JOINT ACOUSTIC ECHO CANCELLATION AND TRANSFER FUNCTION GSC IN THE FREQUENCY DOMAIN

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

[1]Department of Electrical Engineering, Technion - IIT, Haifa 32000, Israel
[2]School of Engineering, Bar-Ilan University, Ramat-Gan, 52900, Israel

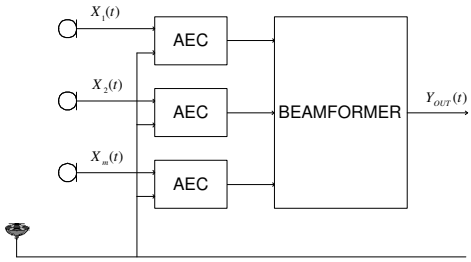
ABSTRACT

In this paper we present two joint acoustic echo and noise cancellation schemes implemented in the frequency domain and used for hands-free communication. In several past contributions equivalent time domain schemes were proposed. However, the frequency domain allows better convergence performance regardless of the condition number of the correlation matrix of the input data, and therefore is more suitable for speech processing. The first joint scheme we propose contains multi-channel acoustic echo canceller (AEC) followed by a beamformer as a second stage (this scheme is denoted AEC-BF). The second scheme contains a beamformer followed by a single channel AEC as a post-filter (denoted BF-AEC). Both schemes include the recently proposed transfer function generalized sidelobe canceller (TF-GSC) beamformer and a block-LMS AEC. The performance of both schemes is evaluated through a series of simulations, using real speech recordings in both room and car environment, and under different types of noise signals. The experimental results show that the AEC-BF scheme usually outperforms the BF-AEC scheme.

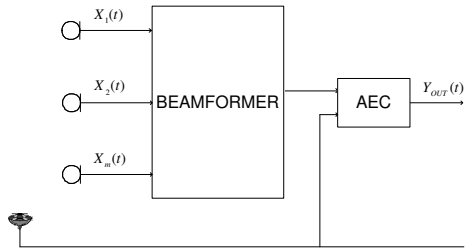
1. INTRODUCTION

Man machine interaction requires acoustic interface in order to provide full duplex hands-free communication. For better speech quality it is required to reduce both acoustic echo and noise. The acoustic echo is due to the coupling of the loudspeaker and microphone in hands-free communication. While echo signals alone can be suppressed successfully by acoustic echo canceller (AEC) and adaptive beamformer can reduce noise, the AEC performance impaired significantly due to the noise and the adaptive beamformer suffers from the echo signal. In [1],[2] Kellermann proposed two generic joint schemes, multi-channel AEC followed by a beamformer and beamformer followed by a single channel AEC. The blocks proposed in these schemes are implemented in the time domain. The time domain is known to be less applicable in real-life scenarios where complex acoustic transfer functions (ATFs) relate the source and the microphone signals. However, we propose to replace the blocks in these schemes by frequency domain blocks which are more suitable for the problem at hand.

The article is organized as follows. In Section 2, we formulate the problem. In Section 3, we introduce the proposed schemes for the joint problem. In Section 4, we describe the experimental study and in section 5 we discuss the results and show that both schemes suffer from disadvantages.



(a) AEC-BF



(b) BF-AEC

Fig. 1. the proposed schemes for joint echo cancellation and noise reduction

2. PROBLEM FORMULATION

The received signal is comprised of three components, the desired signal source, echo signal and interference signal. The m-th microphone signal is

$$x_m(t) = a_m(t)*s(t)+b_m(t)*u(t)+n_m(t)\ ;\ m = 1,\ldots,M \tag{1}$$

where $a_m(t)$ is the impulse response of the filter relating the desired speech source and the m-th microphone; $s(t)$ is the

desired signal source; $b_m(t)$ is the impulse response of the filter relating the echo speech source and the m-th microphone; $u(t)$ is the echo signal measured at the loudspeaker and received by the microphone array through the acoustic path; $n_m(t)$ is the interference signal of the m-th microphone and $*$ denotes convolution. No separate measurement of the noise signal and the desired signal are available.

The analysis frame duration T is chosen such that the source signal and echo signal may be considered stationary over the analysis frame. Although the ATFs change slowly in time, it may typically be considered time invariant over the analysis frame. Multiplying both sides of (1) by a T width window function and applying the discrete time Fourier transform (DTFT) operator yields

$$X_m(t,e^{j\omega}) \approx A_m(e^{j\omega})S(t,e^{j\omega}) + B_m(e^{j\omega})U(t,e^{j\omega}) + N_m(t,e^{j\omega})\ ;\ m=1,\ldots,M \quad (2)$$

The approximation is justified for T sufficiently large. $X_m(t,e^{j\omega}), S(t,e^{j\omega}), U(t,e^{j\omega})$ and $N_m(t,e^{j\omega})$ are the short time Fourier transforms (STFT) of the respective signals. $A_m(e^{j\omega})$ and $B_m(e^{j\omega})$ are the ATF from the local source and remote source to the m-th microphone, respectively. In vector formulation equation set (2) can be written as,

$$\mathbf{X}(t,e^{j\omega}) = \mathbf{A}(e^{j\omega})S(t,e^{j\omega}) + \mathbf{B}(e^{j\omega})U(t,e^{j\omega}) + \mathbf{N}(t,e^{j\omega}) \quad (3)$$

where

$$\mathbf{X}^T(t,e^{j\omega}) = \left[\, X_1(t,e^{j\omega})\ X_2(t,e^{j\omega})\ \cdots\ X_M(t,e^{j\omega}) \,\right]$$
$$\mathbf{A}^T(e^{j\omega}) = \left[\, A_1(e^{j\omega})\ A_2(e^{j\omega})\ \cdots\ A_M(e^{j\omega}) \,\right]$$
$$\mathbf{B}^T(e^{j\omega}) = \left[\, B_1(e^{j\omega})\ B_2(e^{j\omega})\ \cdots\ B_M(e^{j\omega}) \,\right]$$
$$\mathbf{N}^T(t,e^{j\omega}) = \left[\, N_1(t,e^{j\omega})\ N_2(t,e^{j\omega})\ \cdots\ N_M(t,e^{j\omega}) \,\right]$$

3. PROPOSED SCHEMES FOR THE JOINT PROBLEM

We propose two schemes implemented in the frequency domain, namely the AEC-BF and the BF-AEC schemes. Each of the two schemes is comprised of two components: beamformer and acoustic canceller, as depicted in Figure 1. In this section we describe each of these components in details. Note, that opposed to [2] which is implemented in the time domain, these blocks are implemented in the frequency domain.

3.1. Frequency domain implementation

The purpose of using filtering algorithms in the frequency domain is to exploit the computational advantages of performing convolutions using FFT. Furthermore, working in the frequency domain allows better convergence performance regardless of the condition number of the correlation matrix of the input data, and therefore suitable for speech. Moreover, the recently proposed TF-GSC [3], is capable of dealing with complicated ATFs, mainly since it is applied in the frequency domain.

It is well known that FIR adaptive filters can be implemented efficiently in the time domain as well as in the frequency domain by processing data in blocks rather than processing one sample at a time. The block adaptive filter has equivalent convergence properties to those of the LMS adaptive filter for stationary inputs. It has been shown in [4] that the time-domain block adaptive filter implemented in the frequency domain is equivalent to the frequency-domain adaptive filter, when data sectioning is done properly. In [4] it is proven that the overlap-save scheme requires less operations than the overlap-add scheme and therefore we use the former.

Nevertheless, when filtering is realized using multiplication in the frequency domain, aliasing effect due to cyclic convolution must be eliminated by imposing an FIR constraint. Denoted as $\xleftarrow{\text{FIR}}$, the FIR constraint includes the following three stages. First, we transform the multiplication result back to the time domain. Second, we truncate the resulting impulse response to the proper order. Third, we transform the resulting filter to the frequency domain.

3.2. Transfer function GSC

The most widely used beamforming techniques are constrained minimum power adaptive beamforming, suggested by Frost [5], and in particular its generalized sidelobe canceller (GSC) structure derived by Griffiths and Jim [6]. In these algorithms it is assumed that the received signals are simple delayed versions of the source signals. Nevertheless, in complicated acoustic environments where arbitrary ATFs relate the source signal and the microphones, the good interference suppression attained under this assumption is severely degraded. The beamformer we use is the recently proposed TF-GSC algorithm, which is a GSC solution for the arbitrary ATF case. In this algorithm the nonstationarity characteristics of the desired signal is exploited for estimating the ATFs ratios, rather than the ATFs themselves. The TF-GSC algorithm, proposed by Gannot *et al.* [3], is depicted in Figure 2 and summarized in Figure 3.

3.3. Acoustic echo canceller

The AEC receives an input signal which is comprised of the desired signal, echo signal and noise. Using the available remote speaker signal, $U(t,e^{j\omega})$, and an estimate of echo path, $C(t,e^{j\omega})$, the AEC enhances the desired signal by cancelling the echo component.

The AEC in both schemes is using the block least mean square (BLMS) algorithm. The single channel structure is summarized in Figure 4.

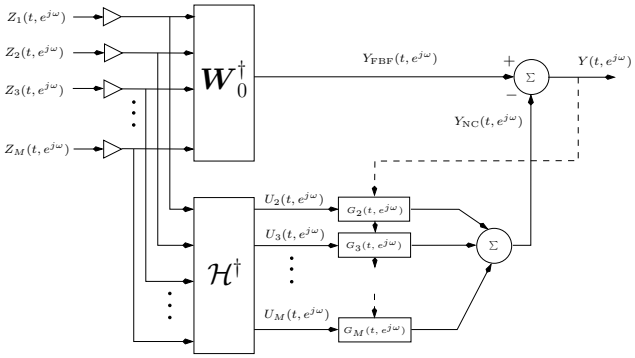


Fig. 2. GSC solution for the general ATFs case (TF-GSC). Three blocks: A fixed beamformer $\boldsymbol{W}_0^{\dagger}(t,e^{j\omega})$; A blocking matrix $\mathcal{H}^{\dagger}(e^{j\omega})$; and a multi-channel noise canceller $\boldsymbol{G}(t,e^{j\omega})$.

The estimated echo component embedded in the received signal $R(t,e^{j\omega})$ is

$$\hat{E}(t,e^{j\omega}) = U(t,e^{j\omega})C(t,e^{j\omega}) \tag{4}$$

where $U(t,e^{j\omega})$ is STFT of the remote speaker signal and $C(t,e^{j\omega})$ is the Fourier transform of the adaptive filter at frame T. The filter is updated by using the following two stages:

$$\tilde{C}(t+1,e^{j\omega}) = C(t,e^{j\omega}) - \frac{\alpha}{\|U(t,e^{j\omega})\|}\hat{D}(t,e^{j\omega})U^{*}(t,e^{j\omega})$$

$$C(t+1,e^{j\omega}) \xleftarrow{\text{FIR}} \tilde{C}(t+1,e^{j\omega}) \tag{5}$$

where α is the convergence constant, $\hat{D}(t,e^{j\omega}) = R(t,e^{j\omega}) - \hat{E}(t,e^{j\omega})$ is STFT of the estimated desired signal and $\xleftarrow{\text{FIR}}$ is imposing the FIR constraint.

4. EXPERIMENTAL STUDY

Both proposed schemes are tested in a room environment and in a car environment, which are different in size and in reflection coefficients of the surfaces. Real recordings of the desired speech signals are used, while the echo signal is synthesized using clean speech sentences drawn from the TIMIT database, and filtered by simulated ATFs [7]. Various noise sources, as will be shown in the sequel, are used for contaminating the microphone inputs. We use two-sided FIR models for all filters. In the AEC, 500 (200 for the car scenario) coefficients are used, in the blocking filters of the TF-GSC the total filter length was set to 181 and filters of the interference cancellers to 251. The sampling frequency is set to 8KHz, and the resolution to 16 bits per sample. SNR is measured at three stages in a time frame consists of both echo and desired signal: at the first microphone signal, at the output of the first stage and at the output of the second stage (total output). The improvement in SNR using the first stage (Δ_{inter}), the second stage (Δ_{out}), and the entire improvement (Δ_{total}) are given in Table 1, for both schemes in two scenarios. In the room scenario we used directional white Gaussian noise. The car scenario, where real noise recordings are used, is depicted in Table 2. Using diffused noise signal in the room environment did not change the results significantly.

1) ATFs ratios: $\mathbf{H}(e^{j\omega}) = \frac{\mathbf{A}(e^{j\omega})}{A_1(e^{j\omega})}$
2) Fixed beamformer:
$Y_{\mathrm{FBF}}(t,e^{j\omega}) = \mathbf{W}_0^{\dagger}(e^{j\omega})\mathbf{Z}(t,e^{j\omega})$
$\mathbf{W}_0(t,e^{j\omega}) = \frac{\mathbf{H}(e^{j\omega})}{\|\mathbf{H}(e^{j\omega})\|^2}\mathcal{F}(e^{j\omega})$
3) Noise reference signals:
$\mathbf{U}(t,e^{j\omega}) = \mathcal{H}^{\dagger}(e^{j\omega})\mathbf{Z}(t,e^{j\omega})$

$$\mathcal{H}(e^{j\omega}) = \begin{bmatrix} -\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} & -\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} & \dots & -\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & \dots & \ddots & \\ 0 & 0 & \dots & 1 \end{bmatrix}$$

4) Output signal:
$Y(t,e^{j\omega}) = Y_{\mathrm{FBF}}(t,e^{j\omega}) - \mathbf{G}^{\dagger}(t,e^{j\omega})\mathbf{U}(t,e^{j\omega})$
5) Filters update, for $m = 1,\dots,M-1$:
$\tilde{G}_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\mathrm{est}}(t,e^{j\omega})}$
$G_m(t+1,e^{j\omega}) \xleftarrow{\text{FIR}} \tilde{G}_m(t+1,e^{j\omega})$
where,
$P_{\mathrm{est}}(t,e^{j\omega}) =$
$\rho P_{\mathrm{est}}(t-1,e^{j\omega}) + (1-\rho)\sum_m |Z_m(t,e^{j\omega})|^2$
6) keep only non-aliased samples.

Fig. 3. TF-GSC Algorithm

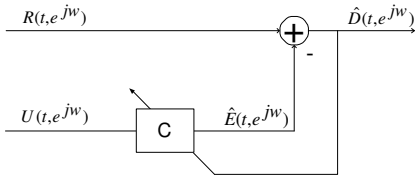


Fig. 4. Frequency domain adaptive filter structure (AEC)

5. DISCUSSION

Analysis of the obtained results shows that the AEC-BF scheme almost always outperforms the BF-AEC scheme while comparing the entire SNR improvement. This phenomenon needs further discussion. When both noise and echo are present, the TF-GSC can eliminate both due to its directivity. However, the performance of the following AEC severely deteriorates due to the variation of the

Input			AEC-BF			BF-AEC		
SNR_{noise}	SNR_{echo}	SNR_{in}	Δ_{inter}	Δ_{out}	Δ_{total}	Δ_{inter}	Δ_{out}	Δ_{total}
-4	-4	-7.2	2.5	5.7	8.2	7.8	0.2	8
8	-4	-4.4	8.2	7.9	16.2	15.2	0.3	15.5
20	-4	-4.1	10.2	7.5	17.7	17	0.4	17.4
-4	8	-4.6	0.2	5.5	5.7	5.6	0	5.6
8	8	4.8	2.4	5.2	7.6	7.1	0.1	7.2
20	8	7.6	8	-1.7	6.3	5.9	0.1	6
-4	20	-4.3	0	5.5	5.5	5.5	0	5.5
8	20	7.4	0.2	5.2	5.3	5.1	0	5.1
20	20	16.8	2	-4.5	-2.5	-2.8	0	-2.7

Table 1. Directional white Gaussian noise in room scenario

Input			AEC-BF			BF-AEC		
SNR_{noise}	SNR_{echo}	SNR_{in}	Δ_{inter}	Δ_{out}	Δ_{total}	Δ_{inter}	Δ_{out}	Δ_{total}
-4	-4	2.2	5.2	2.3	7.5	3.9	3	6.9
8	-4	3.5	12.8	1.4	14.2	7.8	2.5	10.3
20	-4	3.6	15.1	0.1	15.1	10.1	0.8	10.9
-4	8	7.1	0.6	2.2	2.8	2.1	0.7	2.8
8	8	14.2	5.2	-0.6	4.7	2.7	1.4	4.2
20	8	15.5	13.2	-9.3	3.9	2.4	0.6	3
-4	20	7.7	0	2.2	2.2	2.2	0	2.3
8	20	19.1	0.6	-0.7	-0.1	-0.5	0.3	-0.1
20	20	26.2	5.3	-12.1	-6.8	-7.1	0.2	-6.9

Table 2. Recorded noise in car scenario

echo path caused by the beamformer. On the other hand, while the AEC precede the beamformer the degradation of the AEC performance due to the existence of noise signals is partially compensated by the following beamformer. Comparison between the two tables shows that for the car noise scenario, where the noise field tends to be diffused, the obtainable performance is lower. This is due to degradation in the performance of the beamformer. It has been shown [3] that the TF-GSC performance is much better in directional noise field. We conclude the comparison between the two schemes by stressing that the computational burden imposed by the AEC-BF is higher due to the use of several AEC blocks. Note, that even though the AEC-BF scheme outperforms the BF-AEC scheme, both suffer from disadvantages and a better solution is still due to be found.

6. REFERENCES

[1] W. Kellermann, "Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Munich, Apr. 1997, pp. 219–222.

[2] W. Kellermann, "Joint Design of Acoustic Echo Cancellation and Adaptive Beamforming for Microphone Arrays," in *Int. Workshop on Acoustic Echo and Noise Control*, Imperial College, London, UK, Apr. 1997, pp. 81–84.

[3] 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.

[4] G.A. Clark, S.R. Parker and S.K. Mitra, "A Unified Approach to Time- and Frequency-Domain Realization of FIR Adaptive Digital Filters," *IEEE tran. on Acoustics, Speech and Signal Proc.*, vol. 31, no. 5, pp. 1073–1083, Oct. 1983.

[5] O.L. Frost III, "An Algorithm for Linearly Constrained Adaptive Array Processing," *Proc. of the IEEE*, vol. 60, pp. 926–935, Jan 1972.

[6] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Trans. Antennas and Prop.*, vol. AP-30, pp. 27–34, Jan 1982.

[7] 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.