A NEW STRUCTURE FOR COMBINING ECHO CANCELLATION AND BEAMFORMING IN CHANGING ACOUSTICAL ENVIRONMENTS

Trevor Burton and Rafik Goubran

Department of Systems and Computer Engineering, Carleton University, Ottawa, Canada

ABSTRACT

This paper proposes a new robust structure for combining acoustic echo cancellation and microphone array beamforming. The new structure employs short adaptive filters in the acoustic echo cancellers that precede the beamformer, allowing for quick tracking of time variations within the hands-free environment. The beamformer is followed by a single echo canceller to further suppress echo. Simulations show that by taking into consideration the dynamic behaviour of the hands-free environment, the proposed structure reconverges faster compared to standard combined structures during changing acoustical environment conditions. An improvement in ERLE of up to 10 dB is observed during reconvergence.

Index Terms— Acoustic echo cancellation, adaptive filters, array signal processing

1. INTRODUCTION

The classical approach to removing acoustical echo, caused by reverberant room environments, from a full-duplex hands-free communication system requires an acoustic echo canceller (AEC). The AEC is generally implemented via an adaptive filter which is an area of study that has been rigorously researched over many years [1]. The single-microphone approach to acoustic echo cancellation is limited in its effectiveness to suppress echo by background noise sources, non-stationarities in the acoustical environment, and system non-linearities [2]. In order to mitigate the decrease in performance of the single-microphone AEC system, multi-microphone approaches have been proposed [3]-[5]. These methods involve combining microphone array beamforming and acoustic echo cancellation techniques.

While all of the multi-microphone approaches strive to guarantee a superior hands-free conversation exists between parties, the manner in which this is achieved varies. Some techniques are implemented in the time-domain while others in the frequency-domain, with different underlying beamforming and acoustic echo cancellation approaches. In all cases there is generally a trade-off between the amount of achievable echo cancellation and complexity of the strategies used to attain it.

The optimal manner to combine the techniques of microphone array beamforming and acoustic echo cancellation in hands-free communication systems is the subject of much research. An overview of strategies for combining the two techniques is given by Kellermann in [6]. The findings of this work suggest that for a multi-microphone hands-free communication setup, a structure with an AEC per microphone input followed by a single beamformer (BF) is not practical due to its high computational complexity, especially for a large number of microphones. This structure will be referred to as AEC-BF. An opposite structure is also discussed, where a single AEC follows a microphone array BF (BF-AEC). This structure has reduced complexity, due to the single AEC, but the single AEC has to model not only the acoustic echo path but also any time-variations of the BF [6]. This becomes increasingly difficult if an adaptive BF is employed. One way to alleviate the problems introduced by an adaptive BF is to use a fixed BF instead. Another approach for combining acoustic echo cancellation and beamforming is given in [7]. Here the authors present a combined structure that is able to switch between the two basic structures mentioned above while harnessing the advantages of both. However, as more microphones are included in their structure the overall complexity may become prohibitive. A specific AEC-BF structure is presented in [8] that uses reduced length AECs while still achieving good overall echo cancellation performance. Yet, as the size of each AEC is further decreased the echo cancellation performance of their structure drops considerably.

In this paper we develop a combined AEC-BF structure for hands-free communication systems that acts as a compromise between the aforementioned two structures in terms of achievable acoustic echo cancellation, complexity, and robustness to variations in the acoustical environment.

The remainder of this paper is organized as follows. In Section 2 we describe in detail our combined AEC-BF structure. Simulation results are presented in Section 3 under non-stationary acoustical environment conditions that illustrate the robustness of our structure.

2. COMBINED AEC-BF STRUCTURE

A block diagram of our combined structure, hereinafter referred to as AEC-BF-AEC, is shown in Figure 1. The proposed structure employs an AEC per microphone input that is set to model only a small portion of each loudspeaker-room-microphone (LRM) impulse response (IR). Due to the exponentially decaying nature of a typical LRM IR, modeling only the first portion of each IR will allow for significant acoustic echo cancellation at a relatively small computational cost compared to the AEC-BF structure that models the entire echo paths. Each partially echo cancelled microphone signal is then beamformed using simple fixed delay-and-sum beamforming techniques. A final tail-end AEC operates on the beamformed signal to further remove acoustic echo by modeling the remainder of the combined echo paths. Since the tail-end AEC models only the last part of the combined echo path, the overall echo cancellation performance of this structure will be affected less by time variations of the BF compared to the BF-AEC structure that models the entire combined echo path.

Since a fluctuation in the acoustical environment will cause a change in each LRM IR, the AECs in all structures will have to

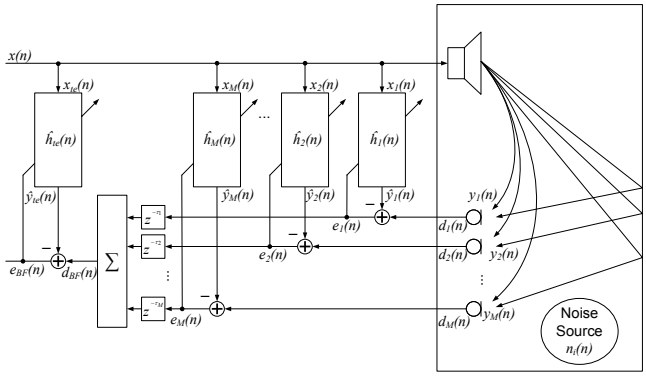

Figure 1 – Block diagram of the AEC-BF-AEC structure

track these changes thereby causing a decrease in echo cancellation performance. Since the front-end AECs of the AEC-BF-AEC structure only model the first part of each echo path, where the highest amount of variation is expected to occur, these AECs will be able to track the changes in each echo path faster due to the shorter adaptive filters used. As a result the AEC-BF-AEC structure's overall acoustic echo cancellation performance will suffer less than the previously mentioned structures.

The normalized least-mean square (NLMS) adaptive algorithm is used to implement the acoustic echo cancellation in the above structures. This algorithm was selected due to its low computational complexity, stability, and simplicity [9]. The beamforming method implemented in the above structures is fixed delay-and-sum beamforming. This beamforming method was also chosen for its simplicity and low computational complexity [10]. Also, by implementing a non-adaptive BF the tail-end AEC of our combined structure does not have to track time variations in the BF output caused by an adaptive algorithm. The basic idea behind delay-and-sum beamforming is to constructively reinforce a desired signal emitting from a specific location while attenuating interference signals. This is achieved by accounting for the delays in the acquired microphone signals to time align them, and then summing the resulting signals to produce a beamformed output.

For our combined structure, shown in Figure 1, the front-end AECs are adapted first followed by beamforming their outputs and then adaptation of the tail-end AEC. This sequential adaptation scheme is used in order to avoid adaptation conflicts that could arise if both the front and tail-end AECs were adapted simultaneously. The tail-end AEC is adapted once a time delay equal to the length of the front-end AECs has been applied to the input and echo signals. This ensures that the tail-end AEC only targets the late part of the overall echo path that is not considered by the front-end AECs. It should also be noted that adaptation of the AECs is only performed under quiet local talker conditions, and that it is assumed a double-talk detector is providing the information on whether or not a local talker is active. Also, for computational complexity reasons linear adaptive filters are used for modeling all echo paths and any nonlinear system effects are ignored. For beamforming purposes it is also assumed that the location of the local talker is known.

Under the conditions outlined above the microphone signals, $d_i(n)$, in Figure 1 are composed of the echo signals, $y_i(n)$, along with the corresponding background noise signals, $n_i(n)$, where M is the total number of microphones in the array:

$$d_i(n) = y_i(n) + n_i(n) \quad i = 1,...,M \tag{1}$$

Defining input signal, $x(n)$, and LRM IR coefficient vectors, $h_i(n)$, at time n as:

$$\boldsymbol{x}(n) = [x(n), x(n-1), ..., x(n-N+1)]^T \tag{2}$$

$$\boldsymbol{h}_i(n) = [h_i(n), h_i(n-1), ..., h_i(n-N+1)]^T \tag{3}$$

We can rewrite (1) in terms of $\boldsymbol{x}(n)$ convolved with the linear LRM IR, of length N, for microphone i as follows:

$$d_i(n) = \boldsymbol{h}_i^T(n)\boldsymbol{x}(n) + n_i(n), \quad i = 1,...,M \tag{4}$$

The error signals, $e_i(n)$, from each front-end AEC can be expressed in terms of $d_i(n)$ and the associated partial echo signal estimates, $\hat{y}_i(n)$:

$$e_i(n) = d_i(n) - \hat{y}_i(n) \quad i = 1,...,M \tag{5}$$

The partial echo signal estimates, $\hat{y}_i(n)$, are computed by each front-end AEC where the underlying adaptive filter taps, $\hat{\boldsymbol{h}}_i(n)$, provide a partial estimate of $\boldsymbol{h}_i(n)$:

$$\hat{y}_i(n) = \hat{\boldsymbol{h}}_i^T(n)\boldsymbol{x}_i(n) \quad i = 1,...,M \tag{6}$$

The length of each front-end AEC adaptive filter, $N_1 << N$, is selected to model only the first portion of the associated LRM IR where the majority of the echo path energy and time variations are expected to occur. The vectors $\hat{\boldsymbol{h}}_i(n)$ and $\boldsymbol{x}_i(n)$ are of length N_1 and defined as follows:

$$\boldsymbol{x}_i(n) = [x(n), x(n-1), ..., x(n-N_1+1)]^T \tag{7}$$

$$\hat{\boldsymbol{h}}_i(n) = [\hat{h}_i(n), \hat{h}_i(n-1), ..., \hat{h}_i(n-N_1+1)]^T \tag{8}$$

The front-end AEC adaptive filter taps are updated using the NLMS algorithm [9]:

$$\hat{\boldsymbol{h}}_i(n+1) = \hat{\boldsymbol{h}}_i(n) + \frac{\mu_i}{a + \left\|\boldsymbol{x}_i(n)\right\|^2}\boldsymbol{x}_i(n)e_i(n) \tag{9}$$

Where a is a small positive constant to help offset numerical difficulties that may occur when the value of the squared norm of the input vector is very small. The adaptation step size constant for each AEC and l^2-norm operator are denoted by u_i, and $\|\cdot\|^2$, respectively.

The output of the fixed delay-and-sum BF is determined by summing and then averaging the error signals from each front-end AEC after the appropriate delays, τ_i, have been applied. The delays to be applied depend on the configuration of the microphone array. In our case we have used a circular microphone array containing six microphones where knowledge of the propagation delays between microphones was known *a priori*. In general, the time delays between microphones can be determined using time-delay estimation (TDE) techniques. A classical method for TDE is described in [11]. The output signal from the BF, $d_{BF}(n)$, can be expressed as:

$$d_{BF}(n) = \left[\frac{1}{M}\sum_{i=1}^{M} e_i(n-\tau_i)\right] \tag{10}$$

The tail-end AEC further suppresses echo from the beamformed signal by modeling the remainder of the combined echo paths with an adaptive filter of length N_2, where $N_2 = N - N_1$. Again, the NLMS algorithm is used to update the tail-end AEC adaptive filter taps, $\hat{\boldsymbol{h}}_{te}(n)$. The output error signal, $e_{BF}(n)$, is determined analogously to $e_i(n)$ using (5) through (9) with the appropriate tail-end AEC signals from Figure 1.

3. SIMULATION RESULTS

3.1. Methodology

In order to carry out meaningful simulations of the structures compared in this paper, real world LRM IRs were experimentally determined. This allowed artificial microphone signals to be created based upon real room acoustics. The LRM IR experiments were performed in an office room at Carleton University measuring 3.8 m by 5.4 m by 2.4 m. A circular prototype hands-free terminal, consisting of six equally spaced omni-directional microphones with a top mounted speaker, was used in the experiments. A thirty second white Gaussian noise signal was used as the reference signal played through the loudspeaker with each microphone signal recorded at an 8 kHz sampling rate. Since the transfer function between a loudspeaker and each microphone of a microphone array in a room enclosure is a system identification problem, NLMS adaptive filtering was used to determine each individual LRM IR. The NLMS adaptive filters were set to adapt to a 1000 tap linear model of each LRM IR in question using a step size of 0.1. It was assumed that a 1000 tap linear model was sufficient to accurately describe each LRM transfer function.

The LRM IRs were measured under two different room configurations. The first set of transfer functions were measured with the microphone array located at an unobstructed position on a desk in the corner of the room. The second set was acquired with the microphone array located in the corner of the desk where it was in close proximity to the back and side walls of the desk as well as to overhead cabinets. Figure 2 shows the LRM IRs obtained under the above conditions for one microphone of the array. The average reverberation times (RT_{60}) for the first and second sets of LRM IRs were estimated to be 0.121 and 0.122 seconds respectively using Schroeder's method [12].

Our AEC-BF-AEC structure, shown in Figure 1, is compared to the AEC-BF and BF-AEC structures in terms of the average overall echo return loss enhancement (ERLE) from all microphones. The overall ERLE between microphone signal, $d_i(n)$, and the output of the structure in question, $e(n)$, is determined as follows:

$$ERLE(n) = 10\log_{10}\frac{E\{d_i^2(n)\}}{E\{e^2(n)\}} \qquad (11)$$

The following sections present simulation results under non-stationary conditions within the hands-free environment.

3.2. Changing BF conditions

In this section the performance of each structure is compared when the BF delay parameters change from their current values to a new set of values halfway through the simulation. This is akin to the BF switching to focus on a local talker in a different location

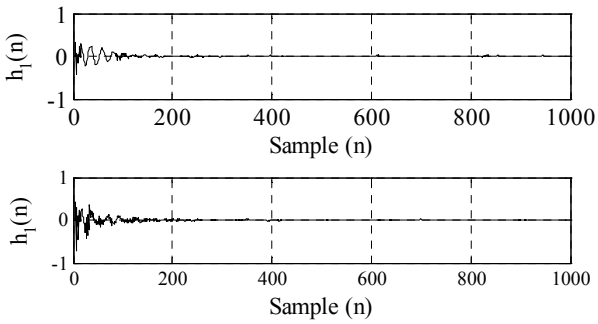

Figure 2 – Measured LRM IRs

within the hands-free environment. White Gaussian noise was used as the input reference signal, $x(n)$, in order to clearly observe the impact of the changing BF conditions. The microphone signals, $d_i(n)$, were created based on the first set of measured LRM IRs discussed in Section 3.1, with background noise added to give a SNR of 20 dB. The echo cancellers in the AEC-BF and BF-AEC structures were set to adapt to the full 1000 tap LRM IRs. While in the AEC-BF-AEC structure the front-end AECs were set to N_1 = 150 taps and the tail-end AEC was set to N_2 = 850 taps.

As shown in Figure 3, the change in the beamformers delay parameters disturbed the echo cancellers in the BF-AEC and our AEC-BF-AEC structure, resulting in a drop in ERLE performance. However, the impact on the performance of our AEC-BF-AEC structure was minimal compared to the large performance drop of the BF-AEC structure. This can be attributed to the shorter tail-end echo canceller being able to adjust to the BF changes faster than the longer AEC used in the BF-AEC structure. Our combined structure also provides faster initial convergence compared to the other structures, due to the short front-end AECs. As expected, the BF variations did not impact the AEC-BF structure since the BF operates after the echo cancellers.

Both the front-end and tail-end AEC sections of our structure provide significant echo cancellation, as shown by the top and bottom ERLE plots respectively in Figure 4. Only the tail-end AEC is slightly impacted by the BF time variations as it appears after the BF in our AEC-BF-AEC structure.

3.3. Changing LRM IR conditions

In this section the performance of each structure is compared under changing acoustical conditions within the hands-free environment. In order to simulate a change in the acoustical environment, each LRM IR is changed from the first measured set to the second measured set, discussed in Section 3.1, at the midpoint of the simulation. Again white Gaussian noise was used for the reference signal with background noise added to the microphone signals to create a 20 dB SNR. Also, the lengths of the echo cancellers were set to the same values as in Section 3.2.

As shown in Figure 5, the change in the LRM transfer functions caused a large drop in the ERLE performance of all structures. Yet, our structure was able to recover from the echo path fluctuations and converge back to steady state operation much quicker than both the AEC-BF and BF-AEC structures, due to the shorter adaptive filters used in the front-end and tail-end AECs. As well, faster initial convergence is observed. Again, this is due mainly to the short front-end echo cancellers used.

The front-end and tail-end AECs of our structure are both influenced by the echo path fluctuations as shown in Figure 6.

However, both AEC sections still provide significant echo cancellation with quicker recovery from the echo path changes than the other structures.

It should also be noted that as the length of the front-end AECs increases, and thus the length of the tail-end AEC decreases, the behaviour of the AEC-BF-AEC structure approaches that of the AEC-BF structure. Similarly, the behaviour of our structure approaches that of the BF-AEC structure as the length of the front-end AECs decreases and the tail-end AEC length increases. Simulations were performed to verify this behaviour under changing LRM IR and changing BF conditions.

4. CONCLUSIONS

A structure was presented combining microphone array beamforming and acoustic echo cancellation for hands-free communication systems. Simulations show that this new structure is able to mitigate the performance drops under changing conditions within the hands-free environment compared to the AEC-BF and BF-AEC structures, with up to a 10 dB improvement in ERLE.

5. ACKNOWLEDGMENTS

The authors would like to thank the Ontario Centres of Excellence (OCE), the Natural Sciences and Engineering Research Council of Canada (NSERC), and Carleton University for their financial support.

6. REFERENCES

[1] M. M. Sondhi and W. Kellermann, "Adaptive echo cancellation for speech signal," *Advances in Speech Signal Processing*, S. Furui and M. Sondhi, Ed. New York: Dekker, ch. 11, 1992.

[2] M. E. Knappe, and R. Goubran "Steady-state performance limitations of full-band acoustic echo cancellers," in *Proc. IEEE ICASSP*, vol. 2, 1994, pp. 73–76.

[3] W. Herbordt, and W. Kellermann, "Limits for generalized sidelobe cancellers with embedded acoustic echo cancellation," in *Proc. IEEE ICASSP*, vol. 5, 2001, pp. 3241–3244.

[4] M. Dahl and I. Claesson, "Acoustic noise and echo canceling with microphone array," *IEEE Trans. Veh. Technol.*, vol. 48, pp. 1518–1526, Sept. 1999.

[5] S. Doclo, M. Moonen, and E. de Clippel, "Combined acoustic echo and noise reduction using GSVD-based optimal filtering," in *Proc. IEEE ICASSP*, vol. 2, 2000, pp. 1061–1064.

[6] W. L. Kellermann, "Acoustic echo cancellation for beamforming microphone arrays," in *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein, and D. Ward, Eds. Berlin: Springer, 2001, pp. 281–306.

[7] K. D. Kammeyer, M. Kallinger, and A. Mertins "New aspects of combining echo cancellers with beamformers," in *Proc. IEEE ICASSP*, vol. 3, 2005, pp. 137–140.

[8] M. Kallinger, J. Bitzer, and K.-D. Kammeyer, "Study on combining multi-channel echo cancellers with beamformers," in *Proc. IEEE ICASSP*, vol. 2, 2000, pp. 797-800.

[9] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Upper Saddle River, NJ: Prentice-Hall, 1996.

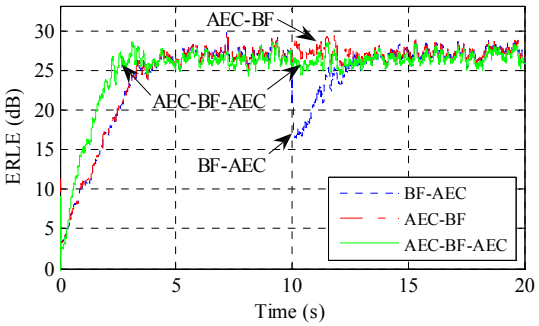

Figure 3 – ERLE performance under changing BF conditions

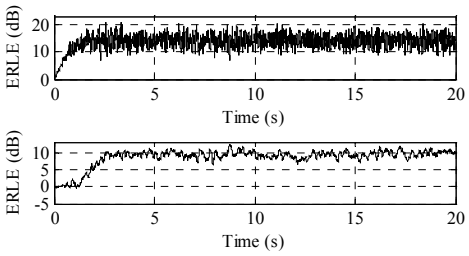

Figure 4 – AEC-BF-AEC front-end and tail-end AEC ERLE performance under changing BF conditions

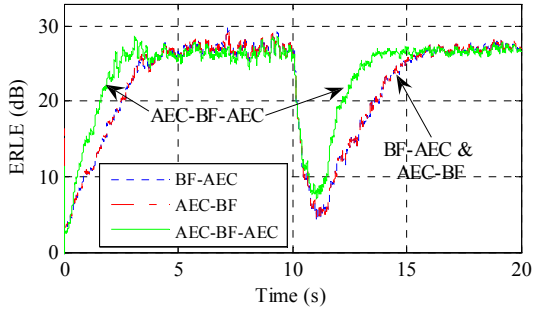

Figure 5 – ERLE performance under changing LRM IR conditions

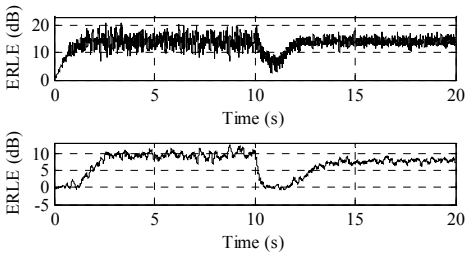

Figure 6 – AEC-BF-AEC front-end and tail-end AEC ERLE performance under changing LRM IR conditions

[10] Y. Grenier, "A microphone array for car environments," in *Proc. IEEE ICASSP*, vol. 1, 1992, pp. 305–308.

[11] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 24, pp. 320–327, Aug. 1976.

[12] M. R. Schroeder, "New method of measuring reverberation time," *J. Acoust. Soc. Am.*, vol. 37, pp. 490–412, 1965.