Available online at www.sciencedirect.com
ScienceDirect
Speech Communication 49 (2007) 623–635
SPEECH COMMUNICATION
www.elsevier.com/locate/specom

Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller ☆

Gal Reuven [a], Sharon Gannot [b], Israel Cohen [a,*]

[a] Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa 32000, Israel
[b] School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel

Received 29 January 2006; received in revised form 1 October 2006; accepted 16 December 2006

Abstract

Man machine interaction requires an acoustic interface for providing full duplex hands-free communication. The transfer-function generalized sidelobe canceller (TF-GSC) is an adaptive beamformer suitable for enhancing a speech signal received by an array of microphones in a noisy and reverberant environment. When an echo signal is also present in the microphone output signals, cascade schemes of acoustic echo cancellation and TF-GSC can be employed for suppressing both interferences. However, the performances obtainable by cascade schemes are generally insufficient. An acoustic echo canceller (AEC) that precedes the adaptive beamformer suffers from the noise component at its input. Acoustic echo cancellation following the adaptive beamformer lacks robustness due to time variations in the echo path affecting beamformer adaptation. In this paper, we introduce an *echo transfer-function generalized sidelobe canceller* (ETF-GSC), which combines the TF-GSC with an acoustic echo canceller. The proposed scheme consists of a primary TF-GSC for dealing with the noise interferences, and a secondary modified TF-GSC for dealing with the echo cancellation. The secondary TF-GSC includes an echo canceller embedded within a replica of the primary TF-GSC components. We show that using this structure, the problems encountered in the cascade schemes can be appropriately avoided. Experimental results demonstrate improved performance of the ETF-GSC compared to cascade schemes in noisy and reverberant environments.

1. Introduction

In many speech communication applications, e.g., audio-conference and hands-free IP telephony, the received multi-microphone speech signals are corrupted by acoustic background noise as well as by echo signals. The noise and echo components significantly degrade the intelligibility of the desired signal, and restrict the performance of subsequent speech processing systems, e.g., speech coding and speech recognition systems. Therefore, efficient methods for joint noise reduction and echo cancellation are generally desirable. The cases where only a single microphone or two microphones are available in the system, have been considered extensively, and a survey of techniques for combined noise and echo reduction can be found in (Jeannes et al., 2001). Here, we consider the case where an array of microphones is available in the system, and address the problem of efficiently combining an adaptive beamformer with acoustic echo cancellation.

Linearly constrained minimum variance (LCMV) beamforming (Frost, 1972) is a method for constructing a beampattern satisfying certain constraints on the array response for a set of directions, while minimizing the array response in all other directions. The LCMV beamformer is efficiently implemented as a generalized sidelobe canceller (GSC) (Griffiths and Jim, 1982) structure, which decouples the constraints and the minimization. The GSC has found numerous applications in the field of speech enhancement (e.g., Affes and Grenier, 1997; Nordholm et al., 1992; Hoshuyama et al., 1999; Bitzer et al., 1999). The GSC

☆ Part of this work will be presented in (Reuven et al., 2007).
* Corresponding author. Tel.: +972 4 8294731; fax: +972 4 8295757.
E-mail addresses: galrv@techunix.technion.ac.il (G. Reuven), gannot@eng.biu.ac.il (S. Gannot), icohen@ee.technion.ac.il (I. Cohen).

0167-6393/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.specom.2006.12.008

flexibility makes it a first choice for microphone array noise reduction, as well as for other interference cancellation tasks, e.g., echo cancellation. Kellermann (1997a,b) addressed the problem of combining acoustic echo cancellation and adaptive beamforming, and proposed two cascade schemes: one is an acoustic echo canceller (AEC) applied to each microphone signal followed by a beamformer (denoted AEC-BF), and the other is a beamformer (BF) followed by a single channel AEC (denoted BF-AEC). These schemes have been implemented in the time-domain. Kellermann further elaborated on these strategies and several other time-domain schemes in (Kellermann, 2001).

The *transfer-function generalized sidelobe canceller* (TF-GSC), introduced by Gannot et al. (2001), is more suitable than the conventional GSC for enhancing speech signals received by an array of microphones in an arbitrary acoustic enclosure. The TF-GSC involves estimation of the relative transfer-functions between distinct microphones with regard to the desired source, and construction of a blocking matrix that is designed to block the desired signal in a reverberant environment. Reuven et al. (2004) adopted Kellermann's approach for combined acoustic echo cancellation and adaptive beamforming, and examined two frequency-domain cascade schemes, comprising the TF-GSC and the block least-mean-square AEC. An experimental study showed that the AEC-BF is more advantageous than the BF-AEC, but both schemes suffer from insufficient echo cancellation. In the AEC-BF scheme, the AEC performance is impaired due to the presence of noise components, while in the BF-AEC scheme the performance of the AEC severely deteriorates due to time variations of the echo path that includes the adaptive beamformer.

Herbordt and Kellermann (2000) introduced a generalized sidelobe AEC (GSAEC) scheme, which contains an AEC embedded in the GSC beamformer. The AEC module is placed in the upper branch of the GSC, behind the *fixed beamformer* (FBF). Therefore, only a single AEC is required for an arbitrary number of array elements, as opposed to the AEC-BF scheme. They compared the performance of the GSAEC to the cascade AEC-BF and BF-AEC schemes, and experimentally confirmed that in reverberant environments the echo reduction can be improved by more than 13 dB relative to that obtainable by using the GSC. However, the AEC in the GSAEC scheme is designed to block the echo signal only in the upper branch, and therefore, the echo may leak into the beamformer output through the lower branch. Affes and Grenier (1997) proposed a GSC structure suitable for double-talk situations, where the desired and echo signals coexist. They presented a distortionless fixed beamformer constrained to cancel the echo, and a blocking matrix constrained to block both the desired signal and the echo signal. The acoustic transfer-functions are estimated using subspace tracking methods, and subsequently employed for constructing the fixed beamformer and blocking matrix. However, it is assumed that the noise signal is white, which restricts the applicability in practical environments. Furthermore, subspace methods are often sensitive to the statistical distribution of the input signal.

Dahl and Claesson (1999) considered a self-calibrating microphone array system for noise reduction and echo cancellation in a car environment. They proposed that the system will be calibrated when the car is parked, i.e., noise sources are excluded. Subsequently, when ambient noise is present and there is no near-end speech, the system utilizes the calibration signals and adapts the beamformer to the actual environment. Low and Nordholm (2005) proposed an alternative scheme that combines a *blind source separation* (BSS) pre-processor and a joint noise and echo canceller. After convergence of the BSS algorithm, the separation process yields two speech-dominant outputs, i.e., near-end target signal and far-end echo, while the remaining outputs are noise-dominant. The speech-dominant outputs are identified based on the kurtosis values (speech signals are assumed to have higher kurtosis than noise), out of which the echo-dominant output is determined based on its relatively higher coherence with the echo signal. The BSS noise-dominant and echo-dominant outputs, as well as the echo signal itself, are employed as reference signals for the subsequent joint adaptive echo and noise canceller (AENC). The AENC enhances the near-end speech-dominant output by cancelling the components that are temporally correlated with the reference signals.

Herbordt et al. (2004) proposed to combine the optimization criterion for both interferences in a GSC structure. As opposed to the cascade schemes, the output signal controls the adaptation of both the noise canceller and the echo cancellation. The combined system is denoted *generalized echo and interference canceller* (GEIC). The efficiency of the proposed solution was demonstrated by speech recognition experiments under conditions of high-level background noise, time-varying echo paths, and frequent double-talk situations (Herbordt et al., 2005). Kammeyer et al. (2005) explored several adaptation strategies for combined noise and echo reduction. They showed that in the AEC-BF structure, the echo canceller can be adapted by the output of the beamformer as well. Hence, the noise sensitivity of the AEC adaptation can be avoided. A comprehensive survey of strategies for combining acoustic echo cancellers and noise reduction systems can be found in (Herbordt, 2005).

Doclo et al. (2000) proposed and compared two additional schemes for noise reduction and echo cancellation. The first scheme includes an M-channel AEC followed by a *generalized singular value decomposition* (GSVD) based beamformer. The second scheme incorporates the far-end echo reference directly into the GSVD beamformer, without cancelling the echo in every received signal. Simulations indicate that the first scheme outperforms the latter. Rombouts and Moonen (2005) combined the speech enhancement and echo cancellation tasks in one integrated scheme, and showed that its performance is superior to the performance of traditional cascade schemes. The optimization problem defined by this scheme is solved adaptively using an QR-decomposition-based least squares lattice algorithm.

In this paper, we introduce an *echo transfer-function generalized sidelobe canceller* (ETF-GSC), for joint echo cancellation and noise reduction in a reverberant environment. The proposed scheme consists of a primary TF-GSC, which is designed for noise suppression, and a secondary modified TF-GSC, which is designed for echo cancellation. The secondary TF-GSC comprises an M-channel echo cancellation embedded within a replica of the primary TF-GSC components. This structure has a twofold advantage. On one hand, it guarantees that no re-estimation of already available components is performed due to the variations of the beamformer (as in the BF-AEC structure). On the other hand, the presence of noise does not deteriorate the performance of the echo canceller (as in the AEC-BF structure). The proposed scheme, which is adapted using the entire system output, decouples the noise and cancellation tasks, and hence overcomes many of the problems encountered in the cascade application of the AEC and TF-GSC blocks. Experimental results demonstrate the improved performance of the ETF-GSC compared to AEC-BF and BF-AEC schemes in noisy and reverberant environments. However, the computational burden imposed by the new scheme is higher than the other two schemes.

The structure of this work is as follows. In Section 2, we formulate the problem of joint noise reduction and acoustic echo cancellation. In Section 3, we review the cascade schemes of acoustic echo cancellation and adaptive beamforming. In Section 4, we introduce the ETF-GSC. In Section 5, we evaluate its performance with comparison to the AEC-BF and BF-AEC cascade schemes.

2. Problem formulation

We consider a microphone array that receives signals from three types of sources: a desired source, an echo source, and interfering sources, as depicted in Fig. 1. Let $s(t)$ represent the desired source signal, and let $e(t)$ represent the transmitted far-end signal. Let $a_m(t)$ denote the acoustic impulse response (AIR) from the desired source to the mth microphone, and let $b_m(t)$ denote the AIR of the loudspeaker-enclosure-microphone (LEM) system corresponding to the mth microphone. Then, the signal received by the mth microphone can be written as

$$z_m(t) = a_m(t) * s(t) + b_m(t) * e(t) + n_m(t), \quad m = 1, \ldots, M, \tag{1}$$

where $n_m(t)$ represents the interference signals in the mth microphone and $*$ denotes convolution.

Typically, the impulse responses $a_m(t)$ and $b_m(t)$ are slowly changing in time and can be considered time-invariant over the analysis interval. In the *short-time Fourier transform* (STFT) domain, we can employ the multiplicative transfer-function (MTF) approximation (Avargel and Cohen, in press), and rewrite (1) as

$$z_m(t, \mathrm{e}^{\mathrm{j}\omega}) \approx a_m(\mathrm{e}^{\mathrm{j}\omega}) s(t, \mathrm{e}^{\mathrm{j}\omega}) + b_m(\mathrm{e}^{\mathrm{j}\omega}) e(t, \mathrm{e}^{\mathrm{j}\omega}) + n_m(t, \mathrm{e}^{\mathrm{j}\omega}), \quad m = 1, \ldots, M, \tag{2}$$

where $z_m(t, \mathrm{e}^{\mathrm{j}\omega})$, $s(t, \mathrm{e}^{\mathrm{j}\omega})$, $e(t, \mathrm{e}^{\mathrm{j}\omega})$, and $n_m(t, \mathrm{e}^{\mathrm{j}\omega})$ are the STFTs of the respective signals, $a_m(\mathrm{e}^{\mathrm{j}\omega})$ is the acoustical transfer-function (ATF) from the desired source to the mth microphone, and $b_m(\mathrm{e}^{\mathrm{j}\omega})$ is the ATF from the echo source to the mth microphone. A vector formulation of (2) is

$$\boldsymbol{z}(t, \mathrm{e}^{\mathrm{j}\omega}) = \boldsymbol{a}(\mathrm{e}^{\mathrm{j}\omega}) s(t, \mathrm{e}^{\mathrm{j}\omega}) + \boldsymbol{b}(\mathrm{e}^{\mathrm{j}\omega}) e(t, \mathrm{e}^{\mathrm{j}\omega}) + \boldsymbol{n}(t, \mathrm{e}^{\mathrm{j}\omega}), \tag{3}$$

where

$$\begin{aligned}
\boldsymbol{z}(t, \mathrm{e}^{\mathrm{j}\omega}) &= \left[z_1(t, \mathrm{e}^{\mathrm{j}\omega}) \quad z_2(t, \mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad z_M(t, \mathrm{e}^{\mathrm{j}\omega}) \right]^{\mathrm{T}}, \\
\boldsymbol{a}(\mathrm{e}^{\mathrm{j}\omega}) &= \left[a_1(\mathrm{e}^{\mathrm{j}\omega}) \quad a_2(\mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad a_M(\mathrm{e}^{\mathrm{j}\omega}) \right]^{\mathrm{T}}, \\
\boldsymbol{b}(\mathrm{e}^{\mathrm{j}\omega}) &= \left[b_1(\mathrm{e}^{\mathrm{j}\omega}) \quad b_2(\mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad b_M(\mathrm{e}^{\mathrm{j}\omega}) \right]^{\mathrm{T}}, \\
\boldsymbol{n}(t, \mathrm{e}^{\mathrm{j}\omega}) &= \left[n_1(t, \mathrm{e}^{\mathrm{j}\omega}) \quad n_2(t, \mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad n_M(t, \mathrm{e}^{\mathrm{j}\omega}) \right]^{\mathrm{T}}.
\end{aligned}$$

Our problem is to reconstruct the desired speech signal $s(t, \mathrm{e}^{\mathrm{j}\omega})$ (or a filtered version thereof) from the noisy observations $\boldsymbol{z}(t, \mathrm{e}^{\mathrm{j}\omega})$ and the available echo signal $e(t, \mathrm{e}^{\mathrm{j}\omega})$.

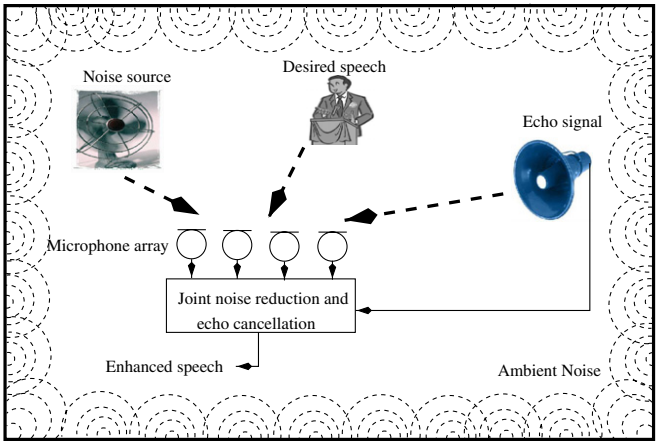

Fig. 1. Desired and echo signals in a noisy and reverberant environment.

3. Cascade schemes

In this section, we review cascade schemes of acoustic echo cancellation and TF-GSC implemented in the frequency-domain (Reuven et al., 2004). The echo signal is suppressed by the AEC, while the background noise is reduced by applying multi-microphone signal enhancement techniques, e.g., fixed and adaptive beamforming algorithms. Similar to the approach in (Kellermann, 1997b), we have presented in (Reuven et al., 2004) two cascade schemes of TF-GSC and acoustic echo canceller for joint noise reduction and echo cancellation. In this section we further elaborate on the schemes, and subsequently in Section 5 present a more comprehensive experimental study together with a comparison to the newly proposed method. As depicted in Figs. 2 and 3, the first scheme consists of M parallel AECs, applied to the microphone signals, followed by a beamformer (denoted AEC-BF) and the second scheme consists of a beamformer followed by a single

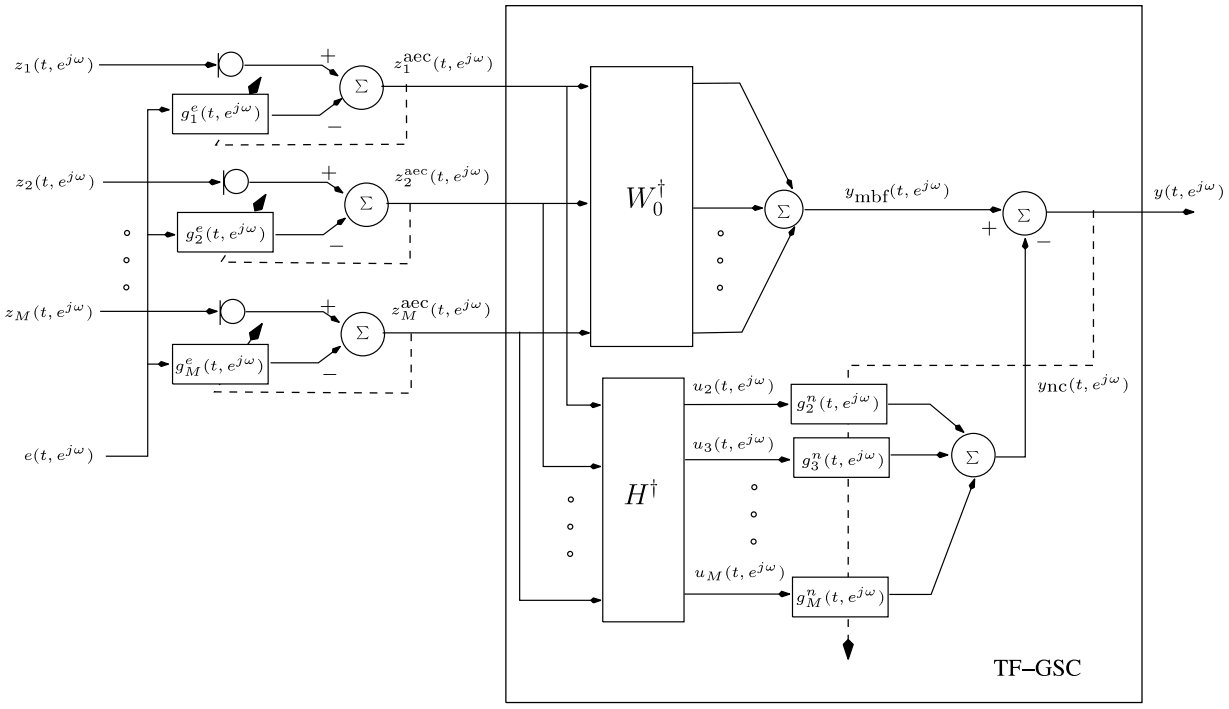

Fig. 2. Cascade scheme of acoustic echo cancellation and TF-GSC.

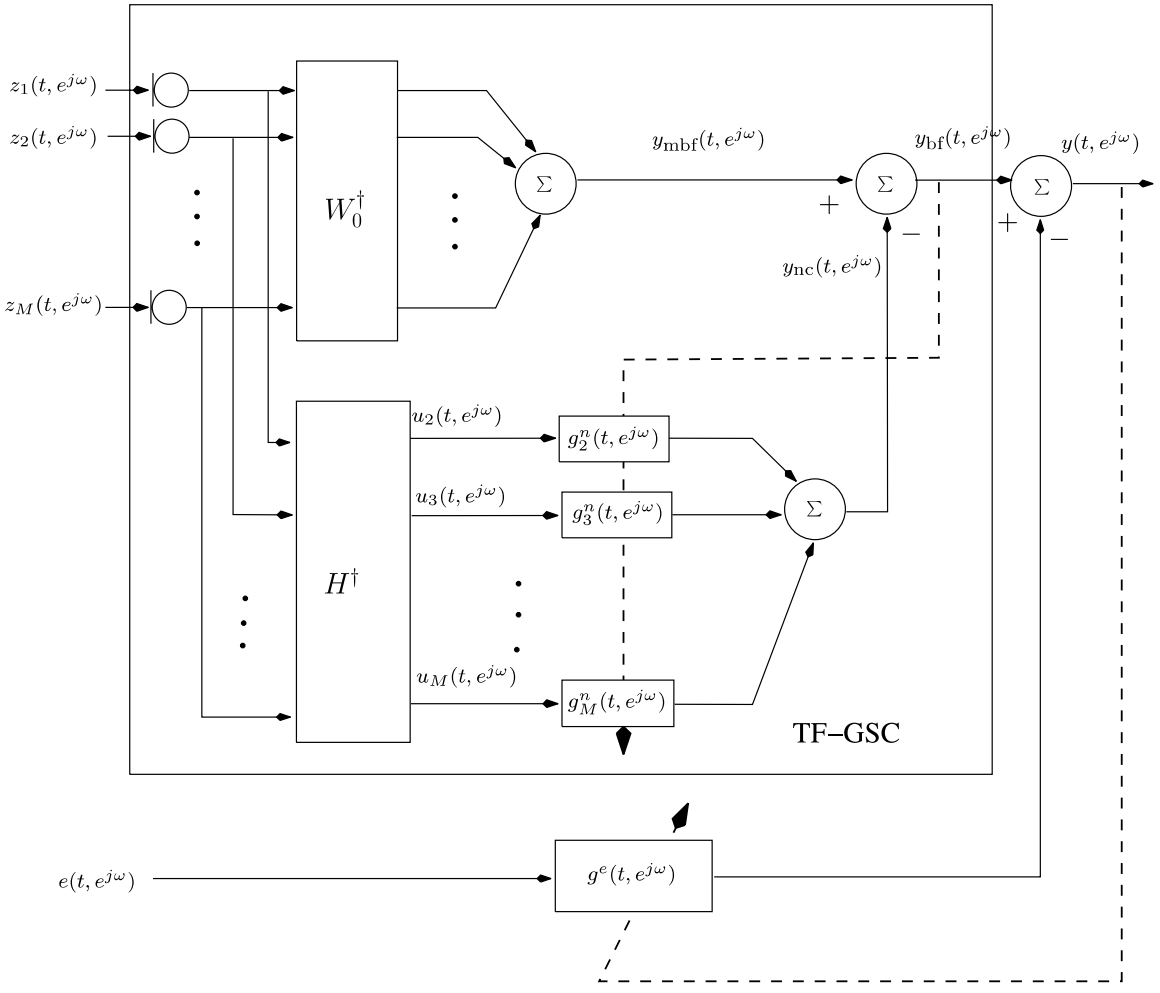

Fig. 3. Cascade scheme of TF-GSC and acoustic echo cancellation.

channel AEC (denoted BF-AEC). Note that the implementation is in the time–frequency domain, which allows faster convergence regardless of the condition number of the input data correlation matrix, and therefore is more suitable for speech processing. In both schemes, the frequency-domain beamformer is implemented by the TF-GSC (Gannot et al., 2001), while the AEC is implemented using the block least-mean-square (BLMS) algorithm (Shynk, 1992).

3.1. Cascade of AEC and TF-GSC

In the AEC-BF scheme, each AEC block receives an input signal that comprises the desired signal, echo signal, and noise. Using the available far-end signal $e(t,\mathrm{e}^{\mathrm{j}\omega})$ and an estimate of the echo path $g_m^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$, the AEC enhances the desired signal by cancelling the echo component. The subsequent beamformer uses the enhanced signals to mitigate the noise by steering the array towards the desired source direction and removing the estimated noise component, while assuming that the echo signal was already cancelled.

The M-channel AEC output signals are

$$z_m^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega}) = z_m(t,\mathrm{e}^{\mathrm{j}\omega}) - (g_m^{\mathrm{e}})^*(t,\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}), \quad m = 1,\ldots,M, \tag{4}$$

where $g_m^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$ is the mth AEC filter, and the superscript * denotes complex conjugation. The filter is updated using the BLMS algorithm,

$$\begin{aligned} \tilde{g}_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &= g_m^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) + \mu^{\mathrm{e}}\frac{e(t,\mathrm{e}^{\mathrm{j}\omega})(z_m^{\mathrm{aec}})^*(t,\mathrm{e}^{\mathrm{j}\omega})}{p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})}, \\ g_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &\overset{\mathrm{FIR}}{\leftarrow} \tilde{g}_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}), \end{aligned} \tag{5}$$

where $p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$ is updated using

$$p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) = \eta^{\mathrm{e}}p_{\mathrm{est}}^{\mathrm{e}}(t-1,\mathrm{e}^{\mathrm{j}\omega}) + (1-\eta^{\mathrm{e}})|e(t,\mathrm{e}^{\mathrm{j}\omega})|^2. \tag{6}$$

η^{e} is the power update forgetting factor, and μ^{e} is the step size of the normalized least-mean-square (NLMS) algorithm. Since filtering is realized using multiplication in the frequency-domain, aliasing effects due to cyclic convolution must be eliminated by imposing an FIR constraint, denoted as $\overset{\mathrm{FIR}}{\leftarrow}$.

The M AEC outputs, $\boldsymbol{z}^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega})$, are then processed by the TF-GSC algorithm. The matched beamformer output signal is given by

$$y_{\mathrm{mbf}}(t,\mathrm{e}^{\mathrm{j}\omega}) = \boldsymbol{1}^{\mathrm{T}}W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega}), \tag{7}$$

where $W_0(\mathrm{e}^{\mathrm{j}\omega})$ is the matched beamformer matrix, as defined later in (18), and $\boldsymbol{1}$ is a column vector of ones, used for calculating the sum of $W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega})$ components. The noise component $y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega})$ is evaluated as

$$y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega}) = (\boldsymbol{g}^{\mathrm{n}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega}), \tag{8}$$

where $\boldsymbol{g}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) = [\,g_2^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) \quad g_3^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad g_M^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})\,]^{\mathrm{T}}$. The blocking matrix $H(\mathrm{e}^{\mathrm{j}\omega})$ is given later by (19), and the noise canceller filters are updated using the BLMS algorithm:

$$\begin{aligned} \tilde{g}_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &= g_m^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) + \mu^{\mathrm{n}}\frac{u_m(t,\mathrm{e}^{\mathrm{j}\omega})y^*(t,\mathrm{e}^{\mathrm{j}\omega})}{p_{\mathrm{est}}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})}, \\ g_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &\overset{\mathrm{FIR}}{\leftarrow} \tilde{g}_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}), \end{aligned} \tag{9}$$

where $p_{\mathrm{est}}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})$ is updated using

$$p_{\mathrm{est}}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) = \eta^{\mathrm{n}}p_{\mathrm{est}}^{\mathrm{n}}(t-1,\mathrm{e}^{\mathrm{j}\omega}) + (1-\eta^{\mathrm{n}})\|\boldsymbol{u}(t,\mathrm{e}^{\mathrm{j}\omega})\|^2 \tag{10}$$

and η^{n} is the forgetting factor.

Similarly to the echo cancellation, only non-aliased samples are kept while calculating the output signal of the beamformer (Shynk, 1992). Finally, the beamformer output signal is obtained as

$$y(t,\mathrm{e}^{\mathrm{j}\omega}) = y_{\mathrm{mbf}}(t,\mathrm{e}^{\mathrm{j}\omega}) - y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega}). \tag{11}$$

When the M-channel AEC precedes the beamformer, the AEC input signals are contaminated by noise, which degrades the echo cancellation performance, especially for medium and low echo-to-noise ratios (Herbordt et al., 2004). On the other hand, the input signals of the BF are free of echo in this scheme, and therefore, the directivity of the BF is utilized for cancelling the noise signal alone.

3.2. Cascade of TF-GSC and AEC

In the BF-AEC scheme, the BF now receives input signals that comprise the desired signal, echo signal, and noise. The beamformer enhances the desired signal by reducing the noise component at the output. Using the available far-end signal $e(t,\mathrm{e}^{\mathrm{j}\omega})$ and the filter $g^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$, the subsequent AEC enhances the desired signal by cancelling the echo component.

The first stage of the combined system is the TF-GSC algorithm. The matched beamformer output signal is

$$y_{\mathrm{mbf}}(t,\mathrm{e}^{\mathrm{j}\omega}) = \boldsymbol{1}^{\mathrm{T}}W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}(t,\mathrm{e}^{\mathrm{j}\omega}). \tag{12}$$

The estimated noise component, $y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega})$, is evaluated as

$$y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega}) = (\boldsymbol{g}^{\mathrm{n}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}(t,\mathrm{e}^{\mathrm{j}\omega}) \tag{13}$$

and the beamformer output signal is

$$y_{\mathrm{bf}}(t,\mathrm{e}^{\mathrm{j}\omega}) = y_{\mathrm{mbf}}(t,\mathrm{e}^{\mathrm{j}\omega}) - y_{\mathrm{nc}}(t,\mathrm{e}^{\mathrm{j}\omega}). \tag{14}$$

The subsequent single-channel AEC then employs the beamformer output as its input signal. The resulting AEC output signal is

$$y(t,\mathrm{e}^{\mathrm{j}\omega}) = y_{\mathrm{bf}}(t,\mathrm{e}^{\mathrm{j}\omega}) - (g^{\mathrm{e}})^*(t,\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}), \tag{15}$$

where $g^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$ is the AEC filter. The filters are updated using the BLMS algorithm,

$$\begin{aligned} \tilde{g}^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &= g^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) + \mu^{\mathrm{e}}\frac{e(t,\mathrm{e}^{\mathrm{j}\omega})y^*(t,\mathrm{e}^{\mathrm{j}\omega})}{p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})}, \\ g^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) &\overset{\mathrm{FIR}}{\leftarrow} \tilde{g}^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}), \end{aligned} \tag{16}$$

where $p^{e}_{est}(t, e^{j\omega})$ is updated using (6). Similarly to the previous scheme, only non-aliased samples are kept while calculating the output signal of the beamformer and the AEC.

When the beamformer precedes the AEC, the echo cancellation performance is less impaired by the noise, since the beamformer has already reduced it. However, the performance of the AEC may deteriorate due to time variations of the echo path caused by the beamformer.

4. Combined scheme of TF-GSC and AEC

In the previous section, we argued that the AEC performance in both the AEC-BF scheme and BF-AEC scheme is insufficient. The AEC performance in the AEC-BF scheme is significantly impaired due to the noise at the received signals, while in the BF-AEC scheme, performance severely deteriorates due to time variations of the echo path that includes the beamformer. In this section, we show that noise and echo can be jointly mitigated by using a dual scheme. The combined scheme consists of a primary TF-GSC, for dealing with the noise cancellation task, and a secondary modified TF-GSC, which is designed for the echo cancellation task. The secondary TF-GSC comprises an M-channel echo cancellation embedded within a replica of the primary TF-GSC components. This structure has a twofold advantage. On one hand, it guarantees that no re-estimation of already available components is necessary during adaptation of the beamformer (as in the BF-AEC structure). On the other hand, the presence of the noise signal does not significantly deteriorate the performance of the echo canceller (as in the AEC-BF structure). Hence, the novel ETF-GSC structure suits the problem at hand better than the cascade schemes. The overall ETF-GSC scheme is depicted in Fig. 4. We now turn to a detailed description of the proposed system.

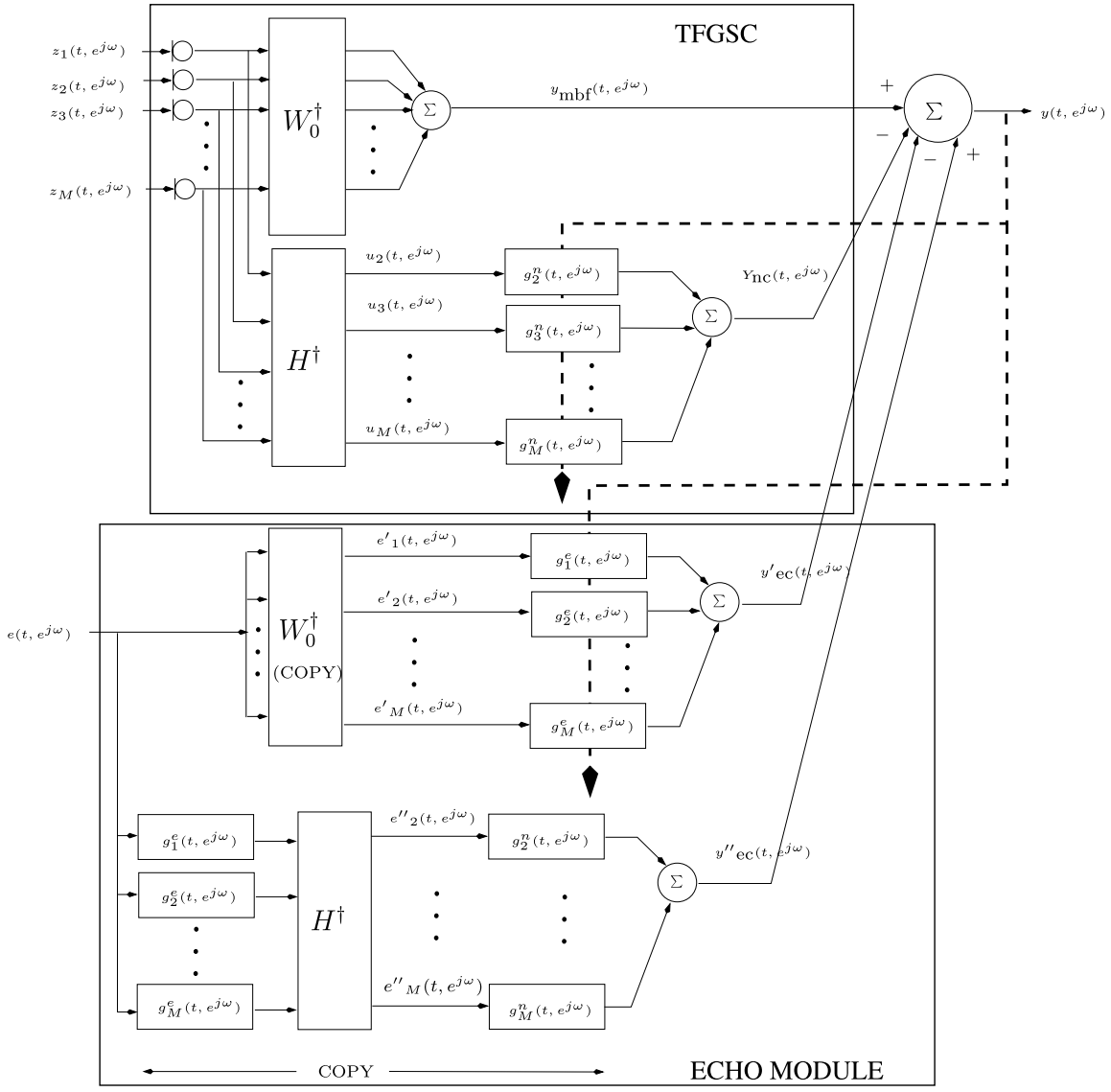

Fig. 4. Echo transfer-function generalized sidelobe canceller (ETF-GSC).

4.1. Primary TF-GSC

4.1.1. Matched beamformer

The design of the matched beamformer (MBF) is essentially identical to the design of the respective block in the conventional TF-GSC. Define the transfer-function ratios by

$$\tilde{\boldsymbol{a}}(\mathrm{e}^{\mathrm{j}\omega}) \triangleq \frac{\boldsymbol{a}(\mathrm{e}^{\mathrm{j}\omega})}{a_1(\mathrm{e}^{\mathrm{j}\omega})}. \tag{17}$$

Then, as shown in (Gannot et al., 2001), the following beamformer can serve as an MBF:

$$W_0(\mathrm{e}^{\mathrm{j}\omega}) = \frac{1}{\|\tilde{\boldsymbol{a}}(\mathrm{e}^{\mathrm{j}\omega})\|^2} \begin{bmatrix} 1 & 0 & \cdots & \cdots & 0 \\ 0 & \tilde{a}_2(\mathrm{e}^{\mathrm{j}\omega}) & 0 & \cdots & 0 \\ \vdots & \cdots & \ddots & & \vdots \\ \vdots & \cdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & \tilde{a}_M(\mathrm{e}^{\mathrm{j}\omega}) \end{bmatrix}, \tag{18}$$

where $\|\tilde{\boldsymbol{a}}(\mathrm{e}^{\mathrm{j}\omega})\|$ is the Euclidean (ℓ_2) norm of the vector $\tilde{\boldsymbol{a}}(\mathrm{e}^{\mathrm{j}\omega})$. The MBF output is given by (12). The role of the MBF block is to maintain the desired signal direction. It is evident that the transfer-function ratios suffice for implementing this task. Some noise reduction and echo suppression can be expected in the MBF output due to the incoherent addition of these interference signals. However, the amount of noise reduction and echo suppression at this point is generally insufficient.

4.1.2. Blocking matrix

The role of the blocking matrix (BM) is to generate noise reference signals by blocking the desired speech signals. The BM is the conventional TF-GSC BM:

$$H(\mathrm{e}^{\mathrm{j}\omega}) = \begin{bmatrix} -\tilde{a}_2^*(\mathrm{e}^{\mathrm{j}\omega}) & -\tilde{a}_3^*(\mathrm{e}^{\mathrm{j}\omega}) & \cdots & -\tilde{a}_M^*(\mathrm{e}^{\mathrm{j}\omega}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \ddots & \\ & \cdots & & \\ 0 & 0 & \cdots & 1 \end{bmatrix}, \tag{19}$$

which is easily verified to be a proper blocking matrix. The output of this blocking matrix is then given by:

$$\begin{aligned} \boldsymbol{u}(t,\mathrm{e}^{\mathrm{j}\omega}) &= H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{z}(t,\mathrm{e}^{\mathrm{j}\omega}) \\ &= H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\left[\boldsymbol{a}(\mathrm{e}^{\mathrm{j}\omega})s(t,\mathrm{e}^{\mathrm{j}\omega}) + \boldsymbol{b}(\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}) + \boldsymbol{n}(t,\mathrm{e}^{\mathrm{j}\omega})\right] \\ &= H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\left[\boldsymbol{b}(\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}) + \boldsymbol{n}(t,\mathrm{e}^{\mathrm{j}\omega})\right], \end{aligned}$$

where the last transition is due to the desired signal blocking. Note, that the signals $\boldsymbol{u}(t,\mathrm{e}^{\mathrm{j}\omega})$ contain both echo and noise components and only the desired signal is blocked.

4.1.3. Adaptive noise canceller

The adaptive noise canceller (ANC) employs the resulting reference signals $\boldsymbol{u}(t,\mathrm{e}^{\mathrm{j}\omega})$ to reduce the noise at the output $y(t,\mathrm{e}^{\mathrm{j}\omega})$. The following, conventional BLMS is used for $m = 2,\ldots,M$:

$$\tilde{g}_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) = g_m^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega}) + \mu^{\mathrm{n}}\frac{u_m(t,\mathrm{e}^{\mathrm{j}\omega})y^*(t,\mathrm{e}^{\mathrm{j}\omega})}{p_{\mathrm{est}}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})}, \tag{20}$$

$$g_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) \overset{\mathrm{FIR}}{\leftarrow} \tilde{g}_m^{\mathrm{n}}(t+1,\mathrm{e}^{\mathrm{j}\omega}), \tag{21}$$

where $p_{\mathrm{est}}^{\mathrm{n}}$ is given by (10), and μ^{n} is the step-size of the BLMS. Since the reference noise signals $\boldsymbol{u}(t,\mathrm{e}^{\mathrm{j}\omega})$ may include echo components, the adaptation must be constrained to echo-free periods. However, this constraint can be easily met since a pure echo reference is available.

4.2. Echo module

The role of the echo module is to cancel out the echo components at the output. This is obtained by applying an M-channel echo canceller as depicted in Fig. 4. The time variations of the echo path during the convergence of beamformer is the main cause for performance degradation of the BF-AEC scheme. To mitigate these problems, we choose to copy the TF-GSC filters into the echo module, which comprises two branches: One branch compensates for the MBF variations, while the other branch compensates for the echo components leaking through the blocking matrix.

4.2.1. MBF compensation

The upper branch of the echo module compensates for the MBF variations by copying the matrix $W_0(\mathrm{e}^{\mathrm{j}\omega})$. The far-end signal $e(t,\mathrm{e}^{\mathrm{j}\omega})$ is filtered by the MBF block, copied from the TF-GSC block. Note, that all the MBF inputs are fed by the same far-end signal $e(t,\mathrm{e}^{\mathrm{j}\omega})$, yielding M distinct reference signals, $\boldsymbol{e}'(t,\mathrm{e}^{\mathrm{j}\omega}) = W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{1}e(t,\mathrm{e}^{\mathrm{j}\omega})$. These signals are fed into the echo canceller vector $\boldsymbol{g}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) = [\,g_1^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) \quad g_2^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) \quad \cdots \quad g_M^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})\,]^{\mathrm{T}}$.

The following multi-channel BLMS is used for updating $g_m^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$ for $m = 1,\ldots,M$:

$$\tilde{g}_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) = g_m^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) + \mu^{\mathrm{e}}\frac{e'_m(t,\mathrm{e}^{\mathrm{j}\omega})y^*(t,\mathrm{e}^{\mathrm{j}\omega})}{p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})}, \tag{22}$$

$$g_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}) \overset{\mathrm{FIR}}{\leftarrow} \tilde{g}_m^{\mathrm{e}}(t+1,\mathrm{e}^{\mathrm{j}\omega}), \tag{23}$$

where

$$p_{\mathrm{est}}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega}) = \eta^{\mathrm{e}}p_{\mathrm{est}}^{\mathrm{e}}(t-1,\mathrm{e}^{\mathrm{j}\omega}) + (1-\eta^{\mathrm{e}})\|\boldsymbol{e}'(t,\mathrm{e}^{\mathrm{j}\omega})\|^2. \tag{24}$$

μ^{e} is the step-size of the BLMS, and η^{e} is the power estimation forgetting factor. The adaptation should be restricted to periods where the echo signal exists, aiming at echo reduction in the output $y(t,\mathrm{e}^{\mathrm{j}\omega})$. It is important to note that, as opposed to the AEC-BF structure, the echo cancellers are adapted using the system output. In that sense, it is similar to other recently proposed structures (Herbordt et al., 2004; Kammeyer et al., 2005).

4.2.2. BM and ANC compensation

The lower branch of the echo module compensates for the echo components that leak through the blocking matrix

to the output. This branch is not adaptive, and consists of copies of the respective blocks, namely, the BM and the ANC from the TF-GSC, and the AEC from the MBF compensation branch of the echo module.

At first glance the reader may argue that this branch is not necessary for the echo cancellation task. However, the following claim should convince the reader that this extra branch can improve the convergence behavior of the algorithm. The echo component at the output of the TF-GSC is given by

$$y^{\mathrm{e}}_{\text{tf-gsc}}(t,\mathrm{e}^{\mathrm{j}\omega})\left[\boldsymbol{I}^{\mathrm{T}}W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})-(\boldsymbol{g}^{\mathrm{n}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\right]\boldsymbol{b}(t,\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}). \tag{25}$$

The echo component at the echo module is given by

$$\begin{aligned} y^{\mathrm{e}}_{\text{echomodule}}(t,\mathrm{e}^{\mathrm{j}\omega}) &= \left[(\boldsymbol{g}^{\mathrm{e}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\boldsymbol{I}-(\boldsymbol{g}^{\mathrm{n}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})(\boldsymbol{g}^{\mathrm{e}})^{*}(t,\mathrm{e}^{\mathrm{j}\omega})\right] \\ &\quad\times e(t,\mathrm{e}^{\mathrm{j}\omega}) \\ &= \left[\boldsymbol{I}^{\mathrm{T}}W_0^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})-(\boldsymbol{g}^{\mathrm{n}})^{\dagger}(t,\mathrm{e}^{\mathrm{j}\omega})H^{\dagger}(\mathrm{e}^{\mathrm{j}\omega})\right](\boldsymbol{g}^{\mathrm{e}})^{*}(t,\mathrm{e}^{\mathrm{j}\omega})e(t,\mathrm{e}^{\mathrm{j}\omega}), \end{aligned} \tag{26}$$

where the last transition is due to the diagonal property of the matrix $W_0(\mathrm{e}^{\mathrm{j}\omega})$. Now, it can be easily verified that the solution

$$(\boldsymbol{g}^{\mathrm{e}})^{*}(t,\mathrm{e}^{\mathrm{j}\omega})=\boldsymbol{b}(t,\mathrm{e}^{\mathrm{j}\omega}) \tag{27}$$

completely eliminates the echo component at the output. Due to this property, convergence of the echo cancellation filters to complicated structures can be avoided. This is a significant advantage of the proposed scheme over other joint echo cancellation and noise reduction schemes (e.g., the BF-AEC and Doclo et al., 2000). The cost is an increased computational burden, as discussed in Section 5.4.

1) Matched beamformer:
$$y_{\text{mbf}}(t,e^{j\omega})=W_0^{\dagger}(e^{j\omega})\mathbf{z}(t,e^{j\omega})$$
where $W_0(t,e^{j\omega})$ is given by (18).

2) Noise reference signals:
$$\mathbf{u}(t,e^{j\omega})=H^{\dagger}(e^{j\omega})\mathbf{z}(t,e^{j\omega})$$
$H(e^{j\omega})$ is defined in (19).

3) Noise canceller output signal:
$$y_{\text{nc}}(t,e^{j\omega})=(\mathbf{g}^{n})^{\dagger}(t,e^{j\omega})\mathbf{u}(t,e^{j\omega})$$

4) NC filters update, for $m=2,\ldots,M$:
$$\tilde{g}^{n}_{m}(t+1,e^{j\omega})=g^{n}_{m}(t,e^{j\omega})+\mu^{n}\frac{u_m(t,e^{j\omega})Y^{*}(t,e^{j\omega})}{P^{n}_{\text{est}}(t,e^{j\omega})}$$
$$g^{n}_{m}(t+1,e^{j\omega})\overset{\text{FIR}}{\longleftarrow}\tilde{g}^{n}_{m}(t+1,e^{j\omega})$$
where, $P^{n}_{\text{est}}(t,e^{j\omega})=\eta^{n}P^{n}_{\text{est}}(t-1,e^{j\omega})+(1-\eta^{n})\|\mathbf{u}(t,e^{j\omega})\|^2$

5) First echo canceller branch:
$$y'_{\text{ec}}(t,e^{j\omega})=(\mathbf{g}^{e})^{\dagger}(t,e^{j\omega})W_0^{\dagger}(t,e^{j\omega})\mathbf{1}e(t,e^{j\omega})$$

6) EC filters update, for $m=1,\ldots,M$:
$$\tilde{g}^{e}_{m}(t+1,e^{j\omega})=g^{e}_{m}(t,e^{j\omega})+\mu^{e}\frac{e'_{m}(t,e^{j\omega})y^{*}(t,e^{j\omega})}{P^{e}_{\text{est}}(t,e^{j\omega})}$$
$$g^{e}_{m}(t+1,e^{j\omega})\overset{\text{FIR}}{\longleftarrow}\tilde{g}^{e}_{m}(t+1,e^{j\omega})$$
where, $P^{e}_{est}(t,e^{j\omega})=\eta^{e}P^{e}_{\text{est}}(t-1,e^{j\omega})+(1-\eta^{e})\|(\mathbf{e}'(t,e^{j\omega})\|^2$

7) Second echo canceller branch:
$$y''_{\text{ec}}(t,e^{j\omega})=(\mathbf{g}^{n})^{\dagger}(t,e^{j\omega})H^{\dagger}(t,e^{j\omega})(\mathbf{g}^{e})^{*}(t,e^{j\omega})e(t,e^{j\omega})$$

8) Output signal:
$$y(t,e^{j\omega})=y_{\text{mbf}}(t,e^{j\omega})-y_{\text{nc}}(t,e^{j\omega})-\left[y'_{\text{ec}}(t,e^{j\omega})-y''_{\text{ec}}(t,e^{j\omega})\right]$$

9) Keep only non-aliased samples in all frequency-domain multiplications [25].

Fig. 5. Summary of the ETF-GSC.

4.3. Algorithm summary

The algorithm is summarized in Fig. 5. Note that the filters $\boldsymbol{g}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})$ and $\boldsymbol{g}^{e}(t,\mathrm{e}^{\mathrm{j}\omega})$ operate separately. They may have different lengths and different step sizes, according to the problem at hand. Moreover, $\boldsymbol{g}^{\mathrm{n}}(t,\mathrm{e}^{\mathrm{j}\omega})$ adapts during noise-only frames, while $\boldsymbol{g}^{\mathrm{e}}(t,\mathrm{e}^{\mathrm{j}\omega})$ adapts during echo-only frames. Detecting activity in the echo signal is carried out using the far-end signal $e(t)$, whereas the availability of a voice activity detector (VAD) is assumed for the desired signal. Note, that the BM $H(\mathrm{e}^{\mathrm{j}\omega})$ and the MBF $W_0(\mathrm{e}^{\mathrm{j}\omega})$ are only estimated once, and then substituted into the echo module. The components of these matrices can be updated whenever the desired speech signal is present and no double-talk is encountered.

5. Experimental study

In this section, we present a comparative experimental study of the ETF-GSC and the AEC-BF and BF-AEC cascade schemes.

5.1. Setup

The proposed algorithms were tested in a simulated room environment. The desired and echo speech signals were drawn from the TIMIT database (Garofolo, 1988), while a speech-like noise from the NOISEX-92 database (Varga and Steeneken, 1993) was used to simulate a directional stationary noise source. All three signals were filtered by simulated room impulse responses, resulting in directional signals, which are received by $M=10$ microphones. The Allen and Berkley *image method* (Allen and Berkley, 1979) was used to simulate the AIRs with reverberation time set to $T_{60}=200$ ms (see Fig. 6 for a typical impulse and frequency responses of the acoustical path). The sampling frequency was 8 kHz and the resolution was set to 16 bits per sample.

Since transfer-function ratios are used in the MBF and BM, and since typical room impulse responses can be non-minimum phase, non-casual (two-sided) FIR models were used for all filters in the cascade and ETF-GSC schemes (see respective discussion in Gannot et al., 2001). In the cascade schemes, the lengths of the AEC filters are 500 taps, the lengths of the BM and MBF filters of the TF-GSC are 181 taps, and the lengths of the interference canceller filters are 251 taps. Segments of 2048 samples are used for implementing the *overlap and save* procedure. In the ETF-GSC scheme, the lengths of the two-sided

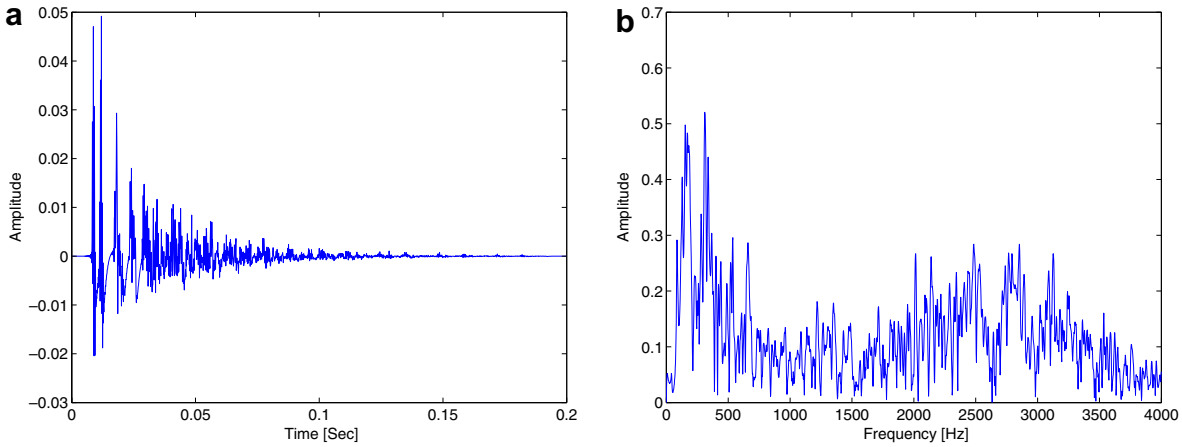

Fig. 6. Response of the first microphone to the desired source. (a) Impulse response. (b) Transfer function.

filters in the MBF and BM are 500 taps, and the lengths of the ANC filters are 1200 taps. For the echo cancellers filters, we used 300 taps on the non-casual side and 1200 taps for the casual side (recall that the AEC filters should converge to the echo acoustic impulse responses, $\boldsymbol{b}(e^{j\omega})$).

5.2. Results

Let $x(t) = x_s(t) + x_e(t) + x_n(t)$ denote one of the signals in the system, which comprises three components, namely signal, echo, and noise components. Define the signal-to-noise ratio (SNR) of the signal $x(t)$ as

$$\mathrm{SNR} = 10\log_{10}\frac{E\{x_s^2(t)\}}{E\{x_n^2(t)\}},$$

and the signal-to-echo ratio (SER) as

$$\mathrm{SER} = 10\log_{10}\frac{E\{x_s^2(t)\}}{E\{x_e^2(t)\}}.$$

The results for the AEC-BF and BF-AEC schemes, in the presence of directional noise for various input SNRs and SERs, are given in Tables 1 and 2, respectively. The two figures of merit are measured in different stages to evaluate the performance of the proposed schemes. In the cascade schemes, the SER is measured three times: at the first microphone, i.e., $x(t) = z_1(t)$, at the output of the first stage, i.e., either $x(t) = y_{bf}(t)$ or $x(t) = z_1^{aec}(t)$ for the BF-AEC and AEC-BF schemes, respectively, and at the total output, $x(t) = y(t)$. The SNR is only measured at the input and output of the cascade schemes. The measurements are taken in a time frame consisting of both echo and desired signals. The improvement in SNR and SER for both schemes, denoted as *noise reduction* and *echo suppression*, respectively, are shown in Tables 1 and 2. The SNR and SER are measured in this scheme twice: at the first microphone and at the output of the system.

Table 1
AEC-BF scheme performance

Input		Echo suppression			Noise reduction
SNR	SER	AEC	BF	Total	Total
5	5	12.8	2.8	15.6	14.6
10	5	13.1	3.1	16.20	15.2
15	5	13.2	3.3	16.5	15.1
5	10	12.1	2.8	14.9	15.5
10	10	12.8	2.9	15.7	15.9
15	10	13.1	3.0	16.1	15.8
5	15	10.2	3.2	13.5	15.7
10	15	12.0	2.9	15.0	16.1
15	15	12.8	2.8	15.6	16.0

Table 2
BF-AEC scheme performance

Input		Echo suppression			Noise reduction
SNR	SER	AEC	BF	Total	Total
5	5	5.6	5.5	11.1	13.1
10	5	6.0	5.4	11.5	13.5
15	5	6.2	5.4	11.6	13.4
5	10	5.7	4.8	10.5	14.7
10	10	6.1	4.7	10.8	15.0
15	10	6.3	4.6	11.0	14.8
5	15	5.2	4.5	9.8	15.3
10	15	5.7	4.4	10.2	15.6
15	15	6.0	4.4	10.5	15.3

We now turn to evaluating the ETF-GSC performance. Fig. 7 depicts the waveforms of one of the input signals and the respective output of the ETF-GSC algorithm in the presence of a directional noise field. The signal-to-noise ratio was set to SNR = 5 dB, and the signal-to-echo ratio was set to SER = 5 dB. The adaptive nature of the algorithm is clearly demonstrated in both noise reduction and echo cancellation behavior. The noise in these experiments was reduced by 21.5 dB, while the echo was attenuated by 16.6 dB. An audio demonstration is available in (Reuven et al., 2000).

Fig. 8 shows sonograms of signals in the ETF-GSC scheme, for the data segment shown in Fig. 7, for which convergence has already been obtained. It can be seen that (for the directional noise field), both noise and interference signals are well suppressed, especially for frequencies above

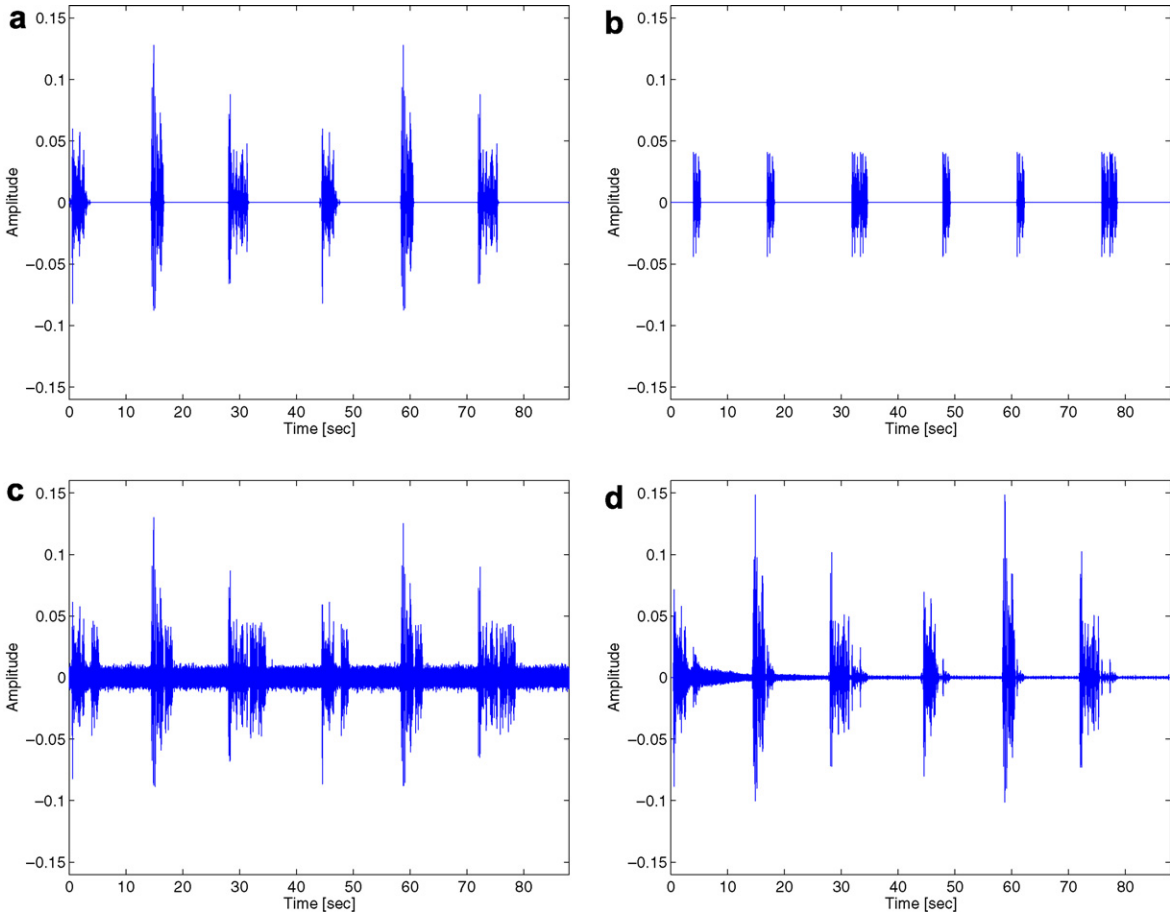

Fig. 7. Speech waveforms in the ETF-GSC scheme. (a) Desired signal. (b) Echo signal. (c) Mic. #1 signal (SNR = 5 dB, SER = 5 dB). (d) ETF-GSC enhanced signal.

500 Hz. Moreover, no self-cancellation or other frequency deviation, is evident in the desired speech signal.

In Table 3, we present the noise reduction and echo suppression for various input SNR and SER levels obtained by using the ETF-GSC algorithm. The calculations were conducted for the time segment depicted in Fig. 8, i.e., while using a directional noise signal, after convergence has been obtained. In Fig. 9 we compare the AEC-BF and the ETF-GSC output sonograms for the same data segment as presented in Fig. 8 (after convergence has been obtained).

5.3. Discussion

Comparison of the results shown in Tables 1 and 2 clearly demonstrate the advantage of the AEC-BF over the BF-AEC scheme, in both noise cancellation and echo suppression performance. This result is in accordance with the results reported by other researchers (Herbordt et al., 2004). From Table 3, we observe that noise reduction obtained by using the ETF-GSC (in the range of 21.5–22.8 dB) is significantly higher than that obtained by using the AEC-BF (14.6–16 dB) or BF-AEC (13.1–15.3 dB). The echo cancellation performance of the AEC-BF is better than that of the BF-AEC by approximately 5 dB. The ETF-GSC echo cancellation performance is further improved compared to that of the AEC-BF by approximately 1.5 dB. The advantage of the ETF-GSC over the AEC-BF in echo cancellation performance is more evident from the Sonograms in Fig. 9 and from informal listening tests (Reuven et al., 2000). We will elaborate now on these phenomena.

5.3.1. Echo suppression

For the BF-AEC scheme, when noise and echo signals are present, the TF-GSC block can eliminate both interferences due to its directivity. However, the performance of the subsequent AEC severely deteriorates due to the time variations of the echo path caused by the beamformer. This is reflected in the fact that both the TF-GSC and the AEC contribute approximately equal amount of echo suppression.

For the AEC-BF scheme, when the AEC precedes the beamformer, the degradation of the AEC performance due to the existence of noise signals is partially compensated by the subsequent beamformer. For that reason, the AEC-BF scheme generally outperforms the BF-AEC scheme when comparing the total echo suppression performance. However, whenever the noise signal becomes more

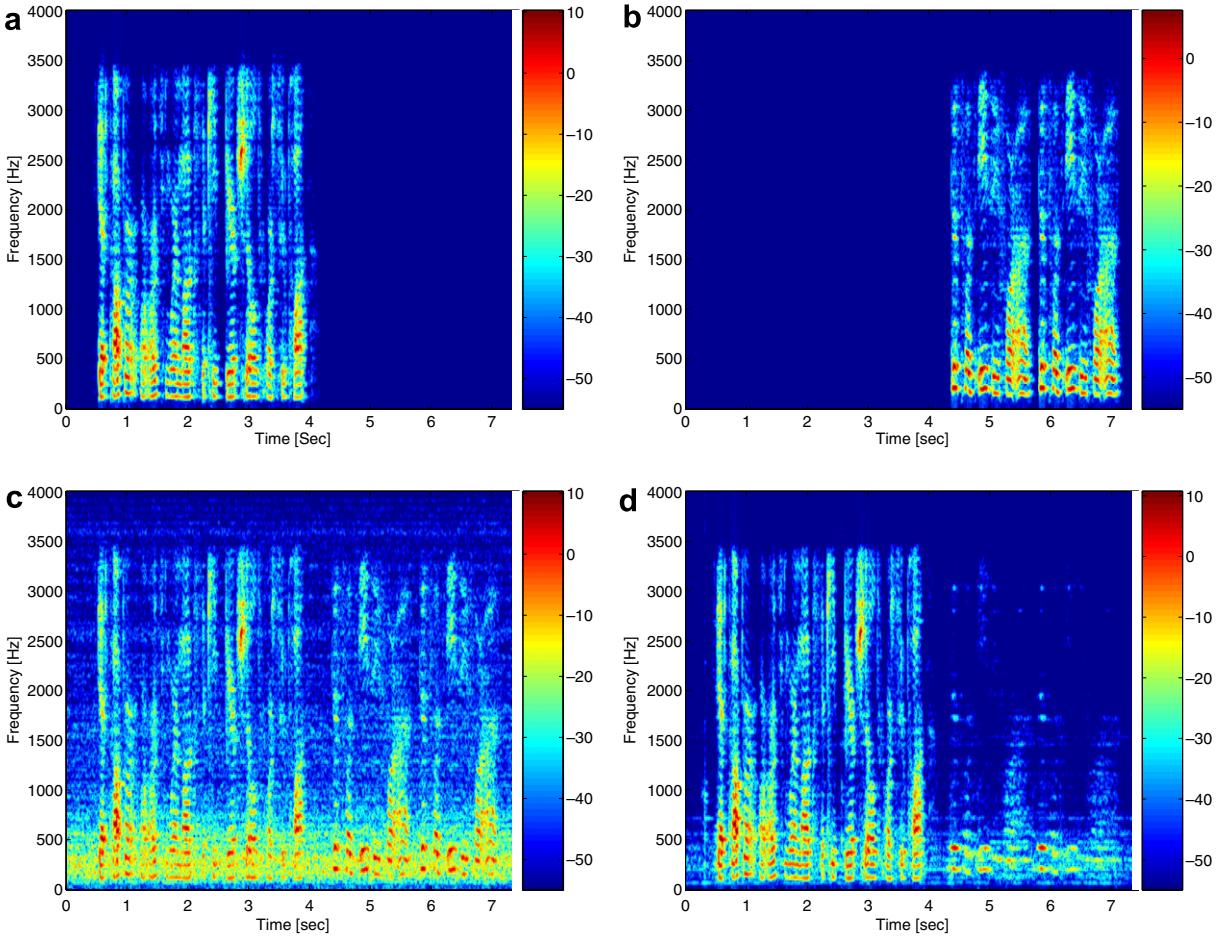

Fig. 8. Sonograms in the ETF-GSC scheme. (a) Desired signal. (b) Echo signal. (c) Mic. #1 signal (SNR = 5 dB, SER = 5 dB). (d) ETF-GSC enhanced signal.

Table 3
ETF-GSC scheme performance

Input		Echo suppression	Noise reduction
SNR	SER	Total	Total
5	5	16.6	21.5
10	5	17.3	22.3
15	5	17.7	21.6
5	10	16.2	21.6
10	10	17.1	22.6
15	10	17.3	22.4
5	15	15.4	21.7
10	15	16.7	22.8
15	15	17.1	22.8

dominant, the performance of the preceding AEC slightly deteriorates. The AEC filters, in this case, are unable to use the output signal for error correction and tracking, since they are masked by the noise.

The echo suppression performance of the ETF-GSC scheme outperforms both the AEC-BF scheme and BF-AEC scheme in all tested SNR and SER combinations. For example, the ETF-GSC scheme achieves 15.4 dB echo suppression when SNR = 5 dB and SER = 15 dB (i.e., noise is more dominant than the echo), while under the same environmental conditions, the AEC-BF and BF-AEC suppress the echo by only 13.5 dB and 9.8 dB, respectively. Although the difference in the echo suppression levels is less significant when the echo becomes stronger (as demonstrated in Table 3), it is much more evident from Fig. 9 and from the sound files (Reuven et al., 2000), obtained for SNR = 5 dB and SER = 5 dB.

The ETF-GSC is clearly advantageous over the two cascade schemes. First, the convergence of the AEC filters in the ETF-GSC scheme is not impaired due to noise presence, since the error feedback is taken from the output signal after the noise is reduced (as in Herbordt et al., 2004; Kammeyer et al., 2005). As opposed to these contributions, the AEC in the ETF-GSC structure requires estimates of the transfer-functions $\boldsymbol{b}(e^{j\omega})$, rather than a complex function thereof. Consequently, the convergence of the proposed scheme is faster than that obtainable by the other schemes. The fast convergence of the algorithm is demonstrated in Fig. 7.

5.3.2. Noise reduction

For the BF-AEC scheme, we showed that the task of echo suppression is shared by both the TF-GSC and

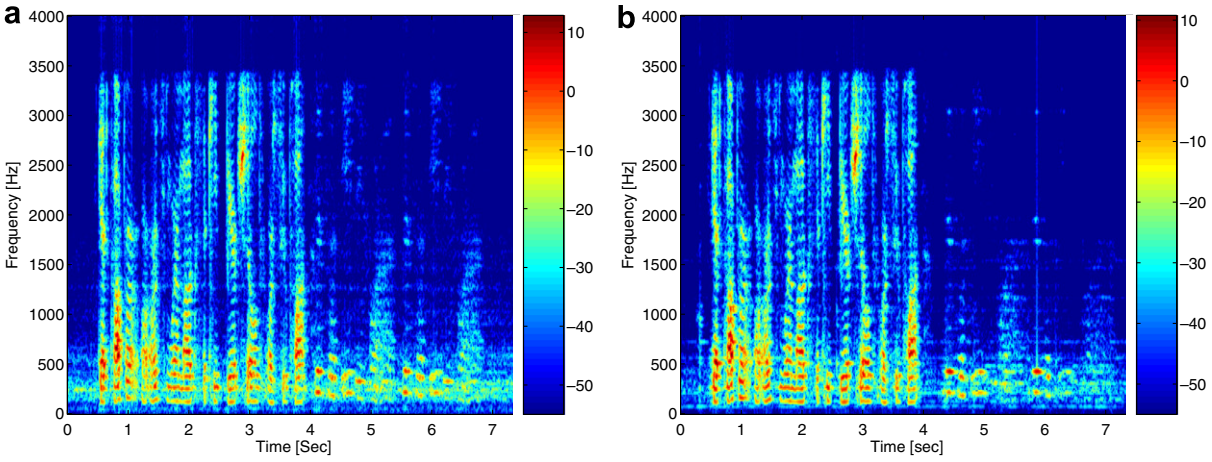

Fig. 9. Sonograms for SNR = 5 dB and SER = 5 dB after convergence. (a) AEC-BF output. (b) ETF-GSC signal.

AEC blocks. Hence, the TF-GSC block is adapted to steer two nulls, each towards one interference direction. Due to this constraint, the expected noise cancellation of the TF-GSC is expected to become limited. This limitation is manifested in the low noise reduction obtained by the BF-AEC scheme, as depicted in Table 2. Notice well, that the TF-GSC beamformer in the BF-AEC scheme may be significantly impaired due to the presence of the echo signal. The ATFs ratio estimation procedure, introduced in (Gannot et al., 2001), relies on the assumption that the desired signal is the only nonstationary component during activity of the desired source. Therefore, the expected performance of the BF-AEC scheme under double talk conditions, may significantly deteriorate as well.

Although the inputs of the TF-GSC in the AEC-BF scheme, $\boldsymbol{z}^{\mathrm{aec}}(t,\mathrm{e}^{\mathrm{j}\omega})$, are echo suppressed signals, the obtained noise reduction is limited as well. This is due to time variations imposed by the AEC on the total ATF which relates the directional noise source and the TF-GSC inputs. By contrast, in the ETF-GSC scheme both restrictions are alleviated, and the obtainable noise reduction is higher. Since the noise and echo signals are considered in parallel, the existence of the echo signal at the input has only marginal influence on the noise reduction obtained by the TF-GSC module.

5.4. Computational burden

The TF-GSC is the core beamformer for all three schemes. We use the TF-GSC rather than the regular GSC (or its variants) due to its ability to cope with complex acoustic environments. Therefore, we will not discuss here the computational burden imposed by the TF-GSC and specifically by the ATFs ratio estimation procedure. Only the additional computational complexity imposed by the echo cancellation blocks are compared.

The BF-AEC scheme requires the least computational burden, as it uses only one AEC applied to the TF-GSC output. The AEC-BF employs M AECs, which are applied directly to the microphone signals. The ETF-GSC utilizes the same number of adaptive AECs. In addition, it requires further filtering operations for the calculation of $\boldsymbol{e}'(t,\mathrm{e}^{\mathrm{j}\omega})$ and $y''_{\mathrm{ec}}(t,\mathrm{e}^{\mathrm{j}\omega})$. Filtering by MBF is required (i.e., M filters are applied) for the calculation of $\boldsymbol{e}'(t,\mathrm{e}^{\mathrm{j}\omega})$. Filtering by BM, ANCs, and AECs is required for the calculation of $y''_{\mathrm{ec}}(t,\mathrm{e}^{\mathrm{j}\omega})$. These filtering operations involve three multiplications in the frequency-domain. However, no extra Fourier transformations are required.

6. Summary

We have addressed the problem of joint echo cancellation and noise reduction in a reverberant environment, and presented a solution based on the TF-GSC. The newly developed ETF-GSC scheme is obtained by using the conventional TF-GSC in parallel with an echo module comprising copies of several TF-GSC blocks. The proposed scheme, which is adapted using the system output, decouples the noise and cancellation tasks, and thus overcomes many of the problems encountered in cascade application of the AEC and TF-GSC blocks. The proposed scheme was evaluated through a series of experiments in single-talk and double-talk situations, and compared to the AEC-BF and BF-AEC cascade schemes. It was demonstrated that the ETF-GSC scheme outperforms the cascade schemes. In particular, noise reduction using the proposed scheme is far greater, while the echo suppression is significantly higher. The enhanced performance of the ETF-GSC scheme is achieved at the expense of increased computational burden imposed by this structure.

Acknowledgement

The authors thank Phoenix Audio Technologies for providing audio equipment and for their helpful technical support. They thank E.A.P. Habets from T.U. Eindhoven for making his implementation of the image method available. They also thank Dr. Dennis R. Morgan from Bell Laboratories and the anonymous reviewers for their constructive comments and helpful suggestions.

References

Affes, S., Grenier, Y., 1997. A source subspace tracking array of microphones for double talk situations. In: Proc. 22nd IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), Munich, Germany, 20–24 April 1997, pp. 269–272.

Affes, S., Grenier, Y., 1997. A signal subspace tracking algorithm for microphone array processing of speech. IEEE Trans. Speech Audio Process. 5 (5), 425–437.

Allen, J.B., Berkley, D.A., 1979. Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Amer. 65 (4), 943–950.

Avargel, Y., Cohen, I., in press. On multiplicative transfer function approximation in the short-time Fourier transform domain. IEEE Signal Process. Lett.

Bitzer, J., Simmer, K.U., Kammeyer, K.D., 1999. Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement. In: Proc. 24th IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), Phoenix, Arizona, 15–19 March 1999, pp. 2965–2968.

Dahl, M., Claesson, I., 1999. Acoustic noise and echo canceling with microphone array. IEEE Trans. Vehicular Technol. 48 (5), 1518–1526.

Doclo, S., Moonen, M., Clippel, E.D., 2000. Combined acoustic echo and noise reduction using GSVD-based optimal filtering. In: Proc. 25th IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), Istanbul, Turkey, June 2000, pp. 1061–1064.

Frost III, O.L., 1972. An algorithm for linearly constrained adaptive array processing. Proc. IEEE 60 (8), 926–935.

Gannot, S., Burshtein, D., Weinstein, E., 2001. Signal enhancement using beamforming and nonstationarity with applications to speech. IEEE Trans. Signal Process. 49 (8), 1614–1626.

Garofolo, J.S., 1988. Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland, Tech. Rep. (prototype as of December 1988).

Griffiths, L.J., Jim, C.W., 1982. An alternative approach to linearly constrained adaptive beamforming. IEEE Trans. Antennas Propag. 30 (1), 27–34.

Herbordt, W., 2005. In: Sound Capture for Human/Machine Interfaces – Practical Aspects of Microphone Array Signal Processing, Vol. 315. Springer, Heidelberg, Germany.

Herbordt, W., Kellermann, W., 2000. GSAEC – acoustic echo cancellation embedded into the generalized sidelobe canceller. In: Proc. European Signal Processing Conference (EUSIPCO), Vol. 3, Tampere, Finland, September 2000, pp. 1843–1846.

Herbordt, W., Kellermann, W., Nakamura, S., 2004. Joint optimization of LCMV beamforming and acoustic echo cancellation. In: Proc. European Signal Processing Conference (EUSIPCO), Vienna, Austria, pp. 2003–2006.

Herbordt, W., Nakamura, S., Kellermann, W., 2005. Joint optimization of LCMV beamforming and acoustic echo cancellation for automatic speech recognition. In: Proc. 30th IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), 18–23 March 2005, Vol. 3. IEEE, Philadelphia, USA, pp. 77–80.

Hoshuyama, O., Sugiyama, A., Hirano, A., 1999. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. IEEE Trans. Signal Process. 47 (10), 2677–2684.

Jeannes, J., Scalart, P., Faucon, G., Beaugent, C., 2001. Combined noise and echo reduction in hands-free systems: a survey. IEEE Trans. Speech Audio Process. 9 (8), 808–820.

Kammeyer, K.-D., Kallinger, M., Mertins, A., 2005. New aspects of combining echo cancellers with beamformers. In: Proc. 30th IEEE Internat. Conf. on Acoustics and Speech Signal Processing (ICASSP), 18–23 March 2005, Vol. 3. IEEE, Philadelphia, USA, pp. 137–140.

Kellermann, W., 1997a. Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays. In: Proc. 22nd IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), Munich, Germany, 20–24 April 1997, pp. 219–222.

Kellermann, W., 1997b. Joint design of acoustic echo cancellation and adaptive beamforming for microphone arrays. In: Proc. 5th Internat. Workshop on Acoustic Echo and Noise Control (IWAENC), Imperial College, London, UK, pp. 81–84.

Kellermann, W., 2001. Acoustic echo cancellation for beamforming microphone arrays. In: Brandstein, M.S., Ward, D.B. (Eds.), Microphone Arrays: Signal Processing Techniques and Applications. Springer, pp. 281–306 (chapter 13).

Low, S.Y., Nordholm, S., 2005. Blind approach to joint noise and acoustic echo cancellation. In: Proc. 30th IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), 18–23 March 2005, Vol. 3. IEEE, Philadelphia, USA, pp. 69–72.

Nordholm, S., Claesson, I., Eriksson, P., 1992. The broadband Wiener solution for Griffiths–Jim beamformers. IEEE Trans. Signal Process. 40 (2), 474–478.

Reuven, G., Gannot, S., Cohen, I., 2000. Audio sample files. [Online]. http://www.eng.biu.ac.il/~gannot/examples1.html.

Reuven, G., Gannot, S., Cohen, I., 2004. Joint acoustic echo cancellation and transfer function GSC in the frequency domain. In: Proc. 23rd IEEE Convention of the Electrical and Electronic Engineers in Israel, Herzlia, Israel, 6–7 September 2004, pp. 412–415.

Reuven, G., Gannot, S., Cohen, I., 2007. Multichannel acoustic echo cancellation and noise reduction in reverberant environments using the transfer-function GSC. In: IEEE Internat. Conf. on Acoustics Speech and Signal Processing (ICASSP), to appear.

Rombouts, G., Moonen, M., 2005. An integrated approach to acoustic noise and echo cancellation. Signal Process. 85 (4), 849–871.

Shynk, J.J., 1992. Frequency-domain and multirate adaptive filtering. IEEE Signal Process. Mag. 9 (1), 14–37.

Varga, A., Steeneken, H.J.M., 1993. Assessment for automatic speech recognition: II. NOISEX-92: a database and an experiment to study the effect of additive noise on speech recognition systems. Speech Comm. 12 (3), 247–251.