ or } \Omega(k,\ell)\le\Omega_{\text{low}}(k)\\ H_1 &: \gamma_s(k,\ell)\ge\gamma_0 \text{ and } \Omega(k,\ell)\ge\Omega_{\text{high}}(k)\\ H_r &: \text{otherwise}\end{aligned} \tag{37}$$

where

$$\gamma_s(k,\ell)\triangleq\frac{|Y(k,\ell)|^2}{\mathcal{M}Y(k,\ell)} \tag{38}$$

represents the *a posteriori* SNR at the beamformer output with respect to the pseudo-stationary noise, γ_0 denotes a constant satisfying $\mathcal{P}\left(\gamma_s(k,\ell)\ge\gamma_0 \mid H_{0s}\right)<\epsilon$ for a certain significance level ϵ, and H_r designates a *reject* option where the conditional error of making a decision between H_{0t} and H_1 is high.

Fig. 3 summarizes a block diagram for the hypothesis testing. The hypothesis testing is carried out in the time-frequency plane for each frame and frequency bin. H_{0s} is accepted when transients have neither been detected at the beamformer output nor at the reference signals. In case a transient is detected at the beamformer output but not at the reference signals, we accept H_1. On the other

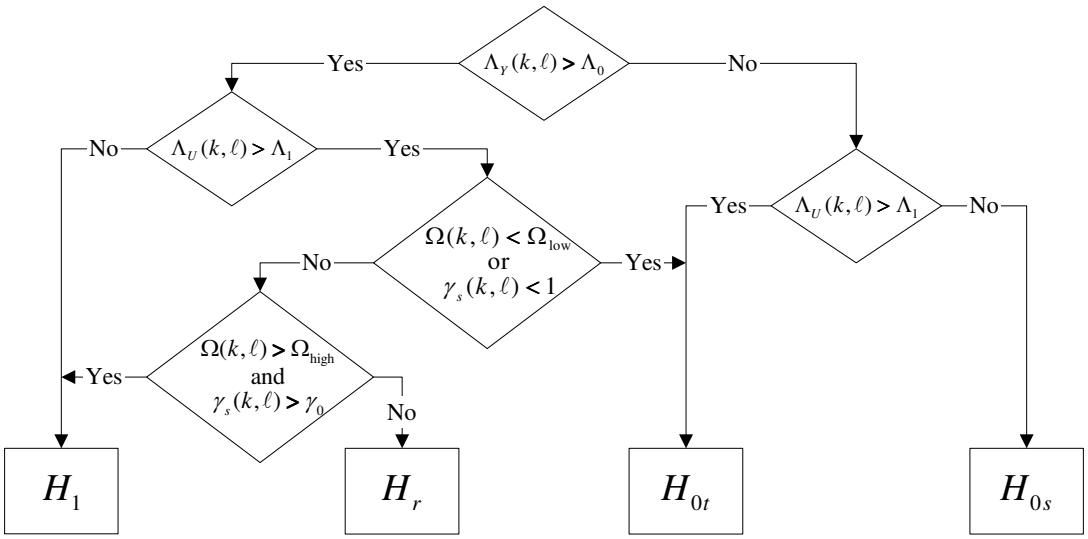


Fig. 3. Block diagram for the hypothesis testing.

hand, if a transient is detected at one of the reference signals but not at the beamformer output, we accept H_{0t}. In case a transient is detected simultaneously at the beamformer output and at one of the reference signals, we compute the TBRR, $\Omega(k,\ell)$, and the *a posteriori* SNR at the beamformer output with respect to the pseudo-stationary noise, $\gamma_s(k,\ell)$, and decide on the hypothesis according to (37).

IV. Multichannel Postfiltering

In this section, we address the problem of estimating the time-varying PSD of the TF-GSC output noise, and present the multichannel postfiltering technique. Fig. 4 describes a block diagram of the multichannel postfiltering. Following the hypothesis testing, an estimate $\hat{q}(k,\ell)$ for the *a priori* signal absence probability is produced. Subsequently, we derive an estimate $p(k,\ell) \triangleq \mathcal{P}\left(H_1 \mid Y, \mathbf{U}\right)$ for the signal presence probability, and an estimate $\hat{\lambda}_d(k,\ell)$ for the noise PSD. Finally, spectral enhancement of the beamformer output is achieved by applying the OM-LSA gain function [18], which minimizes the mean-square error of the log-spectral amplitude under signal presence uncertainty.

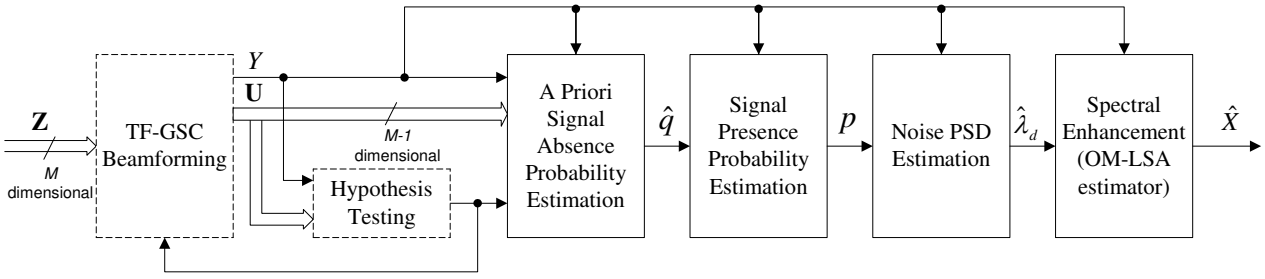


Fig. 4. Block diagram of the multichannel postfiltering.

Based on a Gaussian statistical model [23], the signal presence probability is given by

$$p(k,\ell) = \left\{1 + \frac{q(k,\ell)}{1-q(k,\ell)}(1+\xi(k,\ell))\exp(-\upsilon(k,\ell))\right\}^{-1} \tag{39}$$

where $\xi(k,\ell) \triangleq \lambda_x(k,\ell)/\lambda_d(k,\ell)$ is the *a priori* SNR, $\lambda_d(k,\ell)$ is the noise PSD at the beamformer output, $\upsilon(k,\ell) \triangleq \gamma(k,\ell)\,\xi(k,\ell)/(1+\xi(k,\ell))$, and $\gamma(k,\ell) \triangleq |Y(k,\ell)|^2/\lambda_d(k,\ell)$ is the *a posteriori* SNR. The *a priori* signal absence probability $\hat{q}(k,\ell)$ is set to 1 if signal absence hypotheses (H_{0s} or H_{0t}) are accepted, and is set to 0 if signal presence hypothesis (H_1) is accepted. In case of the reject hypothesis H_r, a soft signal detection is accomplished by letting $\hat{q}(k,\ell)$ be inversely proportional to $\Omega(k,\ell)$ and $\gamma_s(k,\ell)$:

$$\hat{q}(k,\ell) = \max\left\{\frac{\gamma_0 - \gamma_s(k,\ell)}{\gamma_0 - 1}, \frac{\Omega_{\text{high}} - \Omega(k,\ell)}{\Omega_{\text{high}} - \Omega_{\text{low}}}\right\}. \tag{40}$$

The *a priori* SNR is estimated by [18]

$$\hat{\xi}(k,\ell) = \alpha G_{H_1}^2(k,\ell-1)\gamma(k,\ell-1) + (1-\alpha)\max\{\gamma(k,\ell)-1, 0\} \tag{41}$$

where α is a weighting factor that controls the trade-off between noise reduction and signal distortion, and

$$G_{H_1}(k,\ell) \triangleq \frac{\xi(k,\ell)}{1+\xi(k,\ell)}\exp\left(\frac{1}{2}\int_{\upsilon(k,\ell)}^{\infty}\frac{e^{-t}}{t}dt\right) \tag{42}$$

is the spectral gain function of the *Log-Spectral Amplitude* (LSA) estimator when signal is surely present [24]. An estimate for noise PSD is obtained by recursively averaging past spectral power

values of the noisy measurement, using a time-varying frequency-dependent smoothing parameter. The recursive averaging is given by

$$\hat{\lambda}_d(k,\ell+1) = \tilde{\alpha}_d(k,\ell)\hat{\lambda}_d(k,\ell) + \beta \cdot [1-\tilde{\alpha}_d(k,\ell)]|Y(k,\ell)|^2 \tag{43}$$

where the smoothing parameter $\tilde{\alpha}_d(k,\ell)$ is determined by the signal presence probability $p(k,\ell)$,

$$\tilde{\alpha}_d(k,\ell) \stackrel{\triangle}{=} \alpha_d + (1-\alpha_d)\, p(k,\ell)\,, \tag{44}$$

and β is a factor that compensates the bias when signal is absent. The constant α_d $(0 < \alpha_d < 1)$ represents the minimal smoothing parameter value. The smoothing parameter is close to 1 when signal is present, to prevent an increase in the noise estimate as a result of signal components. It decreases when the probability of signal presence decreases, to allow a fast update of the noise estimate.

The estimate of the clean signal STFT is finally given by

$$\hat{X}(k,\ell) = G(k,\ell)Y(k,\ell)\,, \tag{45}$$

where

$$G(k,\ell) = \{G_{H_1}(k,\ell)\}^{p(k,\ell)} \cdot G_{min}^{1-p(k,\ell)} \tag{46}$$

is the OM-LSA gain function and G_{min} denotes a lower bound constraint for the gain when signal is absent. The implementation of the integrated TF-GSC and multichannel postfiltering algorithm is summarized in Fig. 5. Typical values of the respective parameters, for a sampling rate of 8 kHz, are given in Table I.

V. Experimental Results

In this section, we compare under non-stationary noise conditions the performance of the proposed real-time system to a system consisting of an off-line TF-GSC and a single-channel postfilter. The performance evaluation includes objective quality measures, a subjective study of speech spectrograms and informal listening tests.

A linear array, consisting of four microphones with 5 cm spacing, is mounted in a car on the visor. Clean speech signals are recorded at a sampling rate of 8 kHz in the absence of background

Initialize variables at the first frame for all frequency bins k:

$G_{H_1}(k,0) = \gamma(k,0) = 1; \quad P_{\text{est}}(k,0) = \|\mathbf{U}(k.0)\|^2;$
$\mathcal{S}Y(k,0) = \mathcal{M}Y(k,0) = \hat{\lambda}_d(k,0) = |Y(k,0)|^2;$

Let $n_k = 0$; % n_k is a counter for H_1 frames in frequency bin k.

For $i = 2$ to M

$$\mathcal{S}U_i(k,0) = \mathcal{M}U_i(k,0) = |U_i(k,0)|^2; \quad H_i(k,0) = 0; \quad \tilde{A}_i(k,0) = 1;$$

For all time frames ℓ

For all frequency bins k

Compute the reference noise signals $\mathbf{U}(k,\ell)$ using (3), and the TF-GSC output $Y(k,\ell)$ using (4).

Compute the recursively averaged spectrum of the TF-GSC output and reference signals, $\mathcal{S}Y(k,\ell)$ and $\mathcal{S}U_i(k,\ell)$, using (23), and update the MCRA estimates of the background pseudo-stationary noise, $\mathcal{M}Y(k,\ell)$ and $\mathcal{M}U_i(k,\ell)$, using [22] ($i = 2,\ldots,M$).

Compute the local non-stationarities of the TF-GSC output and reference signals, $\Lambda_Y(k,\ell)$ and $\Lambda_{\mathbf{U}}(k,\ell)$, using (24) and (25).

Using the block diagram for the hypothesis testing (Fig. 3), determine the relevant hypothesis; it possibly requires computation of the transient beam-to-reference ratio $\Omega(k,\ell)$ using (34), and the *a posteriori* SNR at the beamformer output with respect to the pseudo-stationary noise $\gamma_s(k,\ell)$ using (38).

Update the estimate for the power of the reference signals $P_{\text{est}}(k,\ell)$ using (8). In case of absence of transients (H_{0s}), update the multichannel adaptive noise canceller $\mathbf{H}(k,\ell+1)$ using (7).

In case of desired signal presence (H_1), update the estimate $\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}^{(R+1)}(k,\ell+1)$ using (14), and increment n_k by 1.

If $n_k \equiv N$, then: store $\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}^{(r+1)}(k,\ell+1)$ as $\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}^{(r)}(k,\ell+1)$ for $r = 1,\ldots,R$; update the ATF ratios $\tilde{\mathbf{A}}(k,\ell)$ using (13); reset $\hat{\boldsymbol{\phi}}_{\mathbf{Z}Z_1}^{(R+1)}(k,\ell+1)$ and n_k to zero.

In case of H_{0s} or H_{0t}, set the *a priori* signal absence probability $\hat{q}(k,\ell)$ to 1. In case of H_1, set $\hat{q}(k,\ell)$ to 0. In case of H_r, compute $\hat{q}(k,\ell)$ according to (40).

Compute the *a priori* SNR $\hat{\xi}(k,\ell)$ using (41), the conditional gain $G_{H_1}(k,\ell)$ using (42), and the signal presence probability $p(k,\ell)$ using (39).

Compute the time-varying smoothing parameter $\tilde{\alpha}_d(k,\ell)$ using (44), and update the noise spectrum estimate $\hat{\lambda}_d(k,\ell+1)$ using (43).

Compute the OM-LSA estimate of the clean signal, $\hat{X}(k,\ell)$, using (45) and (46).

Fig. 5. The integrated TF-GSC and multichannel postfiltering algorithm.

TABLE I

VALUES OF PARAMETERS USED IN THE IMPLEMENTATION OF THE PROPOSED ALGORITHM, FOR A SAMPLING RATE OF 8 KHZ

$\alpha = 0.92$	$\alpha_d = 0.85$	$\alpha_p = 0.9$	$\alpha_s = 0.9$
$\beta = 1.47$	$\gamma_0 = 4.6$	$\mu = 32.2$	$\mu_h = 0.05$
$\Lambda_0 = 1.67$	$\Lambda_1 = 1.81$	$\Omega_{\text{low}} = 1$	$\Omega_{\text{high}} = 3$
$b = [0.25 \quad 0.5 \quad 0.25]$		$N = 10$	$R = 10$
$G_{min} = -20$ dB			

noise (standing car, silent environment). An interfering speaker is recorded while the car speed is about 60 km/h. The input microphone signals are generated by mixing the speech and noise signals at various SNR levels in the range $[-5, 10]$ dB. Two objective quality measures are used. The first is segmental SNR defined by [25]

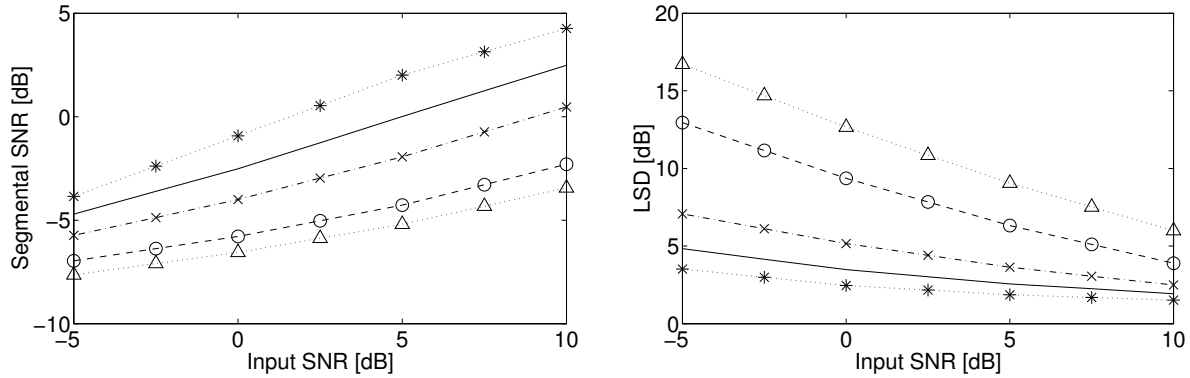
$$\text{SegSNR} = \frac{1}{L}\sum_{\ell=0}^{L-1} 10 \cdot \log \frac{\sum_{n=0}^{K-1} x^2(n+\ell K/2)}{\sum_{n=0}^{K-1}\left[x(n+\ell K/2) - \hat{x}(n+\ell K/2)\right]^2} \quad [\text{dB}] \tag{47}$$

where L represents the number of frames in the signal, and $K = 256$ is the number of samples per frame (corresponding to 32 ms frames, and 50% overlap). The segmental SNR at each frame is limited to perceptually meaningful range between 35 dB and -10 dB [26], [27]. The second quality measure is log-spectral distance (LSD), which is defined by

$$\text{LSD} = \frac{1}{L}\sum_{\ell=0}^{L-1}\left\{\frac{1}{K/2+1}\sum_{k=0}^{K/2}\left[10 \cdot \log \mathcal{C}X(k,\ell) - 10 \cdot \log \mathcal{C}\hat{X}(k,\ell)\right]^2\right\}^{\frac{1}{2}} \quad [\text{dB}] \tag{48}$$

where $\mathcal{C}X(k,\ell) \triangleq \max\left\{|X(k,\ell)|^2, \delta\right\}$ is the spectral power, clipped such that the log-spectral dynamic range is confined to about 50 dB (that is, $\delta = 10^{-50/10} \cdot \max_{k,\ell}\left\{|X(k,\ell)|^2\right\}$).

Fig. 6 shows experimental results obtained for various noise levels. The quality measures are evaluated at the first microphone, the off-line TF-GSC output, the single-channel postfiltering output, and the proposed system output. The single-channel postfiltering output is obtained by applying an OM-LSA estimator to the off-line TF-GSC output. A theoretical limit postfiltering is



(a) (b)

Fig. 6. (a) Average segmental SNR, and (b) average log-spectral distance, at (△) microphone #1, (○) TF-GSC output, (×) single-channel postfiltering output, (solid line) the proposed system output, and (∗) theoretical limit postfiltering output.

achieved by calculating the noise PSD from the noise itself. It can be readily seen that TF-GSC alone does not provide sufficient noise reduction in a car environment, owing to its limited ability to reduce diffuse noise [16]. Furthermore, the proposed real-time system performs considerably better than off-line TF-GSC combined with single-channel postfiltering.

A subjective evaluation of the proposed system was conducted using speech spectrograms and validated by informal listening tests. Typical examples of speech spectrograms are presented in Fig. 7. The TF-GSC output is characterized by a high level of noise. Single-channel postfiltering suppresses pseudo-stationary car noise components, but is inefficient at attenuating transients and interfering speech components. By contrast, the proposed system achieves superior noise attenuation, while preserving the desired speech signal. This is verified by subjective informal listening tests.

VI. Conclusion

We have described an integrated real-time beamforming and postfiltering system, that is particularly advantageous in non-stationary noise environments. The system is based on the TF-GSC beamformer and OM-LSA multichannel postfilter. The TF-GSC beamformer primary output and the reference noise signals are exploited for deciding between speech, stationary noise and transient noise hypotheses. The decisions are used for deriving estimators for the signal presence probability

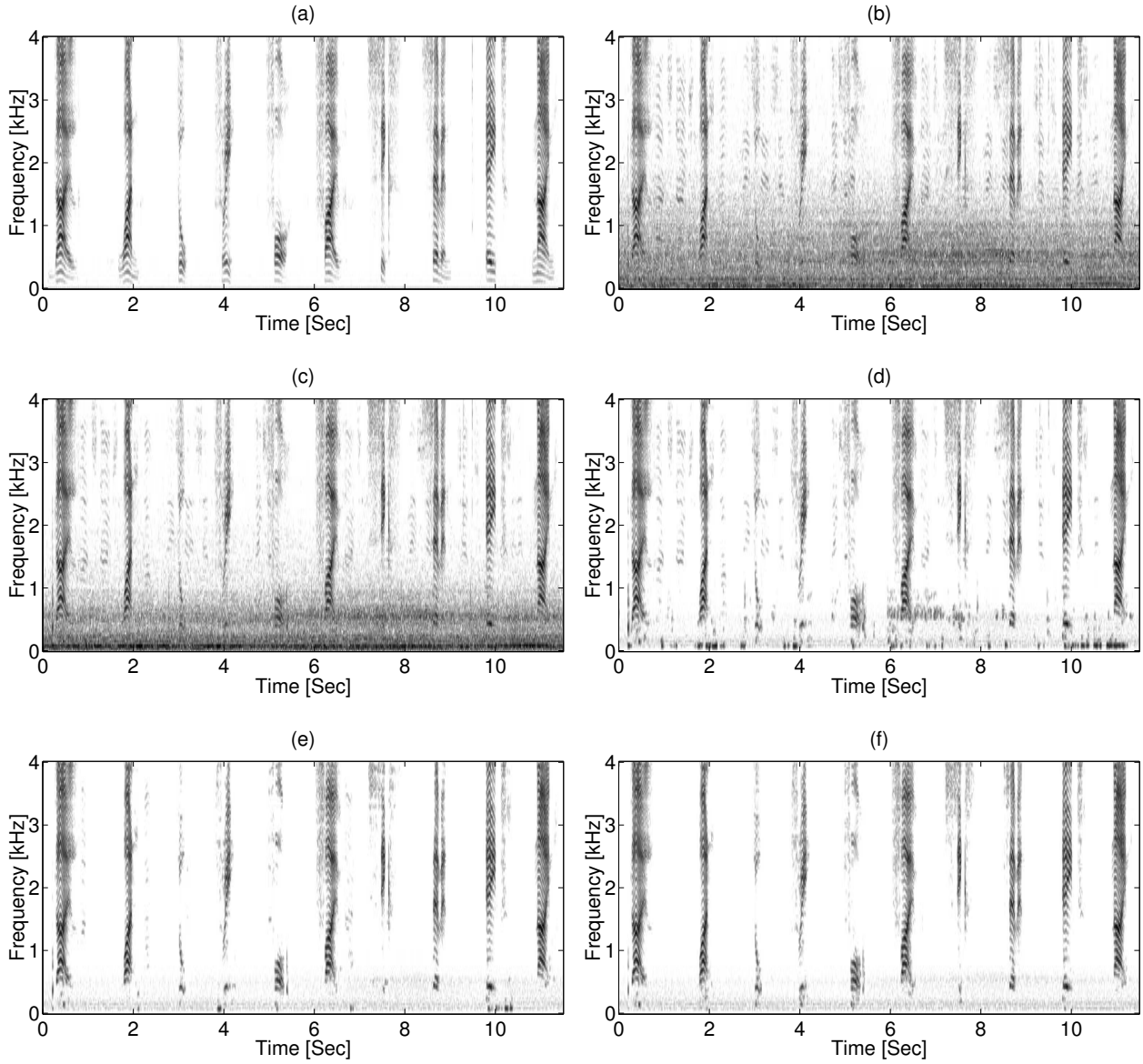


Fig. 7. Speech spectrograms. (a) Original clean speech signal at microphone #1: "Dial one two three four five six seven eight nine."; (b) Noisy signal at microphone #1 (SNR = −4.2 dB, SegSNR = −7.9 dB, LSD = 11.5 dB); (c) TF-GSC output (SegSNR = −7.8 dB, LSD = 10.3 dB); (d) Single-channel postfiltering output (SegSNR = −2.2 dB, LSD = 4.2 dB); (e) Proposed system output (SegSNR = 0.0 dB, LSD = 3.5 dB); (f) Theoretical limit (SegSNR = 0.5 dB, LSD = 3.2 dB).

and for the noise PSD. The signal presence probability modifies the spectral gain function for estimating the clean signal spectral amplitude. The proposed system was tested under non-stationary noise conditions, and its performance was compared to that of a system based on off-line beamforming and single-channel postfiltering. While transient noise components are indistinguishable from desired source components if using a single-channel postfiltering approach, the enhancement of the beamformer output by multichannel postfiltering produces a significantly reduced level of residual transient noise without further distorting the desired signal components.

The novel method has applications in realistic environments, where a desired speech signal is received by several microphones. In typical office environment scenarios, the speech signal is subject to propagation through time-varying acoustical transfer functions (due to talker movements). Stationary noise signals (e.g. noise from an air-conditioning unit), as well as non-stationary interferences (e.g. radio or another talker) are often received by the microphones, contaminating the desired speech. The main contribution of the proposed method is the incorporation of the hypothesis test results back into the beamforming stage, allowing to control the noise canceller branch of the beamformer, as well as the ATF identification.

References

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

[2] K. U. Simmer, J. Bitzer, and C. Marro, *Post-Filtering Techniques*, chapter 3, pp. 39–60, In Brandstein and Ward [1], 2001.

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

[4] R. Zelinski, "A microphone array with adaptive post-filtering for noise reduction in reverberant rooms," in *Proc. 13th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-88*, New York, USA, 11–14 April 1988, pp. 2578–2581.

[5] R. Zelinski, "Noise reduction based on microphone array with LMS adaptive post-filtering," November 1990, vol. 26, pp. 2036–2581.

[6] S. Fischer and K. U. Simmer, "An adaptive microphone array for hands-free communication," in *Proc. 4th International Workshop on Acoustic Echo and Noise Control, IWAENC-95*, Røros, Norway, 21–23 June 1995, pp. 44–47.

[7] S. Fischer and K. U. Simmer, "Beamforming microphone arrays for speech acquisition in noisy environments," *Speech Communication*, vol. 20, no. 3–4, pp. 215–227, December 1996.

[8] S. Fischer and K.-D. Kammeyer, "Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environments," in *Proc. 22th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-97*, Munich, Germany, 20–24 April 1997, pp. 359–362.

[9] J. Meyer and K. U. Simmer, "Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction," in *Proc. 22th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-97*, Munich, Germany, 20–24 April 1997, pp. 21–24.

[10] K. U. Simmer, S. Fischer, and A. Wasiljeff, "Suppression of coherent and incoherent noise using a microphone array," *Annales des Télécommunications*, vol. 49, no. 7–8, pp. 439–446, July 1994.

[11] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multi-microphone noise reduction by post-filter and superdirective beamformer," in *Proc. 6th International Workshop on Acoustic Echo and Noise Control, IWAENC-99*, Pocono Manor, Pennsylvania, 27–30 September 1999, pp. 100–103.

[12] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multi-microphone noise reduction techniques as front-end devices for speech recognition," *Speech Communication*, vol. 34, no. 1, pp. 3–12, April 2001.

[13] I. Cohen and B. Berdugo, "Microphone array post-filtering for non-stationary noise suppression," in *Proc. 27th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-2002*, Orlando, Florida, 13-17 May 2002, pp. 901–904.

[14] I. Cohen, "Multi-channel post-filtering in non-stationary noise environments," Technical Report, CCIT Report 376, EE Pub. 1314, Technion - Israel Institute of Technology, Haifa, Israel, April 2002.

[15] S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," Technical Report, CCIT Report 380, EE Pub. 1318, Technion - Israel Institute of Technology, Haifa, Israel, May 2002.

[16] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, August 2001.

[17] D. Burshtein and S. Gannot, "Speech Enhancement Using a Mixture-Maximum Model," to appear in IEEE Trans. Acoustics, Speech and Signal Processing, Sep. 2002.

[18] I. Cohen and B. Berdugo, "Speech enhancement for non-stationary noise environments," *Signal Processing*, vol. 81, no. 11, pp. 2403–2418, October 2001.

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

[20] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1985.

[21] S. Nordholm, I. Claesson, and P. Eriksson, "The broadband Wiener solution for Griffiths-Jim beamformers," *IEEE Trans. Signal Processing*, vol. 40, no. 9, pp. 474–478, February 1992.

[22] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," Technical Report, EE PUB 1291, Technion - Israel Institute of Technology, Haifa, Israel, October 2001.

[23] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-32, no. 6, pp. 1109–1121, December 1984.

[24] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. ASSP-33, no. 2, pp. 443–445, April 1985.

[25] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1988.

[26] J. R. Deller, J. H. L. Hansen, and J. G. Proakis, *Discrete-Time Processing of Speech Signals*, IEEE Press, New York, 2nd edition, 2000.

[27] P. E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.