CCIT Report #394
July 2002

Analysis of Two-Channel Generalized Sidelobe Canceller (GSC) with Post-Filtering

Israel Cohen

Department of Electrical Engineering, Technion — Israel Institute of Technology, Technion City, Haifa 32000, Israel
E-mail: icohen@ee.technion.ac.il; Tel.: +972 4 8294731; Fax: +972 4 8323041.

Abstract

In this paper, we analyze a two-channel generalized sidelobe canceller with post-filtering in non-stationary noise environments. The post-filtering includes detection of transients at the beamformer output and reference signal, a comparison of their transient power, estimation of the signal presence probability, estimation of the noise spectrum, and spectral enhancement for minimizing the mean-square error of the log-spectra. Transients are detected based on a measure of their local non-stationarity, and classified as desired or interfering based on the transient beam-to-reference ratio. We introduce a *transient discrimination quality* measure, which quantifies the beamformer's capability to recognize noise transients as distinct from signal transients. Evaluating this measure in various noise fields shows that desired and interfering transients can generally be differentiated within a wide range of frequencies. To further improve the transient noise reduction at low and high frequencies in case the signal is wideband, we estimate for each time frame a *global* likelihood of signal presence. The global likelihood is associated with the transient beam-to-reference ratios in frequencies, where the transient discrimination quality is high. Experimental results demonstrate the usefulness of the proposed approach in various car environments.

Keywords

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

I. Introduction

In reverberant and noisy environments, multi-channel systems are designed for spatially filtering interfering signals coming from undesired directions [1]. In case of incoherent or diffuse noise fields, beamforming alone does not provide sufficient noise reduction, and post-filtering is normally required [2], [3]. Post-filtering based on Wiener filtering and the auto and cross spectral densities of the sensor signals was introduced by Zelinski [4], [5]. The noise power density is overestimated, and therefore a modified version was proposed by Simmer and Wasiljeff [6], which employs the power spectral density of the beamformer output rather than the average of the power spectral densities of individual sensor signals. The underlying assumption is that noise components at different sensors are mutually uncorrelated.

To take into account the presence of correlated noise components, Fischer *et al.* [7], [8], [9] suggested a noise reduction system, which is based on the generalized sidelobe canceller (GSC). The GSC suppresses the coherent noise components, while a Wiener filter in the look direction is designed to suppress the incoherent noise components. Bitzer *et al.* showed that in a diffuse noise field, neither the GSC nor adaptive post-filtering performs well at low frequencies [10], [11]. Therefore, at the output of a GSC with standard Wiener post-filtering they used a second post-filter to reduce the spatially correlated noise components [12], [13]. Meyer and Simmer [14] combined Wiener filtering in the high-frequency band with spectral subtraction in the low-frequency band. The Wiener filtering is applied for the suppression of low-coherence noise components, while the spectral subtraction is used for high-coherence noise reduction.

A noise reduction system that is nearly independent of the correlation properties of the noise field was suggested by Fischer and Kameyer [15]. Wiener filtering is applied to the output of a broadband beamformer, that is built up by several harmonically nested subarrays. This structure has been further analyzed by Marro *et al.* [2]. McCowan *et al.* used a near-field super-directive beamforming and investigated the effect of a Wiener post-filter on speech recognition performance [16]. They showed that in the case of nearfield sources and diffuse noise conditions, improved recognition performance can be achieved compared to conventional adaptive beamformers. A theoretical analysis of Wiener multi-channel post-filtering is presented in [3]. Gannot *et al.* [17] addressed the

problem of general transfer functions that relate the source signal to the sensors. They adapted the GSC solution to the general transfer function case, and proposed an algorithm for enhancing an arbitrary non-stationary signal corrupted by stationary noise. To improve the noise reduction performance in a diffuse noise field and at low frequencies, they applied single-channel post-filtering to the beamformer output. However, a single-channel post-filtering approach lacks the ability to attenuate highly non-stationary noise components, since such components are not differentiated from the desired signal components.

Recently, we introduced a multi-channel post-filtering approach for minimizing the log-spectral amplitude distortion in non-stationary noise environments [18], [19]. The ratio between the transient power at the beamformer output and the transient power at the reference noise signals was used for indicating whether such a transient is desired or interfering. We showed that compared to single-channel post-filtering, a significantly reduced level of non-stationary noise can be achieved without further distorting the desired signal components.

In this paper, we analyze a two-channel GSC with post-filtering in non-stationary noise environments. We quantify the beamformer's capability to recognize interfering transients as distinct from source transients by using a *transient discrimination quality* measure. This measure, evaluated in various noise fields, shows that desired and interfering transients can generally be differentiated within a wide range of frequencies. In case the transient or pseudo-stationary noise field is coherent, the direction to the interfering source has to be different from the direction to the desired source by at least twice the uncertainty in the angle of arrival. For low frequencies, the directivity of the beamformer and its spatial filtering capability are lost. For high frequencies, spatial aliasing folds interferences coming from the side to the main lobe. In these cases, the two-channel post-filtering reduces to single-channel post-filtering, since the transient power ratio between the beamformer output and the reference signal is no longer a distinctive characteristic of the transient source.

To further improve the transient noise reduction at low and high frequencies in case the desired signal is wideband (*e.g.*, speech signal), we introduce a *global* likelihood of signal presence. The global likelihood is related to the number of frequency bins that likely contain desired components within a certain range of frequencies and at a given time frame. When the global likelihood is lower than a certain threshold, we conclude that desired components are absent from that frame and set

the *a priori* signal absence probability to one for all frequency bins. This uniformly suppresses the noise in a manner which is more pleasant to a human listener, and better eliminates narrow-band interfering transients, particularly those arriving from the look direction. Experimental results in various car environments confirm that two-channel post-filtering is superior to single-channel post-filtering. The improvement in performance using the proposed post-filtering approach is substantial when the noise spectrum fluctuates.

The paper is organized as follows. In Section II, we review the two-channel generalized sidelobe canceller, and derive relations in the power-spectral domain between the beamformer output, the reference noise signals, the desired source signal, and the input transient interferences. In Section III, we address the problem of estimating the time-varying spectrum of the beamformer output noise, and present the post-filtering approach. Desired source components are detected at the beamformer output and discriminated from transient noise components based on the transient power ratio between the beamformer output and the reference signal. In Section IV, we evaluate in various noise fields the beamformer's discrimination capability to recognize interfering transients as distinct from the source transients. Finally, in Section V, we compare the proposed method to single-channel post-filtering, and present experimental results in various car environments.

II. Two-Channel Generalized Sidelobe Cancelling

Let $x(t)$ denote a desired source signal, and let signal vectors $\mathbf{d}_s(t)$ and $\mathbf{d}_t(t)$ denote uncorrelated interfering signals at the output of two sensors. The vector $\mathbf{d}_s(t)$ represents pseudo-stationary interferences, and $\mathbf{d}_t(t)$ represents undesired transient components. Assuming that the array is presteered to the direction of the source signal, the observed signals are given by

$$z_i(t) = x(t) + d_{is}(t) + d_{it}(t)\,, \quad i = 1, 2 \tag{1}$$

where $d_{is}(t)$ and $d_{it}(t)$ are the interference signals corresponding to the i-th sensor. The observed signals are divided in time into overlapping frames by the application of a window function and analyzed using the short-time Fourier transform (STFT). In the time-frequency domain we have

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

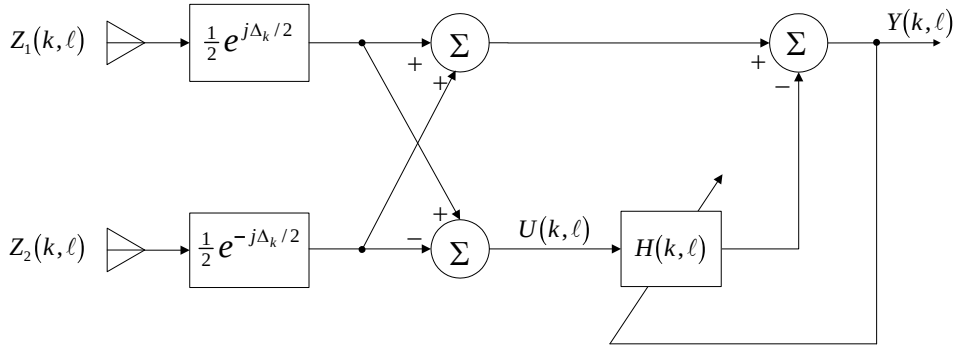


Fig. 1. Two-channel Generalized Sidelobe Canceller.

where $\mathbf{A} \triangleq [\,1 \quad 1\,]^T$, k represents the frequency bin index, ℓ the frame index, and

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

Fig. 1 shows a two-channel generalized sidelobe canceller structure for a linearly constrained adaptive beamformer [20], [21]. The beamformer comprises a fixed beamformer (delay & sum), a blocking channel (delay & subtract) which yields the reference noise signal $U(k,\ell)$, and an adaptive noise canceller $H(k,\ell)$ which eliminates the stationary noise that leaks through the sidelobes of the fixed beamformer. We assume that the noise canceller is adapted only to the stationary noise, and not modified during transient interferences. Furthermore, we assume that some desired signal components may pass through the blocking channel due to steering error.

The uncertainty in the angle of arrival of the signal of interest is represented by

$$\Delta_k = \frac{\omega_k\, l}{c}\sin(\varphi) + \phi \tag{3}$$

where $\omega_k = 2\pi\, f_s\,(k-1)/N$ is the center of the kth frequency bin ($k \in [0, N/2+1]$), N the length of the spectral analysis window, f_s the sampling frequency, l is the distance between the sensors, $c = 340$ m/s the speed of sound, φ the mismatch in the source direction, and ϕ the estimation error in the difference of phase. We let $\mathbf{W}(k) = \frac{1}{2}\left[\; e^{j\Delta_k/2} \quad e^{-j\Delta_k/2} \;\right]^H$ be the weighting vector of the

fixed beamformer, and $\mathbf{B}(k) = \frac{1}{2}\left[\begin{array}{cc} e^{j\Delta_k/2} & -e^{-j\Delta_k/2} \end{array}\right]^H$ the blocking vector. The beamformer output and reference noise signal are thus given by

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

$$U(k,\ell) = \mathbf{B}^H(k)\mathbf{Z}(k,\ell)\,. \tag{5}$$

The optimal solution for the filter $H(k,\ell)$ is obtained by minimizing the output power of the stationary noise [22]. 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:

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

The Wiener-Hopf solution is given by [23]

$$H(k,\ell) = \left[\mathbf{B}^H(k)\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell)\mathbf{B}(k)\right]^{-1}\mathbf{B}^H(k)\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell)\mathbf{W}(k)\,. \tag{7}$$

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 functions, $\Gamma_s(k,\ell)$ and $\Gamma_t(k,\ell)$, by

$$\mathbf{\Phi}_{\mathbf{D}_s\mathbf{D}_s}(k,\ell) = \lambda_s(k,\ell)\left[\begin{array}{cc} 1 & \Gamma_s(k,\ell) \\ \Gamma_s^*(k,\ell) & 1 \end{array}\right] \tag{8}$$

$$\mathbf{\Phi}_{\mathbf{D}_t\mathbf{D}_t}(k,\ell) = \lambda_t(k,\ell)\left[\begin{array}{cc} 1 & \Gamma_t(k,\ell) \\ \Gamma_t^*(k,\ell) & 1 \end{array}\right] \tag{9}$$

where $\lambda_s(k,\ell)$ and $\lambda_t(k,\ell)$ represent the input noise power at a single sensor. In this case, the optimal noise canceller (Eq. (7)) reduces to

$$H(k,\ell) = \frac{j\Im\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}}{1 - \Re\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}}\,. \tag{10}$$

The source signal, the stationary noise and transient noise are assumed to be uncorrelated. Therefore, the input PSD-matrix is given by

$$\mathbf{\Phi_{ZZ}}(k,\ell) = \lambda_x(k,\ell)\mathbf{A}\,\mathbf{A}^T + \mathbf{\Phi_{D_sD_s}}(k,\ell) + \mathbf{\Phi_{D_tD_t}}(k,\ell) \tag{11}$$

where $\lambda_x(k,\ell) \stackrel{\triangle}{=} E\left\{|X(k,\ell)|^2\right\}$ is the PSD of the desired source signal. Using (4) and (5), the PSD's of the beamformer output and the reference signal are obtained by

$$\phi_{YY}(k,\ell) = [\mathbf{W}(k) - \mathbf{B}(k)H(k,\ell)]^H \mathbf{\Phi_{ZZ}}(k,\ell)\,[\mathbf{W}(k) - \mathbf{B}(k)H(k,\ell)] \tag{12}$$

$$\phi_{UU}(k,\ell) = \mathbf{B}^H(k)\mathbf{\Phi_{ZZ}}(k,\ell)\mathbf{B}(k)\,. \tag{13}$$

Substituting Eqs. (8)–(11) into (13) and (12), we have the following linear relations between the PSD's of the beamformer output, the reference signal, the desired source signal, and the input interferences:

$$\phi_{YY}(k,\ell) = C_{11}(k,\ell)\lambda_x(k,\ell) + C_{12}(k,\ell)\lambda_s(k,\ell) + C_{13}(k,\ell)\lambda_t(k,\ell) \tag{14}$$

$$\phi_{UU}(k,\ell) = C_{21}(k)\lambda_x(k,\ell) + C_{22}(k,\ell)\lambda_s(k,\ell) + C_{23}(k,\ell)\lambda_t(k,\ell) \tag{15}$$

where

$$C_{11}(k,\ell) = \left[\cos\left(\tfrac{\Delta_k}{2}\right) - \frac{\Im\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}}{1-\Re\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}}\sin\left(\tfrac{\Delta_k}{2}\right)\right]^2 \tag{16}$$

$$C_{12}(k,\ell) = \frac{1-|\Gamma_s(k,\ell)|^2}{1-\Re\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}} \tag{17}$$

$$C_{13}(k,\ell) = \frac{1}{2}\left[|1+H(k,\ell)|^2 + \Re\left\{e^{j\Delta_k}\Gamma_t(k,\ell)\,[1+H(k,\ell)]^2\right\}\right] \tag{18}$$

$$C_{21}(k) = \sin^2\left(\tfrac{\Delta_k}{2}\right) \tag{19}$$

$$C_{22}(k,\ell) = \frac{1}{2}\left[1-\Re\left\{e^{j\Delta_k}\Gamma_s(k,\ell)\right\}\right] \tag{20}$$

$$C_{23}(k,\ell) = \frac{1}{2}\left[1-\Re\left\{e^{j\Delta_k}\Gamma_t(k,\ell)\right\}\right]\,. \tag{21}$$

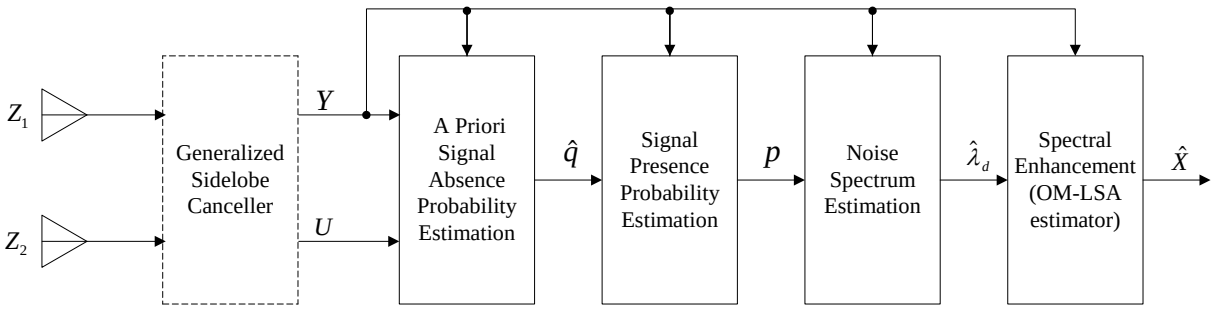


Fig. 2. Block diagram of the post-filtering.

III. Two-Channel Post-Filtering

In this section, we address the problem of estimating the time-varying spectrum of the beamformer output noise, and present the post-filtering approach. Fig. 2 describes the block diagram of the proposed two-channel post-filtering. Desired source components are detected at the beamformer output, and an estimate $\hat{q}(k,\ell)$ for the *a priori* signal absence probability is produced. Based on a Gaussian statistical model [24], and a decision-directed estimator for the *a priori* SNR under signal presence uncertainty [25], we derive an estimator $p(k,\ell)$ for the signal presence probability. This estimator controls the components that are introduced as noise into the PSD estimator. Finally, spectral enhancement of the beamformer output is achieved by applying an *optimally-modified log-spectral amplitude* (OM-LSA) gain function [25]. This gain minimizes the mean-square error of the log-spectra under signal presence uncertainty.

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

$$\mathcal{S}Y(k,\ell) = \alpha_s \cdot \mathcal{S}Y(k,\ell-1) + (1-\alpha_s)\sum_{i=-w}^{w} b_i |Y(k-i,\ell)|^2 \tag{22}$$

where α_s $(0 \le \alpha_s \le 1)$ 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 pseudo-stationary noise, derived using the *Minima Controlled Recursive Averaging* (MCRA) approach [25], [26]. The ratios

$$\Lambda_Y(k,\ell) \triangleq \mathcal{S}Y(k,\ell)/\,\mathcal{M}Y(k,\ell) \tag{23}$$

$$\Lambda_U(k,\ell) \triangleq \mathcal{S}U(k,\ell)/\,\mathcal{M}U(k,\ell) \tag{24}$$

represent the local non-stationarities (LNS) of the beamformer output and reference signal, respectively [19]. The LNS fluctuates about one in the absence of transients, and expected to be well above one in the neighborhood of time-frequency bins that contain transients. The post-filtering includes detection of transients at the beamformer output and reference signal, and a comparison of their transient power. In case we detect transients at the beamformer output but no simultaneous transients at the reference signals, we determine that these transients are likely source components which require a cautious enhancement. On the other hand, simultaneous transients at the beamformer output and the reference signal are handled according to their power ratio. A stronger transient at the beamformer output indicates presence of desired components, and therefore should be preserved. Whereas a stronger transient at the reference signal implies an interfering source, and therefore needs to be suppressed.

A. Detection of transients at the beamformer output

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. Let Λ_0 denote a threshold value of the LNS for the detection of transients at the beamformer output (*i.e.*, decide $H_1 \cup H_{0t}$ if $\Lambda_Y(k,\ell) > \Lambda_0$, and decide H_{0s} otherwise). The false alarm and detection probabilities are defined by

$$P_{f,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \mid H_{0s}\right) \tag{25}$$

$$P_{d,Y}(k,\ell) = \mathcal{P}\left(\Lambda_Y(k,\ell) > \Lambda_0 \mid H_1 \cup H_{0t}\right) . \tag{26}$$

Then for a specified $P_{f,Y}$, the required threshold value and the detection probability are given by [19]

$$\Lambda_0 = \frac{1}{\mu} F^{-1}_{\chi^2;\mu}\left(1 - P_{f,Y}\right) \tag{27}$$

$$P_{d,Y}(k,\ell) = 1 - F_{\chi^2;\mu}\left[\frac{1}{1+\xi_Y(k,\ell)} F^{-1}_{\chi^2;\mu}\left(1 - P_{f,Y}\right)\right] \tag{28}$$

where

$$\xi_Y(k,\ell) \triangleq \frac{C_{11}(k,\ell)\lambda_x(k,\ell) + C_{13}(k,\ell)\lambda_t(k,\ell)}{C_{12}(k,\ell)\lambda_s(k,\ell)} \tag{29}$$

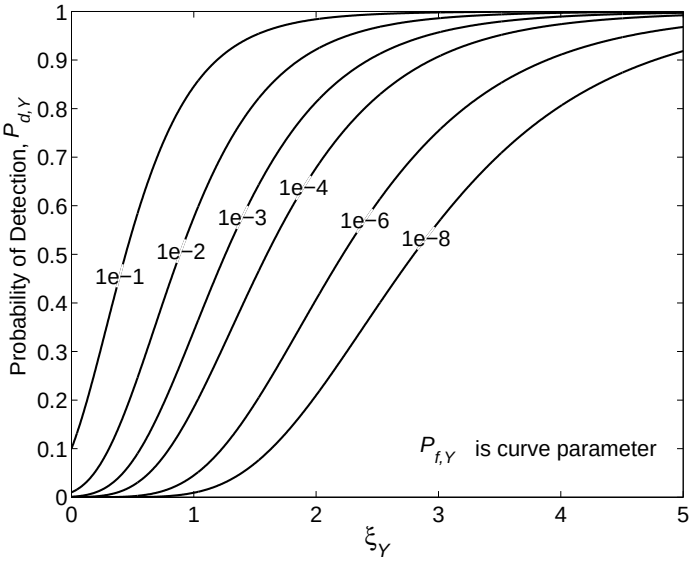


Fig. 3. Receiver operating characteristic curve for the detection of transients at the beamformer output or at the reference noise signal ($\mu = 22.1$).

represents the ratio between the transient and pseudo-stationary power at the beamformer output, and $F_{\chi^2;\mu}(x)$ denotes the standard chi-square distribution function with μ degrees of freedom[1]. Fig. 3 shows the receiver operating characteristic (ROC) curve for detection of transients at the beamformer output, with the false alarm probability as parameter, and μ set to 22.1 (this value of μ was obtained for a smoothing $\mathcal{S}$ of the form (22), with $\alpha_s = 0.8$, and a normalized Hanning window $b = \frac{1}{12}[1 \quad 3 \quad 4 \quad 3 \quad 1]$). Suppose that we require a false alarm probability no larger than $P_{f,Y} = 0.05$, and suppose that transients at the beamformer output are defined by $\xi_Y(k,\ell) \geq 2$. Then, the detection probability obtained using a detector $\Lambda_Y(k,\ell) > \Lambda_0 = 1.54$ is $P_{d,Y}(k,\ell) = 0.97$.

[1]The equivalent degrees of freedom, μ, is determined by the smoothing parameter α_s, the window function b, and the spectral analysis parameters of the STFT (size and shape of the analysis window, and frame-update step). The value of μ is estimated by generating a stationary white Gaussian noise $d(t)$, transforming it to the time-frequency domain, and substituting the sample mean and variance (over the entire time-frequency plane) into the expression $\hat{\mu} \approx 2\,E^2\,\{\mathcal{S}D(k,\ell)\}\,/\mathrm{var}\,\{\mathcal{S}D(k,\ell)\}$.

B. Discrimination between source and interfering transients

Transient signal components are relatively strong at the beamformer output, whereas transient *noise* components are relatively strong at the reference signal. Hence, we expect the transient power ratio between the beamformer output and the reference signal to be large for desired transients, and small for noise components. Let

$$\Omega(k,\ell) = \frac{\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)}{\mathcal{S}U(k,\ell) - \mathcal{M}U(k,\ell)} \tag{30}$$

represent the *transient beam-to-reference ratio* (TBRR), *i.e.*, the ratio between the transient power of the beamformer output and the transient power of the reference signal. Then, given that H_1 or H_{0t} is true,

$$\begin{aligned} \Omega(k,\ell)|_{H_1 \cup H_{0t}} &\approx \frac{\phi_{YY}(k,\ell) - C_{12}(k,\ell)\lambda_s(k,\ell)}{\phi_{UU}(k,\ell) - C_{22}(k,\ell)\lambda_s(k,\ell)} \\ &= \frac{C_{11}(k,\ell)\lambda_x(k,\ell) + C_{13}(k,\ell)\lambda_t(k,\ell)}{C_{21}(k)\lambda_x(k,\ell) + C_{23}(k,\ell)\lambda_t(k,\ell)}\,. \end{aligned} \tag{31}$$

Assuming that H_1 and H_{0t} are exclusive, *i.e.*, assuming that desired and interfering transients do not overlap in the time-frequency domain, and supposing that there exist thresholds $\Omega_{\text{high}}(k)$ and $\Omega_{\text{low}}(k)$ such that

$$\Omega(k,\ell)|_{H_{0t}} \approx \frac{C_{13}(k,\ell)}{C_{23}(k,\ell)} \leq \Omega_{\text{low}}(k) \leq \Omega_{\text{high}}(k) \leq \frac{C_{11}(k,\ell)}{C_{21}(k)} \approx \Omega(k,\ell)|_{H_1} \tag{32}$$

for all ℓ, we can determine that signal is likely present at the kth frequency bin and ℓth frame if $\Omega(k,\ell) \geq \Omega_{\text{high}}(k)$. On the other hand, if $\Omega(k,\ell) \leq \Omega_{\text{low}}(k)$ then we can determine that the detected transient is interfering. To accommodate the uncertainty in the TBRR and to improve the discrimination between source and interfering transients, we define a function $\psi(k,\ell)$ that represents the likelihood of signal presence. The value of $\psi(k,\ell)$ is set to zero if no transients are detected at the beamformer output ($\Lambda_Y(k,\ell) \leq \Lambda_0$). In case a transient is detected at the beamformer output but not at the reference signal ($\Lambda_U(k,\ell) \leq \Lambda_0 < \Lambda_Y(k,\ell)$), $\psi(k,\ell)$ is set to one. In case a transient is detected simultaneously at the beamformer output and at the reference

signal $(\Lambda_U(k,\ell), \Lambda_Y(k,\ell) > \Lambda_0)$, $\psi(k,\ell)$ is proportional to $\Omega(k,\ell)$ according to

$$\psi(k,\ell) = \begin{cases} 0\,, & \text{if } \Omega(k,\ell) \leq \Omega_{\text{low}}(k) \\ \frac{\Omega(k,\ell)-\Omega_{\text{low}}(k)}{\Omega_{\text{high}}(k)-\Omega_{\text{low}}(k)}\,, & \text{if } \Omega_{\text{low}}(k) < \Omega(k,\ell) \leq \Omega_{\text{high}}(k) \\ 1\,, & \text{otherwise.} \end{cases} \tag{33}$$

For a given frame, the global likelihood of signal presence is related to the number of frequency bins that likely contain desired components within a certain range of frequencies. Therefore we define

$$\tilde{\psi}(\ell) = \frac{1}{k_1 - k_0 + 1} \sum_{k=k_0}^{k_1} \psi(k,\ell) \tag{34}$$

where k_0 and k_1 are the lower and upper frequency bin indices representing the frequency range.

Fig. 4 summarizes a block diagram for the estimation of the *a priori* signal presence probability. The detection of desired source components at the beamformer output is carried out in the time-frequency plane for each frame and frequency bin. First we compute the local likelihood of signal presence for all frequency bins. Then, a global likelihood $\tilde{\psi}(\ell)$ is generated, and compared to a certain threshold ψ_0. In case the global likelihood is too low, we conclude that signal is absent from that frame and set the *a priori* signal absence probability $\hat{q}(k,\ell)$ to one for all frequency bins. This prevents from narrow-band interfering transients, particularly those arriving from the look direction, to be confused with desired components. This also helps to reduce musical noise phenomena. In case the global likelihood is above the threshold ψ_0, the *a priori* signal absence probability is related to the likelihood of signal absence at the ℓth frame and kth frequency bin $(1-\psi(k,\ell))$ and to the *a posteriori* SNR at the beamformer output with respect to the pseudo-stationary noise $\gamma_s(k,\ell) \triangleq |Y(k,\ell)|^2/\mathcal{M}Y(k,\ell)$. Specifically, we determine the *a priori* signal absence probability according to

$$\hat{q}(k,\ell) = \begin{cases} 1, & \text{if } \gamma_s(k,\ell) \leq 1 \text{ or } \tilde{\psi}(\ell) \leq \psi_0 \\ \max\left\{\frac{\gamma_0-\gamma_s(k,\ell)}{\gamma_0-1}\,, 1-\psi(k,\ell)\right\}, & \text{otherwise,} \end{cases} \tag{35}$$

where γ_0 denotes a constant satisfying $\mathcal{P}\left(\gamma_s(k,\ell) \geq \gamma_0 \mid H_{0s}\right) < \epsilon$ for a certain significance level ϵ. Since the distribution of $\gamma_s(k,\ell)$ in the absence of transients is exponential [26], the constant γ_0 is related to significance level by $\gamma_0 = -\log(\epsilon)$ (typically we use $\epsilon = 0.01$ and $\gamma_0 = 4.6$).

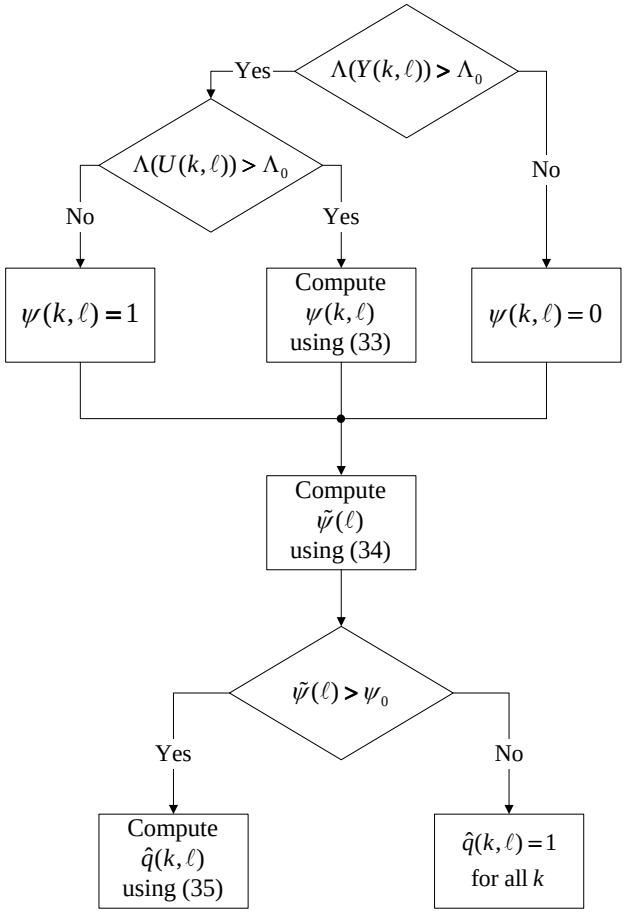


Fig. 4. Block diagram for the *a priori* signal absence probability estimation.

C. Noise estimation and spectral enhancement

Under the assumed statistical model, the signal presence probability is given by

$$p(k,\ell) = \left\{1 + \frac{q(k,\ell)}{1-q(k,\ell)}(1+\xi(k,\ell))\exp(-\upsilon(k,\ell))\right\}^{-1} \tag{36}$$

where $\xi(k,\ell) \triangleq E\left\{|X(k,\ell)|^2\right\}/\lambda_d(k,\ell)$ 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* SNR is estimated by [25]

$$\hat{\xi}(k,\ell) = \alpha G_{H_1}^2(k,\ell-1)\gamma(k,\ell-1) + (1-\alpha)\max\left\{\gamma(k,\ell)-1, 0\right\} \tag{37}$$

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

$$G_{H_1}(k,\ell) \triangleq \frac{\xi(k,\ell)}{1+\xi(k,\ell)} \exp\left(\frac{1}{2}\int_{\upsilon(k,\ell)}^{\infty} \frac{e^{-t}}{t} dt\right) \tag{38}$$

is the spectral gain function of the *Log-Spectral Amplitude* (LSA) estimator when signal is surely present [27]. The MCRA approach for noise spectrum estimation [26] is to recursively average past spectral power values of the noisy measurement, using a smoothing parameter that is controlled by the minima values of a smoothed periodogram. The recursive averaging is given by

$$\hat{\lambda}_d(k,\ell+1) = \tilde{\alpha}_d(k,\ell)\hat{\lambda}_d(k,\ell) + \beta \cdot [1-\tilde{\alpha}_d(k,\ell)]|Y(k,\ell)|^2 \tag{39}$$

where $\tilde{\alpha}_d(k,\ell)$ is a time-varying frequency-dependent smoothing parameter, and β is a factor that compensates the bias when signal is absent. The smoothing parameter is determined by the signal presence probability, $p(k,\ell)$, and a constant α_d $(0 < \alpha_d < 1)$ that represents its minimal value:

$$\tilde{\alpha}_d(k,\ell) \triangleq \alpha_d + (1-\alpha_d)\, p(k,\ell)\,. \tag{40}$$

When signal is present, $\tilde{\alpha}_d$ is close to one, thus preventing the noise estimate from increasing as a result of signal components. As the probability of signal presence decreases, the smoothing parameter gets smaller, facilitating a faster update of the noise estimate.

The estimate for the clean signal STFT is finally given by

$$\hat{X}(k,\ell) = G(k,\ell)Y(k,\ell)\,, \tag{41}$$

where

$$G(k,\ell) = \{G_{H_1}(k,\ell)\}^{p(k,\ell)} \cdot G_{\min}^{1-p(k,\ell)} \tag{42}$$

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 multi-channel post-filtering algorithm is summarized in Fig. 5. Typical values of the respective parameters, for a sampling rate of 8 kHz, are given in Table I. The values of the lower and upper frequency bin indices, $k_0 = 9$ and $k_1 = 113$, which are used in Eq. (34) for the computation of the global likelihood of signal presence, correspond to a frequency range of $[250, 3500]$ Hz.

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

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

For all time frames ℓ

For all frequency bins k

Compute the recursively averaged spectrum of the beamformer output and reference signal, $\mathcal{S}Y(k,\ell)$ and $\mathcal{S}U(k,\ell)$, using Eq. (22), and update the MCRA estimates of the pseudo-stationary noise, $\mathcal{M}Y(k,\ell)$ and $\mathcal{M}U(k,\ell)$, using [26].

Compute the local non-stationarities of the beamformer output and reference signal, $\Lambda_Y(k,\ell)$ and $\Lambda_U(k,\ell)$, using Eqs. (23) and (24), and compute the transient beam-to-reference ratio, $\Omega(k,\ell)$, using Eq. (30).

Using the block diagram in Fig. 4, determine the *a priori* signal absence probability $\hat{q}(k,\ell)$.

Compute the *a priori* SNR $\hat{\xi}(k,\ell)$ using Eq. (37), the conditional gain $G_{H_1}(k,\ell)$ using Eq. (38), and the signal presence probability $p(k,\ell)$ using Eq. (36).

Compute the time-varying smoothing parameter $\tilde{\alpha}_d(k,\ell)$ using Eq. (40), and update the noise spectrum estimate $\hat{\lambda}_d(k,\ell+1)$ using Eq. (39).

Compute the OM-LSA estimate of the clean signal, $\hat{X}(k,\ell)$, using Eqs. (41) and (42).

Fig. 5. The two-channel post-filtering algorithm.

TABLE I

VALUES OF PARAMETERS USED IN THE IMPLEMENTATION OF THE PROPOSED TWO-CHANNEL POST-FILTERING, FOR A SAMPLING RATE OF 8 KHZ

$\Lambda_0 = 1.54$	$\Omega_{\text{low}} = 1$	$\Omega_{\text{high}} = 3$	$\gamma_0 = 4.6$
$\alpha = 0.92$	$\alpha_s = 0.8$	$\alpha_d = 0.85$	$\beta = 1.98$
$k_0 = 9$	$k_1 = 113$	$\psi_0 = 0.25$	$\mu = 22.1$
$b = \frac{1}{12}[1 \quad 3 \quad 4 \quad 3 \quad 1]$		$N = 256$	$G_{min} = -20$ dB

IV. THEORETICAL ANALYSIS

In this section we assume that the spatial coherence functions of the pseudo-stationary and transient noise, $\Gamma_s(k,\ell)$ and $\Gamma_t(k,\ell)$, are independent of the frame index ℓ. We define a *transient discrimination quality*, which indicates a beamformer's capability to recognize interfering transients as distinct from source transients, and evaluate this quality in various noise fields.

According to the inequalities in (32), the discrimination quality between desired and interfering

transients is high whenever the range of the TBRR values given that H_1 is true ($\Omega(k,\ell)|_{H_1}$) is readily distinguishable from the range given that H_{0t} is true ($\Omega(k,\ell)|_{H_{0t}}$). Otherwise, the TBRR alone is insufficient for determining the origin of transients that are simultaneously detected at the beamformer output and at the reference signal. Let the *transient discrimination quality* of a beamformer at the kth frequency bin be defined by

$$Q(k) = \frac{C_{11}(k)C_{23}(k)}{C_{21}(k)C_{13}(k)} \approx \frac{\Omega(k,\ell)|_{H_1}}{\Omega(k,\ell)|_{H_{0t}}} \tag{43}$$

where $\{C_{ij}(k) \mid i = 1,2;\ j = 1,2,3\}$ as specified in Eqs. (16)–(21) are independent of ℓ, since Γ_s and Γ_t are assumed independent of ℓ. Then from (32) it follows that a reliable discrimination between transient noise and desired signal components requires $Q(k) \gg 1$. In practice, given that H_1 is true, the distributions of the nominator and denominator in Eq. (30) are approximated by the chi-square distributions with μ degrees of freedom, and the distribution of the TBRR is approximated by the F-distribution:

$$\begin{aligned}
\mathcal{P}\left([\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)]|_{H_1} \le \epsilon\right) &= F_{\chi^2;\mu}\left(\frac{\mu\,\epsilon}{C_{11}(k)\lambda_x(k,\ell)}\right) \\
\mathcal{P}\left([\mathcal{S}Y(k,\ell) - \mathcal{M}Y(k,\ell)]|_{H_1} \le \epsilon\right) &= F_{\chi^2;\mu}\left(\frac{\mu\,\epsilon}{C_{21}(k)\lambda_x(k,\ell)}\right) \\
\mathcal{P}\left(\Omega(k,\ell)|_{H_1} \le \epsilon\right) &= F_{F;\mu,\mu}\left(\epsilon\frac{C_{21}(k)}{C_{11}(k)}\right)
\end{aligned}$$

where

$$F_{F;a,b}(x) \triangleq 1 - I_{(1+a\,x/b)^{-1}}\left(\frac{a}{2},\frac{b}{2}\right)$$

is the standard F distribution function, and $I_x(a,b)$ is the incomplete beta function [28]. We require that the probability of the TBRR be smaller than the thresholds $\Omega_{\text{high}}(k)$ and $\Omega_{\text{low}}(k)$, given that H_1 is true, to be 0.1 and 0.01, respectively, at the most:

$$\begin{aligned}
\mathcal{P}\left(\Omega(k,\ell)|_{H_1} \le \Omega_{\text{high}}(k)\right) &\le 0.1 \\
\mathcal{P}\left(\Omega(k,\ell)|_{H_1} \le \Omega_{\text{low}}(k)\right) &\le 0.01\,.
\end{aligned}$$

Hence, the thresholds are given by

$$\Omega_{\text{high}}(k) \;=\; F^{-1}_{F;\mu,\mu}\,(0.1)\;\frac{C_{11}(k)}{C_{21}(k)} = 0.57\,\frac{C_{11}(k)}{C_{21}(k)} \tag{44}$$

$$\Omega_{\text{low}}(k) \;=\; F^{-1}_{F;\mu,\mu}\,(0.01)\;\frac{C_{11}(k)}{C_{21}(k)} = 0.63\,\Omega_{\text{high}}(k) \tag{45}$$

where we used $\mu = 22.1$. This, together with the requirement that $\Omega_{\text{low}}(k)$ be larger than $C_{13}(k)/C_{23}(k)$, implies that a satisfactory discrimination performance can be obtained in frequency bins which are characterized by

$$Q(k) \geq 1/\,F^{-1}_{F;\mu,\mu}\,(0.01) = 2.78\,. \tag{46}$$

Substituting Eqs. (10) and (16)–(21) into (44) and (43), we have explicit expressions for the transient discrimination quality and for the upper threshold of the TBRR in terms of the spatial coherence functions and the uncertainty in the angle of arrival:

$$Q(k) = \frac{\left\{\cot\left(\frac{\Delta_k}{2}\right)\left[1-\Re\left\{e^{j\Delta_k}\Gamma_s(k)\right\}\right] - \Im\left\{e^{j\Delta_k}\Gamma_s(k)\right\}\right\}^2\left[1-\Re\left\{e^{j\Delta_k}\Gamma_t(k)\right\}\right]}{\left|1-e^{j\Delta_k}\Gamma_s(k)\right|^2 + \Re\left\{e^{j\Delta_k}\Gamma_t(k)\left[1-e^{-j\Delta_k}\Gamma_s^*(k)\right]^2\right\}} \tag{47}$$

$$\Omega_{\text{high}}(k) = 0.57\,\left[\cot\left(\tfrac{\Delta_k}{2}\right) - \frac{\Im\left\{e^{j\Delta_k}\Gamma_s(k)\right\}}{1-\Re\left\{e^{j\Delta_k}\Gamma_s(k)\right\}}\right]^2 . \tag{48}$$

We note that $\Omega_{\text{high}}(k)$ is independent of the transient noise field, since its value is determined by the confidence level associated with the TBRR given that H_1 is true, and we assumed that desired and interfering transients do not overlap in the time-frequency domain ($H_1 \cap H_{0t} = \emptyset$).

To realistically evaluate the discrimination capability of the proposed approach in various acoustic environments, we let the distance between the sensors be $l = 10$ cm, the mismatch in the source direction $\varphi = 5^\circ$, and the estimation error in the difference of phase $\phi = 5^\circ$. Figs. 6–8 show the transient discrimination quality for incoherent, diffuse and coherent noise fields. The respective upper thresholds for the TBRR are depicted in Fig. 9. Analytical expressions are derived in Appendix A.

Generally, the discrimination between desired and interfering transients is attainable within a certain frequency band. The requirement (46) that the transient discrimination quality should be large enough is satisfied over a wide range of frequencies. For low frequencies, the directivity of the beamformer and its spatial filtering capability are lost. For high frequencies, spatial aliasing folds interferences coming from the side to the main lobe. In these cases, the two-channel post-filtering reduces to single-channel post-filtering, since the transient power ratio between the beamformer output and the reference signal is no longer a distinctive characteristic of the transient source. In case of coherent noise fields, the discrimination is possible only if the interfering signals are coming from different directions than the look direction. Due to the error φ in the estimation of the angle of arrival, the direction to an interfering source should be at least $2\,\varphi$ away from the direction to the desired source.

V. Experimental Results

In this section, the proposed post-filtering approach is compared to a single-channel post-filtering in various car environments. The performance evaluation includes objective quality measures, as well as a subjective study of speech spectrograms and informal listening tests.

Two microphones with 10 cm spacing are mounted in a car on the visor. Clean speech signals are recorded at a sampling rate of 8 kHz in the absence of background noise (standing car, silent environment). An interfering speaker and 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 input microphone signals are generated by mixing the speech and noise signals at various SNR levels in the range $[-5, 10]$ dB.

Two-channel GSC beamforming is applied to the noisy signals. The beamformer output is enhanced using the OM-LSA estimator [25], and is referred to as the single-channel post-filtering output. Alternatively, the beamformer output, enhanced using the procedure described in Section III, is referred to as the two-channel post-filtering output. Three different objective quality measures are used in our evaluation. The first is segmental SNR defined by [29]

$$\mathrm{SegSNR} = \frac{1}{L}\sum_{\ell=0}^{L-1} 10\cdot\log\frac{\sum_{n=0}^{N-1} x^2(n+\ell N/2)}{\sum_{n=0}^{N-1}\left[x(n+\ell N/2)-\hat{x}(n+\ell N/2)\right]^2} \quad [\mathrm{dB}] \tag{49}$$

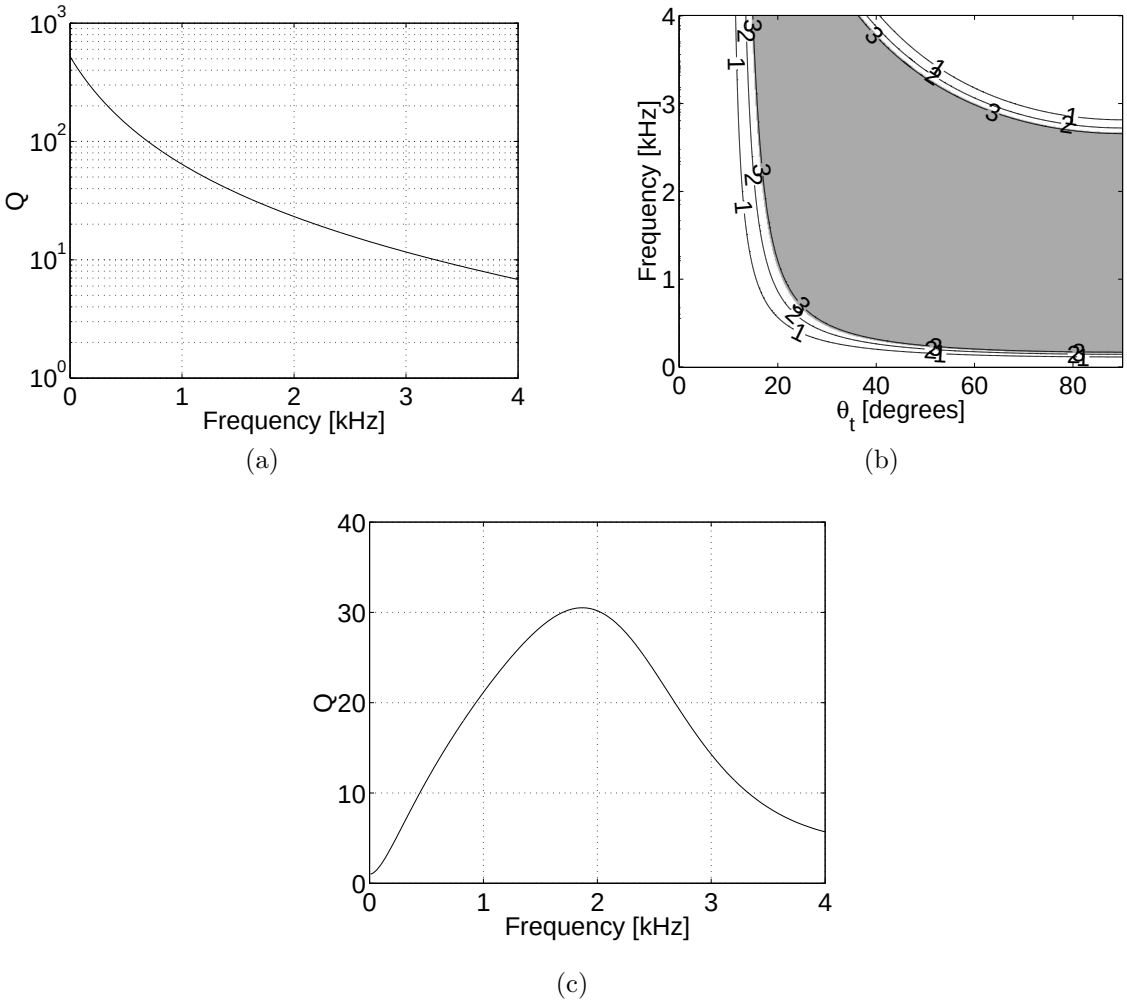


Fig. 6. Transient discrimination quality for incoherent pseudo-stationary noise and (a) incoherent, (b) coherent, and (c) diffuse transient noise fields. Referring to (b), θ_t is the angle of arrival of the transient noise field, and the dark area represents the region where Q is larger than 2.78 (region of satisfactory discrimination performance).

where L represents the number of frames in the signal, and $N = 256$ 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 [30], [31]. This measure takes into account both residual noise and speech distortion. The second quality measure is noise

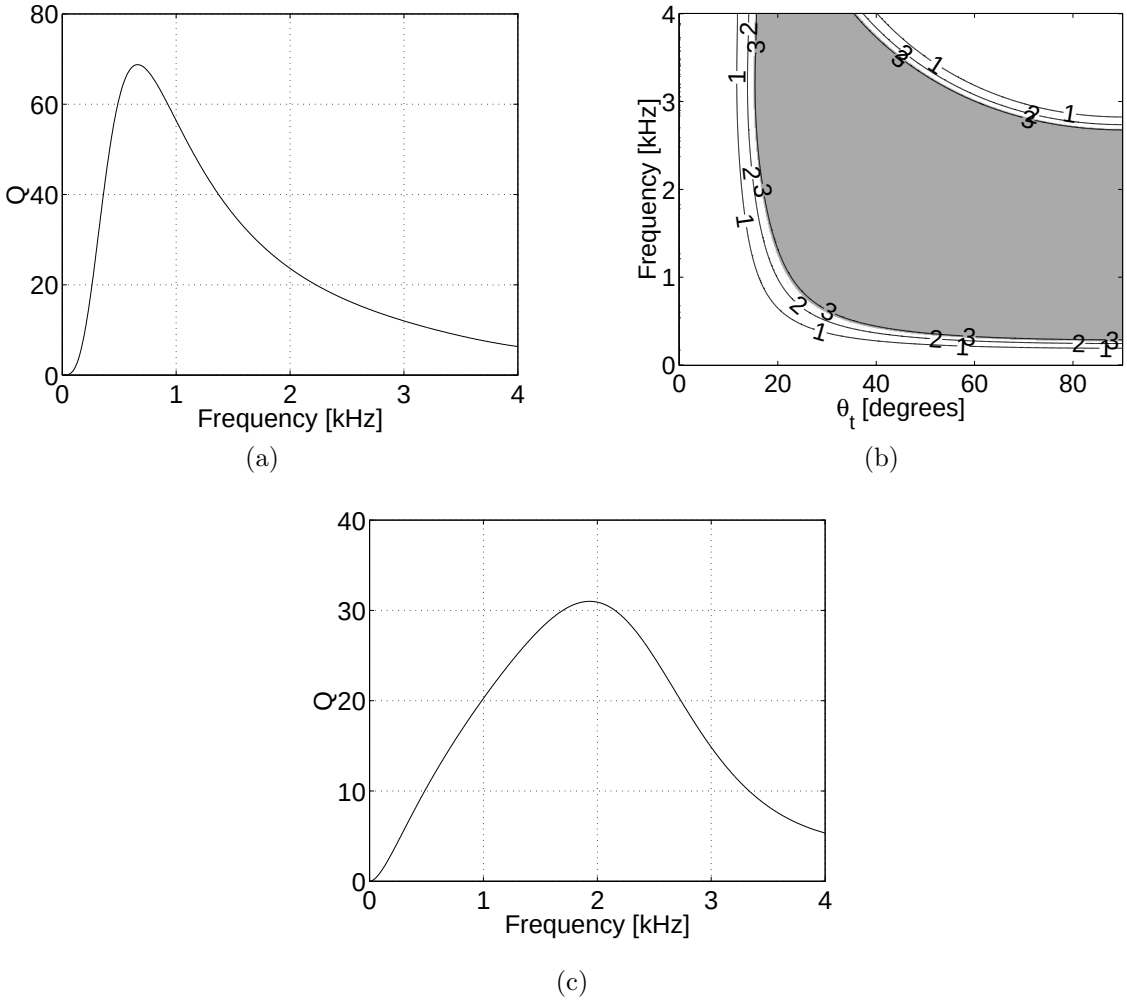


Fig. 7. Transient discrimination quality for diffuse pseudo-stationary noise and (a) incoherent, (b) coherent, and (c) diffuse transient noise fields. Referring to (b), θ_t is the angle of arrival of the transient noise field, and the dark area represents the region of satisfactory discrimination performance ($Q \geq 2.78$).

reduction (NR), which is defined by

$$\mathrm{NR} = \frac{1}{|\mathcal{L}'|} \sum_{\ell \in \mathcal{L}'} 10 \cdot \log \frac{\sum_{n=0}^{N-1} z_1^2(n + \ell N/2)}{\sum_{n=0}^{N-1} \hat{x}^2(n + \ell N/2)} \quad [\mathrm{dB}] \tag{50}$$

where $\mathcal{L}'$ represents the set of frames that contain only noise, and $|\mathcal{L}'|$ its cardinality. The NR measure compares the noise level in the enhanced signal to the noise level recorded by the first

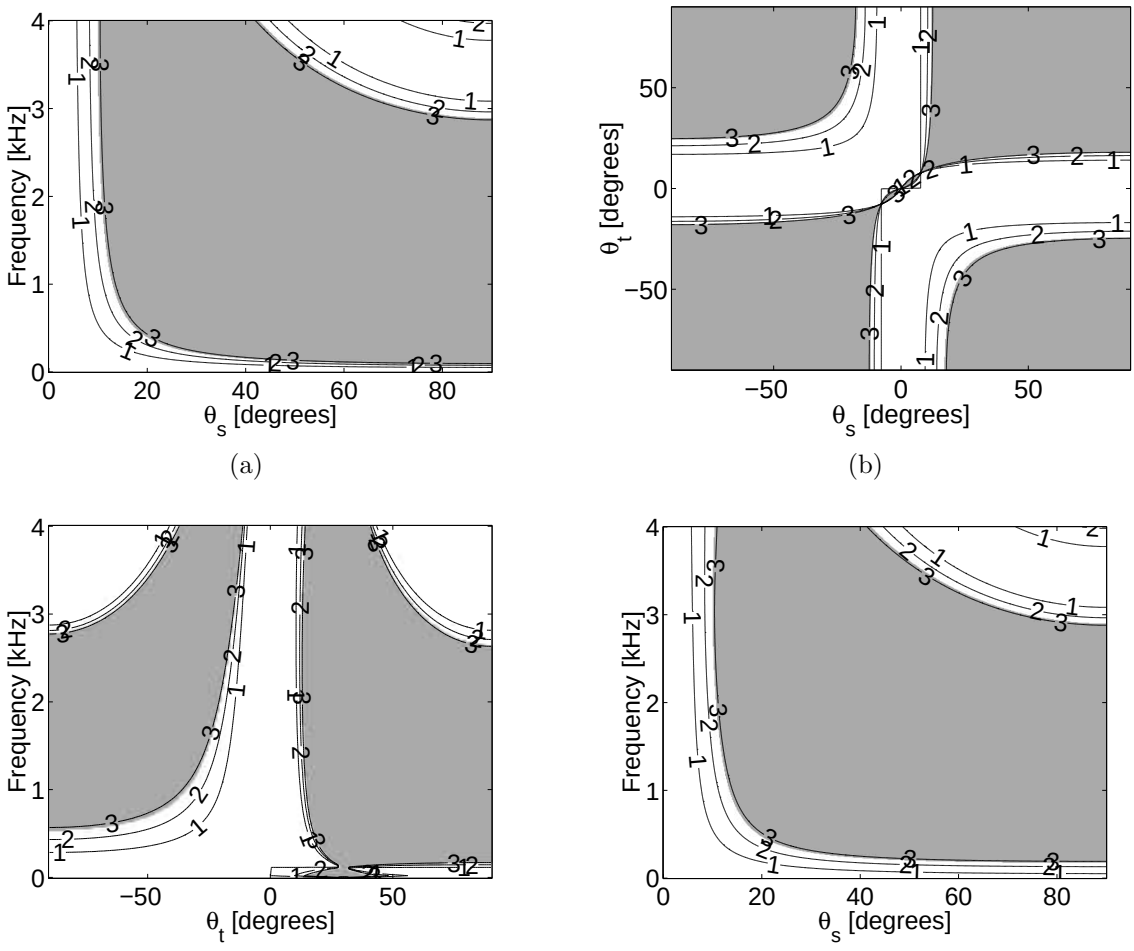


Fig. 8. Transient discrimination quality for coherent pseudo-stationary noise field whose angle of arrival is θ_s: (a) Transient noise is incoherent; (b) Transient noise is coherent and frequency is 1 kHz; (c) Transient noise is coherent and θ_s is 30 degrees; (d) Transient noise is diffuse. The dark areas represent the regions of satisfactory discrimination performance ($Q \geq 2.78$).

microphone. The third quality measure is log-spectral distance (LSD), which is defined by

$$\text{LSD} = \frac{1}{L}\sum_{\ell=0}^{L-1}\left\{\frac{1}{N/2+1}\sum_{k=0}^{N/2}\Big[10\cdot\log\mathcal{A}X(k,\ell) - 10\cdot\log\mathcal{A}\hat{X}(k,\ell)\Big]^2\right\}^{\frac{1}{2}} \quad [\text{dB}] \tag{51}$$

where $\mathcal{A}X(k,\ell) \triangleq \max\Big\{|X(k,\ell)|^2, \delta\Big\}$ is the spectral power, clipped such that the log-spectrum

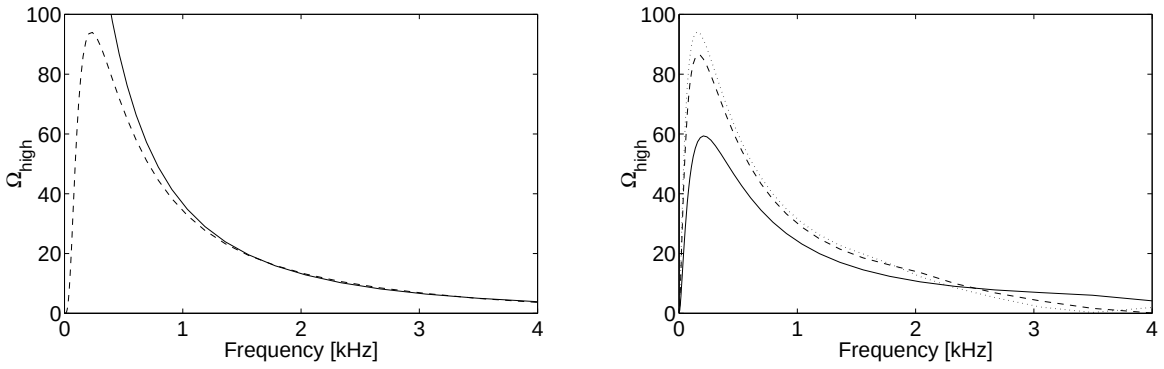


Fig. 9. Upper threshold for the transient beam-to-reference ratio in case the pseudo-stationary noise is (a) incoherent (solid), diffuse (dashed), or (b) coherent at $\theta_s = 30^o$(solid), $\theta_s = 60^o$(dashed), or $\theta_s = 90^o$(dotted).

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. 10 shows experimental results of the average segmental SNR, obtained for various noise types and at various noise levels. The segmental SNR is evaluated at one of the microphones, at the beamformer output, and at the post-filtering outputs. A theoretical limit post-filtering, achievable by calculating the noise spectrum from the noise itself, is also considered. Results of the NR and LSD measures are presented in Figs. 11 and 12, respectively. It shows that beamforming alone does not provide sufficient noise reduction in a car environment, owing to its limited ability to reduce diffuse noise [17]. Furthermore, two-channel post-filtering is consistently better than single-channel post-filtering under all noise conditions. The improvement in performance of the former over the latter is expectedly high in non-stationary noise environments (specifically, in case of open windows or an interfering speaker), but is insignificant otherwise, since two-channel post-filtering reduces to single-channel in pseudo-stationary noise environments.

A subjective comparison between two-channel and single-channel post-filtering was conducted using speech spectrograms and validated by informal listening tests. Typical examples of speech spectrograms are presented in Fig. 13 for the case of non-stationary noise at SNR = 0 dB. The window next to the driver is slightly open, inducing transient low-frequency noise due to wind blows, and wide-band transient noise due to passing cars. The beamformer output (Fig. 13(c)) is

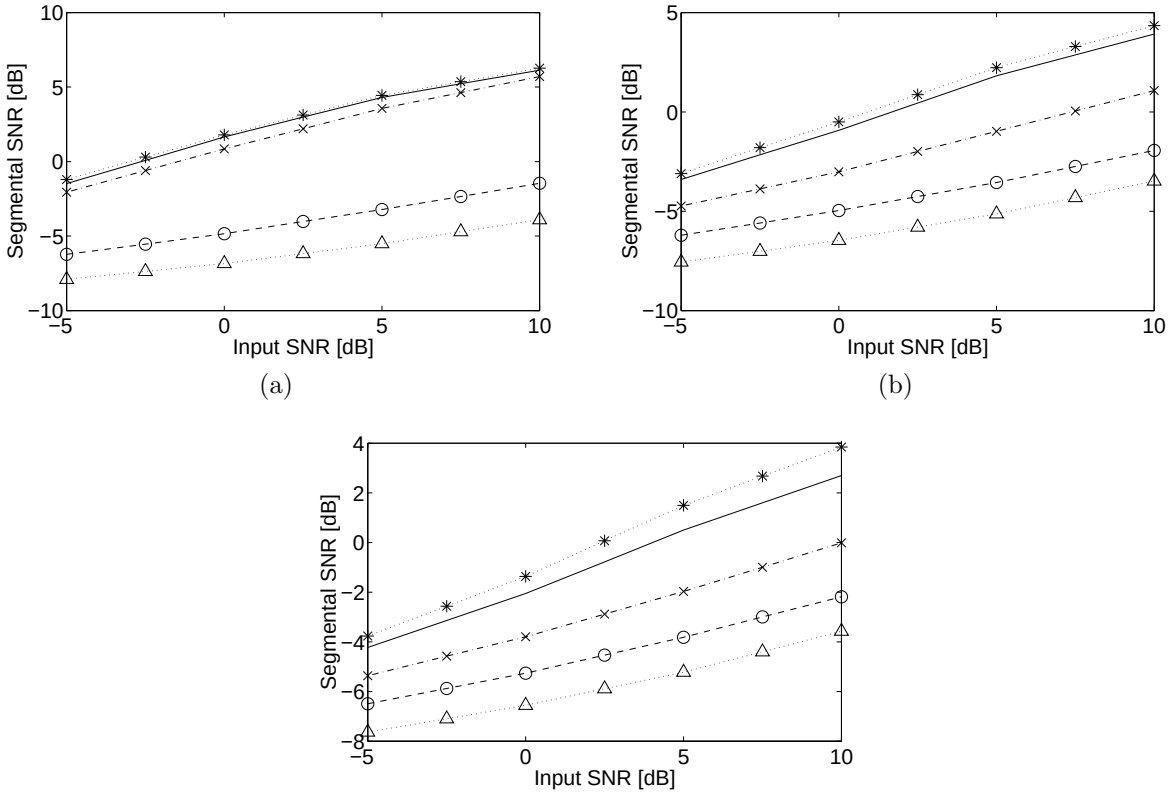


Fig. 10. Average segmental SNR at (△) microphone #1, (○) beamformer output, (×) single-channel post-filtering output, (solid line) two-channel post-filtering output, and (∗) theoretical limit post-filtering output, for various car noise conditions: (a) Closed windows; (b) Open windows; (c) Interfering speaker.

clearly characterized by a high level of noise. Its enhancement using single-channel post-filtering well suppresses the pseudo-stationary noise, but adversely retains the transient noise components. By contrast, the enhancement using two-channel post-filtering results in superior noise attenuation, while preserving the desired source components.

Fig. 14 shows traces of the improvement in segmental SNR and LSD measures, gained by the two-channel post-filtering and theoretical limit, in comparison with a single-channel post-filtering. The traces are averaged out over a period of about 400 ms (25 frames of 32 ms each, with 50% overlap). The improvement in performance over the single-channel post-filtering is obtained when the noise spectrum fluctuates. In some instances the increase in segmental SNR surpasses as much as 4 dB, and the decrease in LSD is greater than 5 dB. Clearly, a single-channel post-filter is inefficient

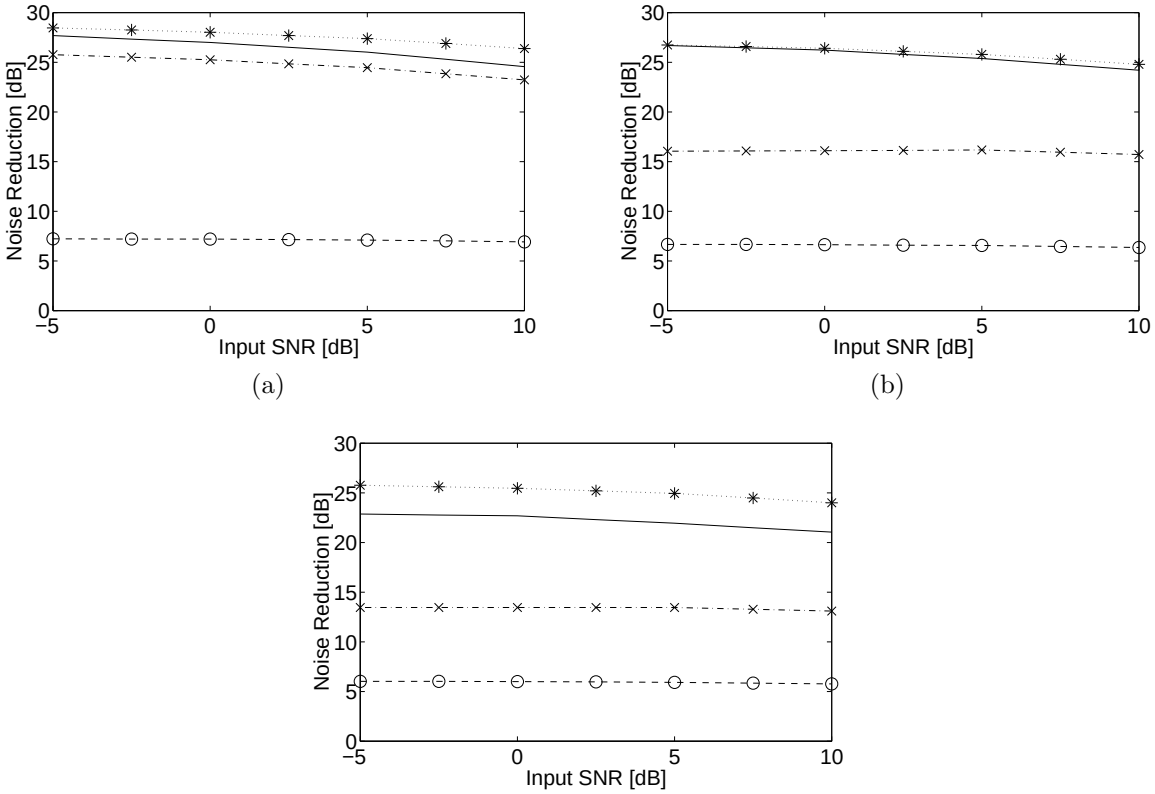


Fig. 11. Average noise reduction at (∘) beamformer output, (×) single-channel post-filtering output, (solid line) two-channel post-filtering output, and (∗) theoretical limit post-filtering output, for various car noise conditions: (a) Closed windows; (b) Open windows; (c) Interfering speaker.

at attenuating highly non-stationary noise components, since it lacks the ability to differentiate such components from the speech components. On the other hand, the proposed two-channel post-filtering approach achieves a significantly reduced level of background noise, whether stationary or not, without further distorting speech components. This is verified by subjective informal listening tests.

VI. Conclusion

We have analyzed a two-channel post-filtering approach for generalized sidelobe cancellers, that is particularly advantageous in non-stationary noise environments. The post-filtering includes detection of transients at the beamformer output and reference signal, a comparison of their transient

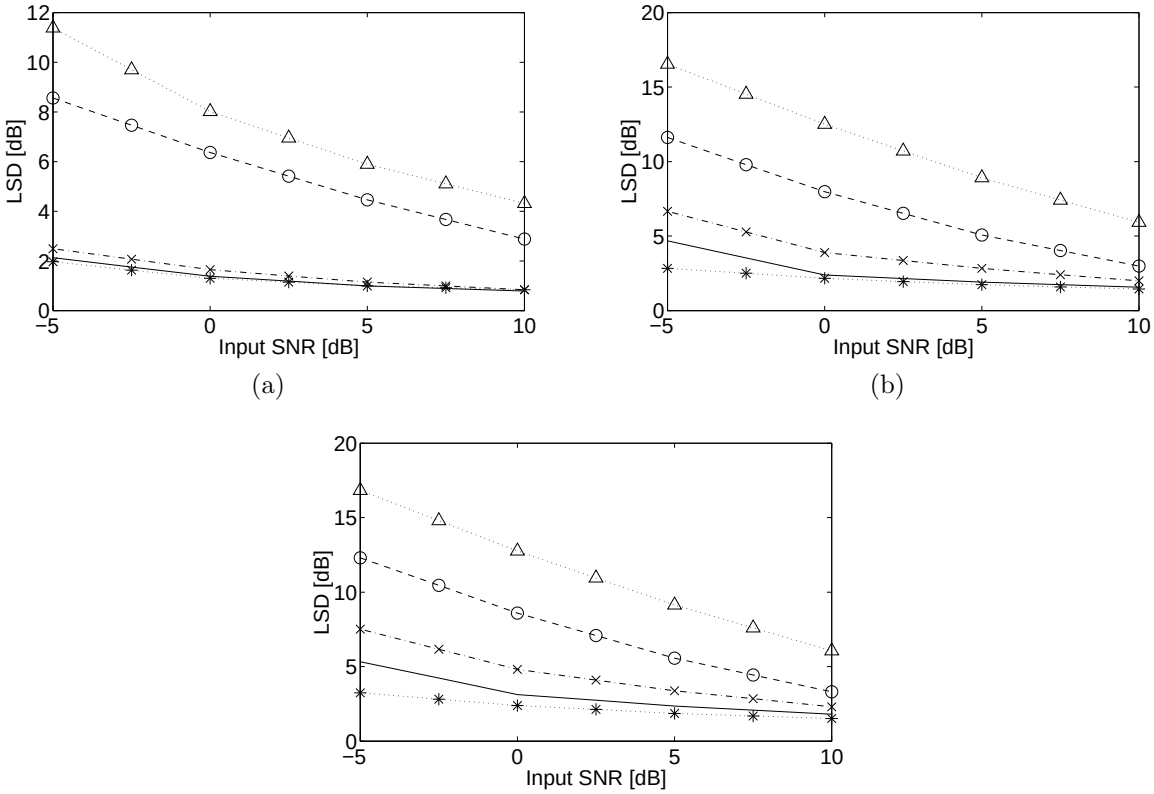


Fig. 12. Average log-spectral distance at (△) microphone #1, (○) beamformer output, (×) single-channel post-filtering output, (solid line) two-channel post-filtering output, and (∗) theoretical limit post-filtering output, for various car noise conditions: (a) Closed windows; (b) Open windows; (c) Interfering speaker.

power, estimation of the signal presence probability, estimation of the PSD of the beamformer output noise, and spectral enhancement for minimizing the mean-square error of the log-spectra. Transients are detected based on a measure of their local non-stationarity, and classified as desired or interfering based on the transient beam-to-reference ratio.

We introduced a *transient discrimination quality* measure, which quantifies the beamformer's capability to recognize interfering transients as distinct from source transients. Evaluating this measure in various noise fields shows that differentiating between desired and interfering transients is practicable within a wide range of frequencies. In case of coherent noise fields, such a discrimination is only possible if the interfering signals are coming from different directions than the desired source direction by at least twice the uncertainty in the angle of arrival. For low frequencies, the

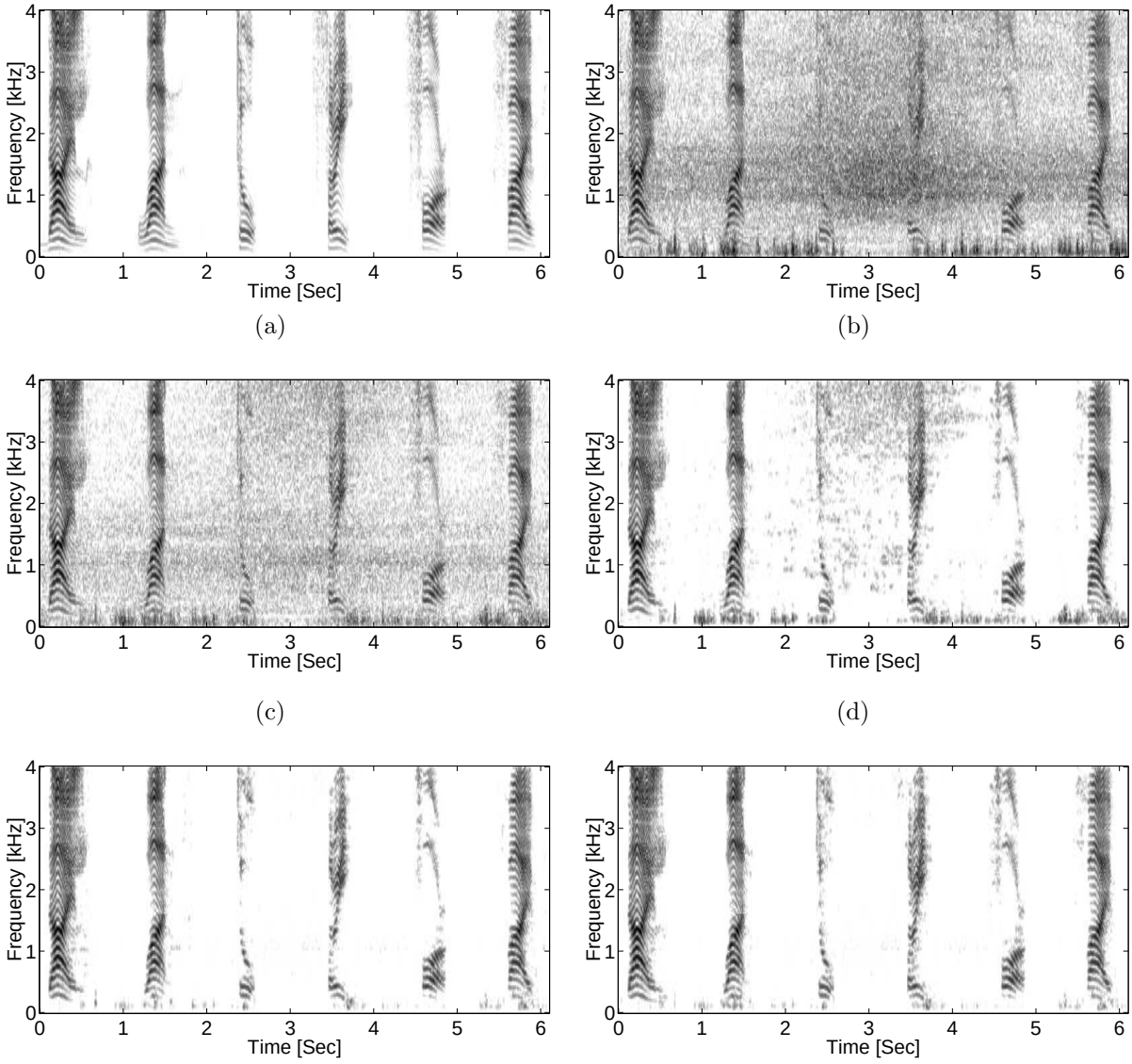
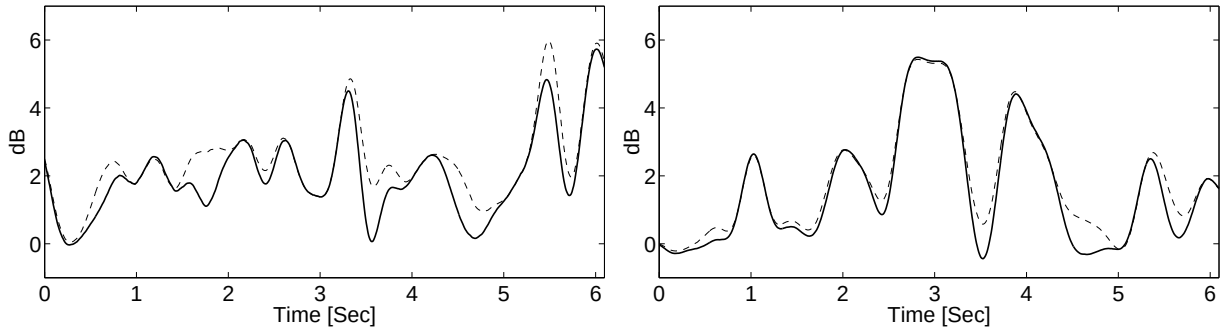


Fig. 13. Speech spectrograms. (a) Original clean speech signal at microphone #1: "Dial one two three four five."; (b) Noisy signal at microphone #1 (car noise, open window, interfering speaker. SNR = 0 dB, SegSNR = −6.5 dB, LSD = 12.5 dB); (c) Beamformer output (SegSNR = −5.0 dB, NR = 6.6 dB, LSD = 8.0 dB); (d) Single-channel post-filtering output (SegSNR = −3.0 dB, NR = 16.1 dB, LSD = 3.9 dB); (e) Multi-channel post-filtering output (SegSNR = −0.9 dB, NR = 26.2 dB, LSD = 2.4 dB); (f) Theoretical limit (SegSNR = −0.5 dB, NR = 26.4 dB, LSD = 2.1 dB).



(a) (b)

Fig. 14. Trace of the improvement over a single-channel post-filtering gained by the proposed two-channel post-filtering (solid) and theoretical limit (dashed): (a) Increase in segmental SNR; (b) Decrease in Log-Spectral Distance.

directivity of the beamformer is lost, and for high frequencies, the transient beam-to-reference ratio is no longer a distinctive characteristic of the transient source due to spatial aliasing.

In case the desired signal is wideband (*e.g.*, speech signal), we improve the transient noise reduction at low and high frequencies by considering a *global* likelihood of signal presence. The global likelihood is related to the number of frequency bins that likely contain desired components within a certain range of frequencies and at a given time frame. Whenever the global likelihood is lower than a certain threshold, the *a priori* signal absence probability is reset to one for all frequency bins. This also helps to eliminate narrow-band interfering transients arriving from the look direction, and uniformly suppresses the noise in a manner which is more pleasant to a human listener.

The proposed post-filtering approach is compared to state-of-the-art single-channel post-filtering in various car environments. We show that beamforming alone is insufficient in a car environment, due to its limited ability to reduce diffuse noise. Single-channel post-filtering well suppresses the pseudo-stationary noise. However, transient noise components that leak through the beamformer proceed through the post-filter. A single-channel post-filter is inefficient at attenuating highly non-stationary noise components, since it lacks the ability to differentiate such components from the speech components. By contrast, two-channel post-filtering results in a significantly reduced level of background noise, whether stationary or not, while preserving the desired source components.

Appendix

I. Computation of $Q(k)$ and $\Omega_{\text{high}}(k)$ for Various Acoustic Environments

In this appendix we compute the transient discrimination quality $Q(k)$ and the threshold $\Omega_{\text{high}}(k)$ for various acoustic environments. The pseudo-stationary and transient noise fields are assumed incoherent, coherent or diffuse. For incoherent noise field, the spatial coherence function is zero for all frequencies. In case a noise field is coherent, its spatial coherence function is $\Gamma(k) = \exp\left(-j\frac{\omega_k\, l}{c}\sin\theta\right)$, where θ is the angle of arrival. For a diffuse noise field, the spatial coherence function is $\Gamma(k) = \frac{\sin(\omega_k\, l/c)}{\omega_k\, l/c} = \text{sinc}\,(\omega_k\, l/c)$. Therefore, $Q(k)$ and $\Omega_{\text{high}}(k)$ are computed for various pseudo-stationary and transient noise fields by substituting the corresponding spatial coherence functions into Eqs. (47) and (48).

A. Incoherent Pseudo-Stationary Noise

Assuming the pseudo-stationary noise is incoherent ($\Gamma_s(k) = 0$), we have

$$Q(k) = \cot^2\left(\tfrac{\Delta_k}{2}\right)\frac{1 - \Re\left\{e^{j\Delta_k}\Gamma_t(k)\right\}}{1 + \Re\left\{e^{j\Delta_k}\Gamma_t(k)\right\}} \tag{52}$$

$$\Omega_{\text{high}}(k) = 0.57\,\cot^2\left(\tfrac{\Delta_k}{2}\right). \tag{53}$$

In case the transient noise is also incoherent ($\Gamma_t(k) = 0$), the transient discrimination quality reduces to

$$Q(k) = \cot^2\left(\tfrac{\Delta_k}{2}\right). \tag{54}$$

For coherent transient noise field, the spatial coherence function is $\Gamma_t(k) = \exp\left(-j\frac{\omega_k\, l}{c}\sin\theta_t\right) \triangleq \exp(-j\omega_k\tau_t)$, where θ_t is the angle of arrival of the interfering transient noise field. In this case, the transient discrimination quality is given by

$$Q(k,\theta_t) = \cot^2\left(\tfrac{\Delta_k}{2}\right)\frac{1 - \cos\left(\omega_k\tau_t - \Delta_k\right)}{1 + \cos\left(\omega_k\tau_t - \Delta_k\right)}. \tag{55}$$

For diffuse transient noise field, we have

$$Q(k) = \cot^2\left(\tfrac{\Delta_k}{2}\right) \frac{1 - \operatorname{sinc}(\omega_k\, l/c)\cos\Delta_k}{1 + \operatorname{sinc}(\omega_k\, l/c)\cos\Delta_k}\,. \tag{56}$$

B. Diffuse Pseudo-Stationary Noise

Assuming the pseudo-stationary noise is diffuse, we have

$$Q(k) = \frac{\cot^2(\Delta_k/2)\left[1 - \operatorname{sinc}(\omega_k\, l/c)\right]^2 \left[1 - \Re\left\{e^{j\Delta_k}\Gamma_t(k)\right\}\right]}{\left|e^{j\Delta_k} - \operatorname{sinc}(\omega_k\, l/c)\right|^2 + \Re\left\{e^{-j\Delta_k}\Gamma_t(k)\left[e^{j\Delta_k} - \operatorname{sinc}(\omega_k\, l/c)\right]^2\right\}} \tag{57}$$

$$\Omega_{\text{high}}(k) = 0.57\,\frac{\cot^2(\Delta_k/2)\left[1 - \operatorname{sinc}(\omega_k\, l/c)\right]^2}{\left[1 - \operatorname{sinc}(\omega_k\, l/c)\cos\Delta_k\right]^2}\,. \tag{58}$$

For incoherent transient noise field

$$Q(k) = \frac{\cot^2(\Delta_k/2)\left[1 - \operatorname{sinc}(\omega_k\, l/c)\right]^2}{1 - 2\operatorname{sinc}(\omega_k\, l/c)\cos\Delta_k + \operatorname{sinc}^2(\omega_k\, l/c)}\,. \tag{59}$$

For coherent transient noise field

$$Q(k,\theta_t) = \frac{\cot^2(\Delta_k/2)\left[1 - \operatorname{sinc}(\omega_k\, l/c)\right]^2\left[1 - \cos(\omega_k\tau_t - \Delta_k)\right]}{\left|e^{j\Delta_k} - \operatorname{sinc}(\omega_k\, l/c)\right|^2 + \Re\left\{e^{-j(\omega_k\tau_t + \Delta_k)}\left[e^{j\Delta_k} - \operatorname{sinc}(\omega_k\, l/c)\right]^2\right\}}\,. \tag{60}$$

For diffuse transient noise field

$$Q(k) = \cot^2\left(\tfrac{\Delta_k}{2}\right) \frac{1 - \operatorname{sinc}(\omega_k\, l/c)}{1 + \operatorname{sinc}(\omega_k\, l/c)}\,. \tag{61}$$

C. Coherent Pseudo-Stationary Noise

Assuming the pseudo-stationary noise is coherent, its spatial coherence function is $\Gamma_s(k) = \exp\left(-j\frac{\omega_k\, l}{c}\sin\theta_s\right) \triangleq \exp(-j\omega_k\tau_s)$, where θ_s is the angle of arrival. In this case,

$$Q(k,\theta_s) = \frac{\sin^2(\omega_k\tau_s/2)}{\sin^2(\Delta_k/2)}\,\frac{1 - \Re\left\{e^{j\Delta_k}\Gamma_t(k)\right\}}{1 - \Re\left\{e^{j\omega_k\tau_s}\Gamma_t(k)\right\}} \tag{62}$$

$$\Omega_{\text{high}}(k) = 0.57\,\frac{\sin^2(\omega_k\tau_s/2)}{\sin^2(\Delta_k/2)\sin^2(\omega_k\tau_s/2 - \Delta_k/2)}\,. \tag{63}$$

For incoherent transient noise field

$$Q(k,\theta_s) = \frac{\sin^2(\omega_k\tau_s/2)}{\sin^2(\Delta_k/2)}\,. \tag{64}$$

For coherent transient noise field

$$Q(k,\theta_s,\theta_t) = \frac{\sin^2(\omega_k\tau_s/2)}{\sin^2(\Delta_k/2)}\,\frac{1-\cos(\omega_k\tau_t-\Delta_k)}{1-\cos(\omega_k\tau_t-\omega_k\tau_s)}\,. \tag{65}$$

For diffuse transient noise field

$$Q(k,\theta_s) = \frac{\sin^2(\omega_k\tau_s/2)}{\sin^2(\Delta_k/2)}\,\frac{1-\operatorname{sinc}(\omega_k\,l/c)\cos\Delta_k}{1-\operatorname{sinc}(\omega_k\,l/c)\cos(\omega_k\tau_s)}\,. \tag{66}$$

References

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

[2] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 3, pp. 240–259, May 1998.

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

[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] K. U. Simmer and A. Wasiljeff, "Adaptive microphone arrays for noise suppression in the frequency domain," in *Proc. 2nd Cost-229 Workshop on Adaptive Algorithms in Communications*, Bordeaux, France, 30 September–2 October 1992, pp. 185–194.

[7] 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.

[8] 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.

[9] 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.

[10] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multichannel noise reduction - algorithms and theoretical limits," in *Proc. European Signal Processing Conference, EUSIPCO-98*, Rhodes, Greece, 8–11 September 1998, pp. 105–108.

[11] J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," in *Proc. 24th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-99*, Phoenix, Arizona, 15–19 March 1999, pp. 2965–2968.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] I. A. McCowan, C. Marro, and L. Mauuary, "Robust speech recognition using near-field superdirective beamforming with post-filtering," in *Proc. 25th IEEE Internat. Conf. Acoust. Speech Signal Process., ICASSP-2000*, Istanbul, Turkey, 5–9 June 2000, pp. 1723–1726.

[17] 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.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

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

[23] 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.

[24] 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.

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

[26] I. Cohen, "Noise spectrum estimation in adverse environments: Improved minima controlled recursive averaging," Technical Report, CCIT Report 357, EE Pub. 1291, Technion - Israel Institute of Technology, Haifa, Israel, October 2001.

[27] 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.

[28] R. N. McDonough and A. D. Whalen, *Detection of Signals in Noise*, Academic Press, San Diego, California, 2nd edition, 1995.

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

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

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