ACOUSTIC ECHO CANCELLATION FOR MICROPHONE ARRAYS USING SWITCHED COEFFICIENT VECTORS

Rainer Martin, Stefan Gustafsson, and *Mario Moser*
Institute of Communication Systems and Data Processing
Aachen University of Technology, 52056 Aachen, Germany
Tel: +49 241 806984; fax: +49 241 8888186, e-mail: martin@ind.rwth-aachen.de

ABSTRACT

In this contribution we propose and experimentally verify an algorithm for the cancellation of acoustic echoes for signals derived from a beamforming microphone array. Almost independent of the number of microphones this algorithm has the computational complexity of only a single echo canceller. It requires, however, additional memory to store several coefficient vectors for this echo canceller. Experimental results show that after a brief training period the new approach can outperform the computationally expensive conventional approach, where one echo canceller is used for each of the microphones.

1 INTRODUCTION

Adaptive microphone arrays and adaptive beamformers are successfully used for sound pickup in noisy and reverberant environments [1, 2, 3, 4]. It is therefore attractive to incorporate a beamforming microphone array into a hands-free telephone or teleconferencing device. Besides the environmental noise, the hands-free device must also reduce acoustic echoes which arise from the feedback of the loudspeaker signal into the microphones. Although the microphone array will attenuate acoustic echoes to some extent (depending on the directivity of the array, the position of the loudspeaker, and the position of the near end speaker) [5], an echo canceller will be necessary to achieve sufficient echo attenuation and double talk capability. It is, however, not obvious how adaptive echo cancellation can be combined with adaptive beamforming without creating an excessive computational load.

In this contribution we will explore the possibility of using just one canceller to compensate the echo component within the output signal of an adaptive beamformer. This canceller is equipped with several coefficient sets, each corresponding to a steering direction of the adaptive beamformer. To exemplify our results we will use a simple model of a two microphone delay and sum beamformer. The results can be easily generalized to other types of beamforming algorithms. We assume that the beamforming algorithm is able to track a moving speaker and can accurately determine the speaker position.

2 COMBINED BEAMFORMING AND ECHO CANCELLATION

To explain the difficulties which arise from the use of a microphone array in a hands-free telecommunication device we consider the two-microphone system shown in Fig. 1 ("conventional approach", see also [3, 6, 7]). This system combines two echo cancellers with an adaptive beamformer which steers the main beam in the direction of the near end sound source. In our simple model the beamformer is just an adaptive time delay adjustment as shown in Fig. 1 ("delay and sum beamformer"). Without loss of generality we use only integer delay values and denote the maximum delay value by N.

In Fig. 1 the microphone signals $y_1(k)$ and $y_2(k)$ are cancelled independently by echo cancellers with coefficient vectors $\mathbf{c}_1(k)$ and $\mathbf{c}_2(k)$, where k denotes the discrete time index. We thus achieve at least the same echo reduction as a single microphone hands-free device but, depending on the number of microphones, this approach might be computationally very expensive. The cancellation of the sum signal $d(k)$, as shown in Fig. 2, is not easily accomplished because any variation of the steering direction, i.e. the time delay adjustment $t(k)$, will cause sudden and significant variations of the array impulse response which has to be modelled by the single echo canceller at the array output. As a result, whenever the near end sound source moves, the canceller will have to adapt anew.

We therefore propose a new algorithm which uses only one echo canceller to cancel the sum signal $\widehat{s}(k)$ but employs $2N+1$ coefficient vectors $\mathbf{c}^{t(k)}(k)$, one for each steering direction $0 \leq t(k) \leq 2N$, to circumvent the problem of a time varying array impulse response as shown in Fig. 2. For the sake of minimal computational complexity the echo compensator will adapt only the coefficient vector belonging to the current steering direction. Whenever the delay values

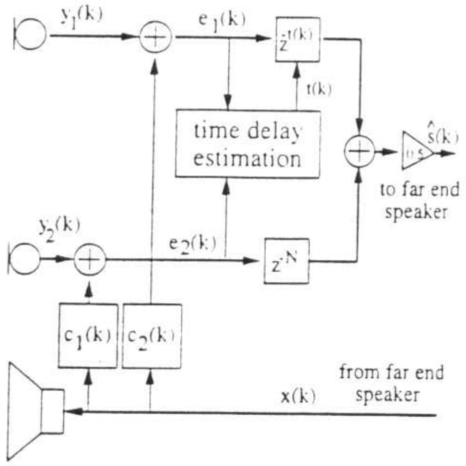

Figure 1: Conventional approach: one echo canceller for each microphone signal.

are modified to steer the main beam into another direction, the canceller switches to the coefficient vector which was previously used for this direction and continues the adaptation of this coefficient vector. In this way, the same coefficient vector is always used for the same time delay value and no coefficient vector is subject to time varying delay adjustments. As a result, we avoid the breakdown of echo attenuation whenever the steering direction changes.

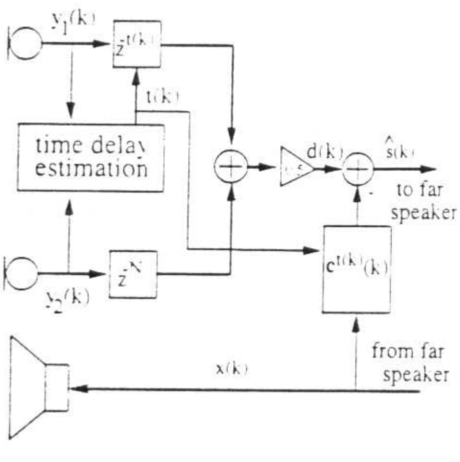

Figure 2: New approach: one echo canceller with switched coefficient vectors for the sum signal.

Our echo canceller is an NLMS adapted FIR filter with adaptive step size control as described in [8, 9]. The main advantage of this canceller is its fast convergence and high robustness against double talk and noise. In our application a number of modifications are necessary to adapt the step size control algorithm to the coefficient vector switching. A set of step size control parameters are stored for each coefficient set. When the step size for the current vector is increased (e.g. because of variations of the acoustic environment) this increase is applied to all control parameter sets. In any case, whenever the algorithm switches to a new coefficient set, the step size is increased to allow a faster initial convergence. In situations where the beamforming system is used to track just one speaker, the control unit of the beamformer thus helps the step size control to detect variations of the acoustic environment.

3 EXPERIMENTAL RESULTS

The algorithm was tested with an array of two microphones in a reverberant office room. The distance between the microphones was 40 cm, the sampling frequency 8 kHz. To facilitate the evaluation of results the speaker localization problem was separated from the echo cancellation problem. For the experiments recordings were made where the speaker positions were exactly known. In the simulation the switching between the canceller coefficient vectors was then controlled by an additional signal which indicated the speaker positions.

Fig. 3 shows the microphone positions and the nine possible speaker positions. Correspondingly, the canceller was equipped with nine coefficient sets. Near end speech was uttered from varying positions, always in the order 1-2-3-4-5-6-7-8-9-8-7-6-5-4-3-2-1-2-3-4-5-6-7-8-9.

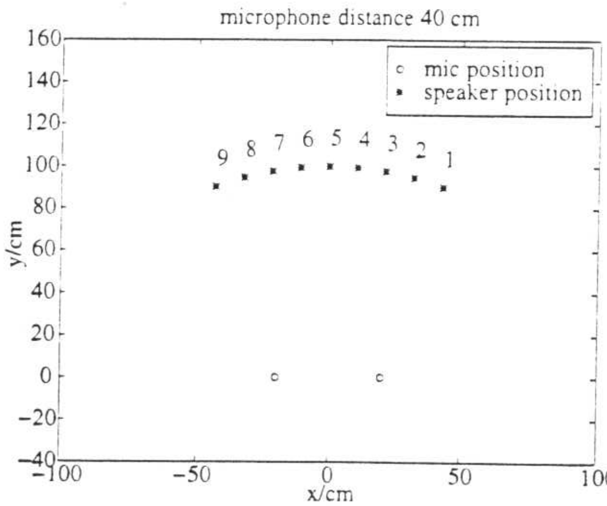

Figure 3: Microphone and near end speaker positions.

To be able to measure the system distance, the time varying room impulse responses (due to speaker movements) of both channels were measured using a technique described in [10]. Thus, the far end echo could be simulated using arbitrary speech or noise signals and either the time varying room impulse responses or, for reference purposes as described below, time invariant room impulse responses.

3.1 Experiment 1: Multiple Speakers, No Movement

In this first experiment the room impulse responses of the two microphone channels were kept constant. The sum response of the array then varies only because of time delay adjustments. For example this corresponds to a situation were several speakers sit around a conference table and do not move much. When the array is steered to the currently active speaker, the compensator switches to the corresponding coefficient set and adapts it. We assume that there are nine potential speakers placed in the positions shown in Fig. 3. Each of these positions are selected for two seconds in the order given in the previous section (i.e. 1...8-9-8...2-1-2...9) and the coefficient sets are switched accordingly. There was, however, no near end signal present (single talk experiment). The result of this experiment is shown in Fig. 4. It can be seen that after initial convergence each coefficient set is improved whenever it is selected again.

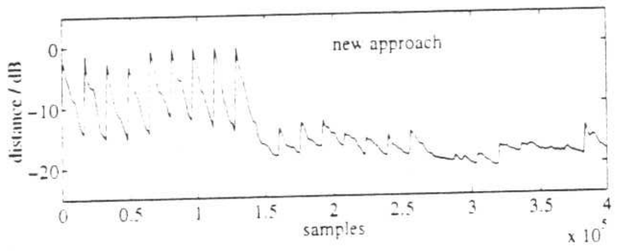

Figure 4: System Distance for Experiment 1 (time invariant acoustic environment).

3.2 Experiment 2: Single Moving Person, White Noise

We now consider the case where a single person moves to the various positions as depicted in Fig. 3. Now, the impulse responses of the microphone channels are time variant. The far end signal is a white noise signal: there is no double talk. The canceller thus adapts continuously to the time varying environment without being disturbed by near end speech. For the conventional approach (Fig. 1) and the new approach (Fig. 2) the results are shown in Fig. 5. Because of the strong movements of the person the system distance oszillates around -10dB for the white noise excitation for both approaches after the initialization phase of 18 seconds. Whenever a new coefficient set is selected during the initialization phase of $9 * 2$ seconds the system distance collapse to 0 dB for the new approach, as expected since a new coefficient vector has to be adapted. Because of the strong speaker movements the average system distance is large. The variance of the system distance is larger for the new approach than for the conventional approach. As a result the auditive impression is better for the conventional approach.

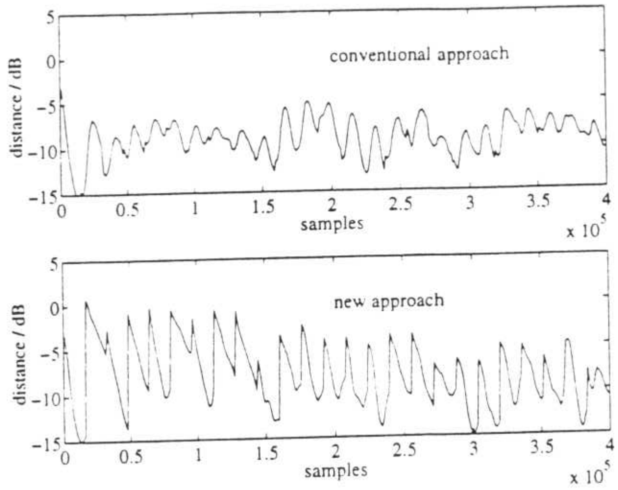

Figure 5: System distance for experiment 2: Conventional approach (upper graph) and new approach (lower graph).

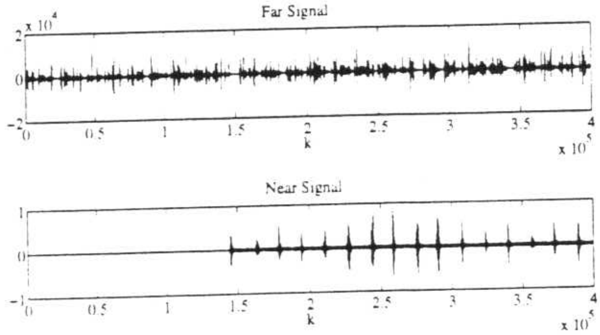

Figure 6: Far end speech (upper graph) and near end speech (lower graph) signals.

3.3 Experiment 3: Single Moving Speaker, Double Talk

In experiment 3 we consider a realistic case where a single speaker moves to the various positions as depicted in Fig. 3. The near end speaker spends two seconds in each of the nine different positions. As in experiment 2 the impulse responses of the microphone channels are time varying. During the first trip from position 1 to position 9 only the far end speech signal was present so that the echo canceller could adapt the nine coefficient sets. After that, double talk was simulated by adding the near speaker signal to the far end echo. Whenever the near end speaker moved to a new position, he uttered the position number, i.e. eight, seven, etc. The speech signals of the far end and the near end speaker are shown in Fig. 6.

The results are shown in Fig. 7. Here we find a strong degradation of the system distance for the conventional approach. This is due to the strong movements of the near end speaker and the dou-

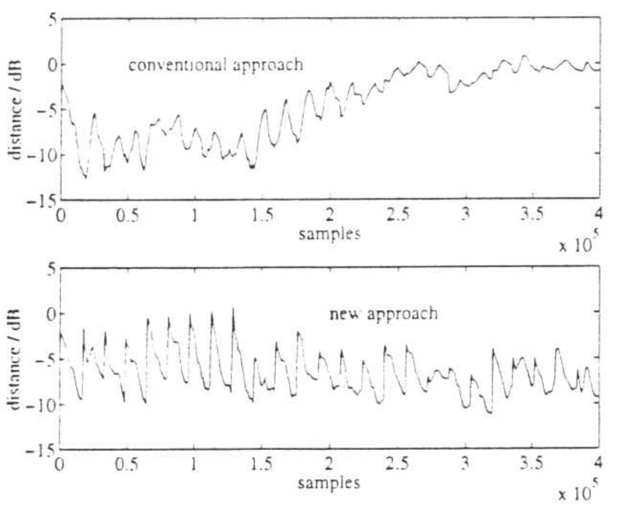

Figure 7: System distance for experiment 3: Conventional approach (upper graph) and new approach (lower graph).

ble talk. Since the step size of the canceller is kept low the canceller cannot follow the variations of the acoustic environment. Remarkably, the results are better for the new approach. Since the impulse responses and hence the coefficient sets corresponding to the various near end speaker positions do not change much over time, the new approach is better suited to cope with the time variant environment and double talk. This is essentially a consequence of the direct coupling between the speaker movements, as detected by the time delay estimator, and the variations of the acoustic environment.

4 DISCUSSION AND CONCLUSIONS

It has been shown in this paper that the echo canceller with switched coefficient vectors can be used to compensate the echo in the output signal of a beamforming microphone array. It requires, however, an increased amount of memory (RAM) to store the coefficient sets corresponding to the various near end speaker locations. It is obvious that when the speaker location must be determined with a high resolution and the space segment in which the speaker may move is large, the necessary amount of RAM becomes prohibitive. Also, in this case the speaker may spend a relatively short time in the various locations so that the time to train the echo canceller coefficient sets becomes rather short. The method is therefore best suited for hands-free situations where the near end speaker can move only in a restricted area, preferable only in one dimension (e.g. in a car or at a conference table). In the conference situation with multiple speakers and only little variations of the acoustic environment the results can be improved by adapting more than one coefficient set at the same time (provided that the additional computational load can be handled). In the situation with only one moving speaker and significant variations of the acoustic environment it turned out that adapting only the coefficient set corresponding to the speaker location and freezing the others outperforms the approach of adapting all coefficient sets at the same time.

5 ACKNOWLEDGMENTS

This research project was funded by the German Telekom AG, Berlin. The authors thank Dr. H. Schütze, Mr. E. Diedrich, and Dr. R. Zelinski for their support.

References

[1] J.B. Allen, D.A. Berkley, and J. Blauert, "Multimicrophone Signal-Processing Technique to Remove Room Reverberation from Speech Signals", J. Acoust. Soc. Am., Vol. 62, No. 4, pp. 912-915, October 1977.

[2] R. Zelinski, "Noise Reduction Based on Microphone Array with LMS Adaptive Post-Filtering", Elect. Lett., Vol. 26, No. 24, pp. 2036-2037, November 1990.

[3] R. Martin and P. Vary, "Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction: A Two Microphone Approach", Annales des Télécommunications, Vol. 49, No. 7-8, pp. 429-438, 1994.

[4] M. Dörbecker and S. Ernst, "Combination of Two-Channel Spectral Subtraction and Adaptive Wiener Post-Filtering for Noise Reduction and Dereverberation", Proc. *EUSIPCO-96, 1996*, Triest, 1996.

[5] W. Kellermann, "Some Properties of Echo Path Impulse Responses of Microphone Arrays and Consequences for Acoustic Echo Cancellation", Proc. Fourth Int. Workshop on Acoustic Echo and Noise Control, pp. 39-43, Røros, Norway, June 1995.

[6] R. Martin and P. Vary, "Combined Acoustic Echo Cancellation, Dereverberation, and Noise Reduction: A Two Microphone Approach", Proc. Third Int. Workshop on Acoustic Echo Control, pp. 125-132, Lannion, September 7-8, 1993.

[7] W. Kellermann, "Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays", Proc. Int. Conf. Acoustics, Speech, Signal Processing '97, pp. 219-222, Munich, 1997.

[8] R. Frenzel, "Freisprechen in gestörter Umgebung", Fortschritt-Berichte VDI, Reihe 10, VDI-Verlag, Düsseldorf, 1992.

[9] C. Antweiler, "Orthogonalizing Algorithms for Digital Compensation of Acoustic Echoes", Dissertation (in German), Aachen University of Technology, April 1995.

[10] C. Antweiler and H.-G. Symanzik, "Simulation of Time-Variant Room Impulse Responses", Proc. Int. Conf. Acoust., Speech, Signal Processing, pp.3031-3034, 1997.