MULTICHANNEL ACOUSTIC ECHO CANCELLATION AND NOISE REDUCTION IN REVERBERANT ENVIRONMENTS USING THE TRANSFER-FUNCTION GSC

Gal Reuven[1], *Sharon Gannot*[2] *and Israel Cohen*[1]

[1] Department of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa 32000, Israel
{galrv@tx,icohen@ee}.technion.ac.il

[2] School of Engineering
Bar-Ilan University
Ramat-Gan, 52900, Israel
gannot@eng.biu.ac.il

ABSTRACT

In this paper, we present a multi-channel acoustic echo canceller that is integrated into the transfer-function generalized sidelobe canceller (TF-GSC). 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. Experimental results demonstrate improved performance compared to cascade schemes of acoustic echo cancellation and adaptive beamforming.

***Index Terms*—** Echo suppression, array signal processing, adaptive signal processing, acoustic noise

1. INTRODUCTION

The transfer-function generalized sidelobe canceller (TF-GSC) [1] 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, cascade schemes are usually used to mitigate both inferences [2, 3, 4]. Unfortunately, when the acoustic echo canceller (AEC) precedes the beamformer (this scheme is denoted AEC-BF), it suffers from the noise component at its input. When the beamformer precedes the AEC (this scheme is denoted BF-AEC), the latter suffers from the time variations in the echo path, due to the beamformer convergence [5].

In this paper, we present *echo transfer function generalized sidelobe canceller* (ETF-GSC), for joint echo cancellation and noise reduction in a reverberant environment [5]. 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 the 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 scheme). On the other hand, the presence of noise does not deteriorate the performance of the echo canceller (as in the AEC-BF scheme). The proposed scheme, which is adapted using the entire system output, decouples the noise reduction and echo cancellation tasks, and hence overcoming 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.

2. PROBLEM FORMULATION

Let $s(t)$ represent the desired source signal, and let $e(t)$ represent the echo signal measured at the loudspeaker. Let $a_m(t)$ denote the acoustic impulse response (AIR) of the mth microphone to the desired source, and let $b_m(t)$ denote the AIR of the loudspeaker-enclosure-microphone (LEM) system corresponding to the mth microphone. Then, the signal received in 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. In the *short time Fourier transform* (STFT) domain, (1) can be approximately (see [6]) rewritten as

$$\mathbf{z}(t,e^{j\omega}) = \mathbf{a}(e^{j\omega})s(t,e^{j\omega}) + \mathbf{b}(e^{j\omega})e(t,e^{j\omega}) + \mathbf{n}(t,e^{j\omega}) \tag{2}$$

where

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

$z_m(t,e^{j\omega}), s(t,e^{j\omega}), e(t,e^{j\omega})$ and $n_m(t,e^{j\omega})$ are the STFT of the respective signals; $a_m(e^{j\omega})$ and $b_m(e^{j\omega})$ are the acoustical transfer-functions (ATFs) from the desired source and echo source to the mth microphone, respectively, which are assumed hereinafter time invariant over the observation period. Our problem is to reconstruct the desired speech signal $s(t, e^{j\omega})$ from the noisy observations $\mathbf{z}(t, e^{j\omega})$ and the available echo signal $e(t, e^{j\omega})$.

3. ECHO TRANSFER FUNCTION GENERALIZED SIDELOBE CANCELLER

The combined scheme of TF-GSC and AEC is depicted in Figure 1. It 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 better suits the problem at hand than the cascade schemes [5].

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. 1. The time variations of the echo path during the convergence of the beamformer is the main cause for performance degradation of the BF-AEC scheme. To mitigate this problem we choose to copy the TF-GSC filters into the echo module. Hence, the echo module comprises two branches: The upper branch compensates for the variations of the matched beamformer (MBF), $W_0(e^{j\omega})$; The lower branch compensates for the echo components leaking through the blocking matrix (BM) of the TF-GSC.

In the upper branch of the echo module, the echo signal $e(t, e^{j\omega})$ is filtered by the MBF block, copied from the TF-GSC block. Note, that all the MBF inputs are fed by the same echo signal $e(t, e^{j\omega})$ yielding M distinct reference signals, $\mathbf{e}'(t, e^{j\omega}) = W_0^\dagger(e^{j\omega})\mathbf{e}(t, e^{j\omega})$. These signals are fed into the echo canceller filters $\mathbf{g}^e(t, e^{j\omega})$.

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

$$\begin{aligned} \tilde{g}_m^e(t+1, e^{j\omega}) &= g_m^e(t, e^{j\omega}) + \mu^e \frac{e'_m(t, e^{j\omega}) y^*(t, e^{j\omega})}{P_{\text{est}}^e(t, e^{j\omega})} \\ g_m^e(t+1, e^{j\omega}) &\xleftarrow{\text{FIR}} \tilde{g}_m^e(t+1, e^{j\omega}) \end{aligned} \tag{3}$$

where

$$P_{\text{est}}^e(t, e^{j\omega}) = \eta^e P_{\text{est}}^e(t-1, e^{j\omega}) + (1-\eta^e)\|(\mathbf{e}'(t, e^{j\omega})\|^2 \tag{4}$$

and 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 $y(t, e^{j\omega})$. It is important to note that the echo cancellers are adapted using the system's output, as opposed to the AEC-BF structure. In that sense, it is similar to other recently proposed structures [7, 8].

The lower branch of the echo module compensates for the echo components which are leaking through the blocking matrix of the TF-GSC to the output. This branch is not adaptive and it consists of copies of the respective blocks, namely, the BM and the adaptive noise canceller from the TF-GSC, and the AEC from the MBF compensation branch of the echo module.

The echo component at the output of the TF-GSC is given by

$$\begin{aligned} y_{\text{tf-gsc}}^e(t, e^{j\omega}) = & \left[\mathbf{1}^T W_0^\dagger(e^{j\omega}) - (\mathbf{g}^n)^\dagger(t, e^{j\omega}) H^\dagger(e^{j\omega})\right] \\ & \times \mathbf{b}(t, e^{j\omega})\, e(t, e^{j\omega}). \end{aligned} \tag{5}$$

The echo component at the echo module is given by

$$\begin{aligned} y_{\text{echo module}}^e(t, e^{j\omega}) = & \left[\mathbf{1}^T W_0^\dagger(e^{j\omega}) - (\mathbf{g}^n)^\dagger(t, e^{j\omega}) H^\dagger(e^{j\omega})\right] \\ & \times \mathbf{g}^e(t, e^{j\omega})\, e(t, e^{j\omega}). \end{aligned} \tag{6}$$

It can be verified that the solution

$$\mathbf{g}^e(t, e^{j\omega}) = \mathbf{b}(t, e^{j\omega}) \tag{7}$$

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. [9]). The cost is an increased computational burden.

4. EXPERIMENTAL RESULTS AND DISCUSSION

The proposed scheme was tested in a simulated room environment. The desired and echo speech signals were drawn from the TIMIT database, while a speech-like noise from the NOISEX-92 database was used to simulate directional stationary noise source. The signals were filtered by simulated room impulse responses, resulting in directional signals, which are received by $M = 10$ microphones. The *image method* [10] was used to simulate the ATFs with reverberation time set to T_{60} = 200 ms. The sampling frequency was 8 KHz and the resolution was set to 16 bits per sample.

In the cascade schemes, the length of the AEC filters was set to 500 taps, the length of the BM and MBF filters of the TF-GSC was set to 181 taps, and the length of the interference canceller filters was set to 251 taps. Segments of 2048 samples were used for implementing the *overlap and save* procedure. For the ETF-GSC scheme, the lengths of the filters in the MBF and BM was set 500 taps, the length of the adaptive noise canceller (ANC) filters was set to 1200 taps. For the echo cancellers filters we used 300 taps at the non-causal side and 1200 taps for the causal side (recall that the AEC filters should converge to the echo acoustic impulse responses, $\mathbf{b}(e^{j\omega})$).

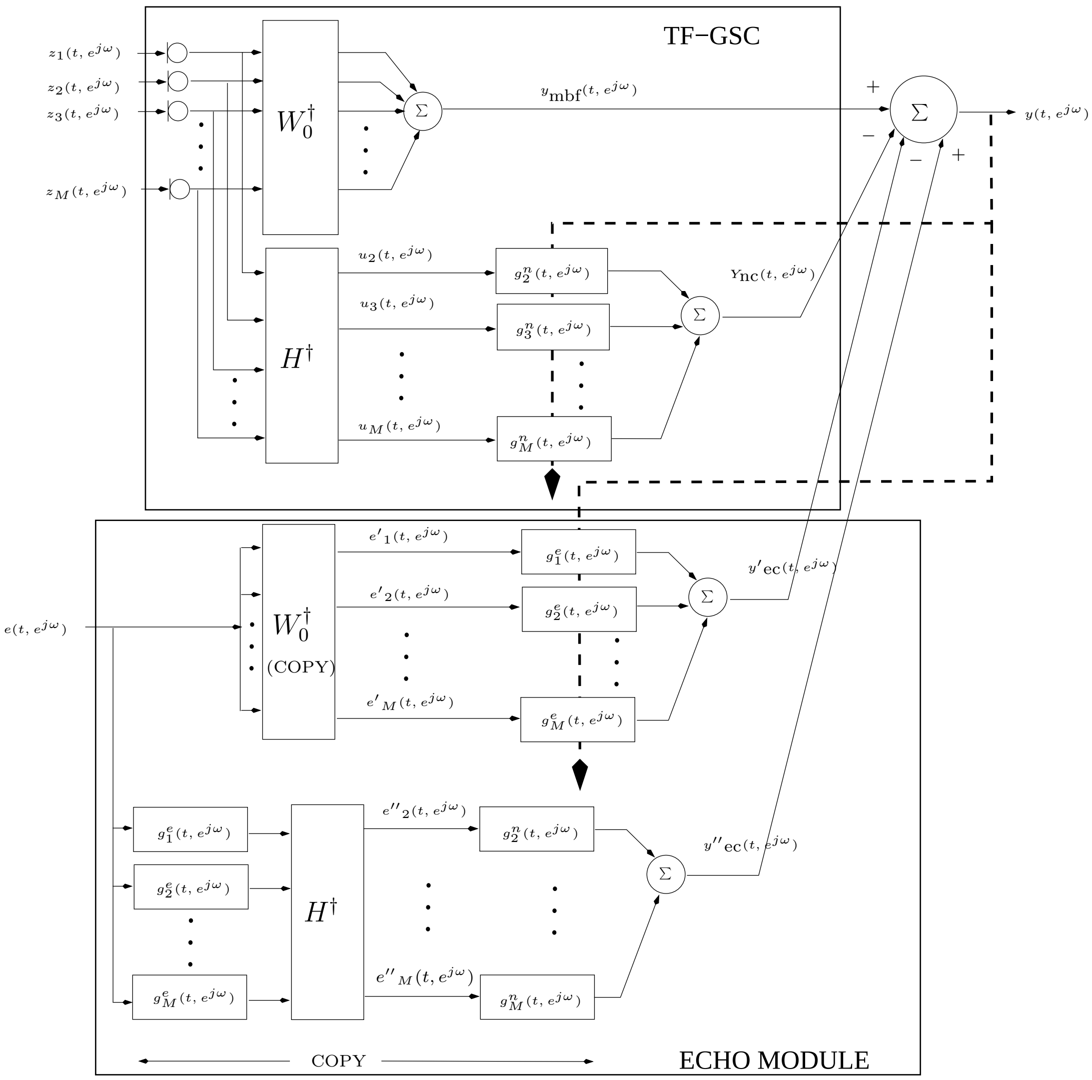

Fig. 1. Echo transfer function generalized sidelobe canceller (ETF-GSC), combining acoustic echo cancellation with the transfer-function generalized sidelobe canceller

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

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

In Table 1 we present the noise reduction and echo suppression for various input SNR and SER levels, obtained by using the ETF-GSC scheme compared to using the cascade schemes (AEC-BF and BF-AEC). The SNR and SER are measured in these schemes twice, at the first microphone signal and at the output of the system. The improvement in SNR is denoted *noise reduction*, and the improvement in SER is denoted *echo suppression*. The results were obtained while using directional noise signal, after convergence of the adaptive filters.

The results in Table 1 clearly demonstrate the advantage of the AEC-BF over the BF-AEC scheme, in both noise reduction and echo suppression performance. The echo cancellation performance of the AEC-BF is better than that of the BF-AEC by approximately 5 dB. These results are in accordance with the results presented by other researchers [7].

The noise reduction demonstrated by ETF-GSC (at the range of 21.5–22.8 dB) is significantly higher than that

Input		Echo suppression			Noise reduction		
SNR	SER	AEC-BF	BF-AEC	ETF-GSC	AEC-BF	BF-AEC	ETF-GSC
5	5	15.6	11.1	16.6	14.6	13.1	21.5
10	5	16.2	11.5	17.3	15.2	13.5	22.3
15	5	16.5	11.6	17.7	15.1	13.4	21.6
5	10	14.9	10.5	16.2	15.5	14.7	21.6
10	10	15.7	10.8	17.1	15.9	15.0	22.6
15	10	16.1	11.0	17.3	15.8	14.8	22.4
5	15	13.5	9.8	15.4	15.7	15.3	21.7
10	15	15.0	10.2	16.7	16.1	15.6	22.8
15	15	15.6	10.5	17.1	16.0	15.3	22.8

Table 1. Echo suppression and noise reduction for various input SNR and SER levels, obtained by using the ETF-GSC scheme, compared to using cascade schemes (AEC-BF and BF-AEC).

demonstrated by the AEC-BF (14.6–16 dB) or the BF-AEC (13.1-15.3 dB). The echo suppression performance of the ETF-GSC is greater than that obtained by the AEC-BF or BF-AEC in all tested SNR and SER combinations. For example, the ETF-GSC achieves 15.4 dB echo suppression when the 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 the 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 1), it is perceptually meaningful in the audio sample files [11].

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 was reduced (as in [7, 8]). As opposed to these contributions, the AEC in the ETF-GSC structure requires estimates of the transfer functions $\mathbf{b}(e^{j\omega})$, rather than a complicated function thereof. Consequently, the convergence of the proposed scheme is faster than other schemes.

5. ACKNOWLEDGEMENT

The authors thank Phoenix Audio Technologies for providing audio equipment and for their helpful technical support.

6. REFERENCES

[1] 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, Aug. 2001.

[2] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Munich, Germny, Apr. 1997, pp. 219–222.

[3] W. Kellermann, *Microphone Arrays: Signal Processing Techniques and Applications*, chapter Acoustic echo cancellation for beamforming microphone arrays, pp. 281–306, Springer, Berlin, Germany, 2001.

[4] G. Reuven, S. Gannot, and I. Cohen, "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.

[5] G. Reuven, S. Gannot, and I. Cohen, "Joint noise reduction and acoustic echo cancellation using the transfer-function generalized sidelobe canceller," to appear in Special Issue of Speech Communication on Speech Enhancement.

[6] Y. Avargel and I. Cohen, "On multiplicative transfer function approximation in the short-time Fourier transform domain," *to appear in IEEE Signal Processing Letters*.

[7] W. Herbordt, W. Kellermann, and S. Nakamura, "Joint optimization of lcmv beamforming and acoustic echo cancellation," in *Proc. of the European Signal Processing Conference (EUSIPCO)*, Vienna, Austria, Sep. 2004, pp. 2003–2006.

[8] K.-D. Kammeyer, M. Kallinger, and A. Mertins, "New aspects of combining echo cancellers with beamformers," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Philadelphia, USA, Mar. 2005, IEEE, vol. 3, pp. 137–140.

[9] S. Doclo, M. Moonen, and E.D. Clippel, "Combined acoustic echo and noise reduction using GSVD-based optimal filtering," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Istanbul, Turkey, June 2000, pp. 1061–1064.

[10] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[11] G. Reuven, S. Gannot, and I. Cohen, "Audio Sample Files," `http://www.biu.ac.il/~gannot`, Jun. 2006.