STUDY ON COMBINING MULTI-CHANNEL ECHO CANCELLERS WITH BEAMFORMERS

Markus Kallinger, Joerg Bitzer, Karl-Dirk Kammeyer

University of Bremen, FB–1, Dept. of Telecommunications
P.O. Box 330 440, D–28334 Bremen, Germany
email: kallinger@comm.uni-bremen.de

ABSTRACT

In this paper, we will compare different combinations of a multi-channel non-adaptive noise reduction unit (NRU) and an acoustic echo cancellation unit (AEC) for a standard single-channel voice transmission. The results show that the NRU and the AEC-unit can be interchanged without increasing the computational complexity of the combined system. The length of an AEC's adaptive filter can be greatly reduced, if a multi-channel AEC-unit in front of a multi-channel NRU is used. Furthermore, the control information, like the step-size of the NLMS-algorithm, has to be computed only once.

1. INTRODUCTION

In a hands-free communication system we need two different devices to ensure a high quality of speech transmission. A noise reduction unit (NRU) is needed, especially in environments with strong background noise, e.g. in a car. The other device is the acoustic echo canceller, which compensates the far-end speaker's signal.

Several authors (e.g. [1, 2, 3]) have examined the combination of the two devices for the single-channel case using one microphone. Their results indicate that the AEC has to come first, as the time-varying noise reduction filter would disturb the adaption of the AEC. However, it is possible to switch this arrangement, if the AEC was adjusted to the known noise reduction filter. But this procedure will involve high computational complexity.

The combination of multi-channel NRUs and AECs and their possible interaction was published by Kellermann [4] as a theoretical overview. The setup of a multi-channel AEC-unit preceding the NRU was discussed and rejected, since the calculation power needed for one AEC is multiplied by the number of microphones.

Martin [5] suggested a combined system with two microphones where the AEC- and the NRU can be switched. He found out that the length of the AEC's compensation filters can be reduced to the half, if the two AECs are placed in front of the NRU. In addition to a NR post-filter, he uses a simple delay-and-sum beamformer. We will show that his suggestion can work well with a four-channel setup, too. Here, the length of the adaptive filters after each microphone can be reduced even more.

In section 2 we introduce our combined system. Besides the mentioned reduction of the filter length in a multi-channel AEC-unit, we will show more possibilities to make the usage of multi-channel echo cancellation more convenient. Section 3 shows the simulation environment, followed by the discussion of our results in section 4.

2. COMBINED SYSTEM

Figure 1 and figure 2 show the two examined arrangements of a multi-channel AEC- and NR-unit. As the focus of this

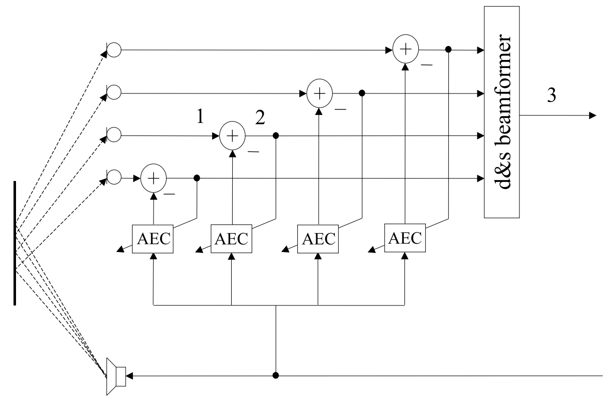

Figure 1: A unit of four AECs (one for each microphone channel) is followed by the multi-channels NRU (**setup** 1).

paper lies on the multi-channel AEC-unit, we have chosen a simple delay-and-sum beamformer as the NRU. Since we need fractional delays for the delay-and-sum beamformer, we implemented the NRU in the frequency domain. For the AEC-unit we tested two different algorithms to adapt the filters in the AEC to the according loudspeaker room microphone (LRM) impulse response. A simple time-domain

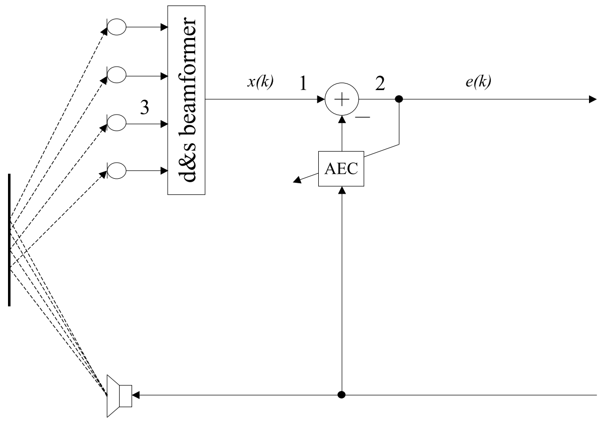

Figure 2: One AEC is positioned at the output of the multi-channels NRU (**setup** 2).

NLMS-algorithm works fine for a first experiment using white noise (see 3.1). For the following simulations with speech signals an APA-algorithm [6] with a projection order of 10 (APA10) was used.

In a real-world environment with double-talk situations a reliable control of the AEC is needed [7]. As an example the step-size of the NLMS is one key-parameter for a sufficient adjustment. We have chosen the suggestion by Antweiler [8] for the NLMS-algorithm.

3. SIMULATION

First, we want to illustrate the spatial geometry of our simulated experimental setup. Figure 3 gives a survey over the position of the sources and sensors. We obtained the twelve needed LRM-impulse responses (each source to each sensor) using the image method by Allen and Berkley [9] implemented in the frequency domain in order to get fractional delays. The reverberation time is $\tau_{60} = 200ms$. Each LRM-impulse response has got a length of 4096 taps.

3.1. Studies with White Noise

In a first experiment we simulated the near-end and the far-end speaker as white noise sources, uncorrelated to each other. This will show an estimation of the behaviour of the two setups. We used the linear array in endfire steering, since the results using broadside steering did not show distinct differences. We measured the echo return loss enhancement (ERLE) for each AEC between points 1 and 2. The ERLE of the whole system can be obtained using the points 1 and 3 (see figure 1 for setup 1, for setup 2 (fig. 2) use points 3 and 2). The ERLE is given by

$$\mathrm{ERLE}_{dB} = 10\log_{10}\frac{\mathrm{E}\{x^2(k)\}}{\mathrm{E}\{e^2(k)\}} \tag{1}$$

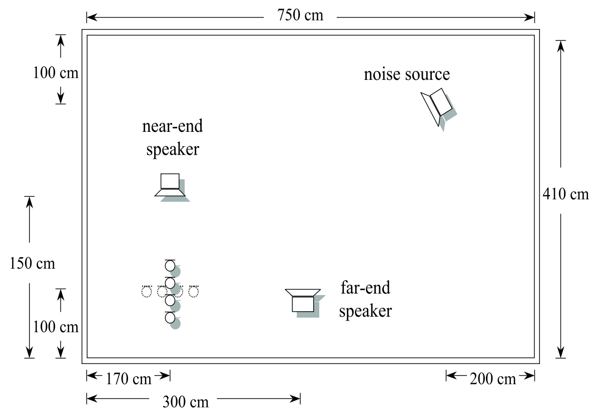

Figure 3: Placement of signal sources and sensors in the simulated room. The linear arrangement of the sensors are shown in both endfire and broadside steering to the near-end speaker. The distance of adjacent sensors is $5cm$.

and $\mathrm{E}\{\cdot\}$ represents the expectation operator.

3.2. Studies with speech signals

The second experiment was done using speech sources for far-end and near-end speaker as well as white noise for the background noise. For all simulations, we chose a far-end speech signal to white noise ratio of $20dB$. We still used the endfire steering of the microphone array. Figure 4 shows the two speech signals. The SNR of the near-end speaker to the white noise was set to $23dB$. Therefore the near-end speaker is $3dB$ louder than the far-end speaker, which represents a realistic double-talk situation.

4. RESULTS

Figure 5 shows the maximum ERLE for a given SNR. The SNR is defined as the ratio between the near-end and the far-end speaker. Both sources were white noise. Therefore, a negative SNR means a stronger far-end speaker. Figure 5a shows the maximum ERLE at each AEC, when we use the same filter length for both setup 1 and 2. We can see that setup 2 always leads to a worse ERLE. This behaviour can be explained by the additional reduction of the far-end speaker caused by the preceding NRU. In setup 1 the adaption in each AEC is improved by the better (i.e. smaller) near-end to far-end signal ratio. Increasing the number of adaptive coefficients from 1024 to 2048 will not increase the ERLE significantly. This result holds for both setups. Even at a SNR of $-20dB$ the disturbing near-end signal is stronger than the residual error caused by the filter length being too short (note that we use LRM-impulse responses of a length of 4096 taps). A theoretical explanation and an estimation of the resulting error, if the number of coefficients

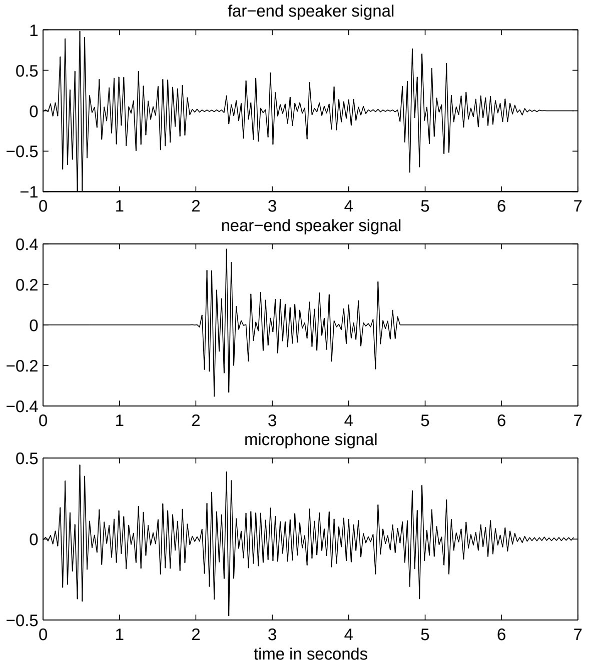

Figure 4: The upper plot shows the far-end speaker signal, the middle plot the near-end speaker signal. In the lower plot we can see a complete microphone signal, which includes the reverberated far-end speaker, the near-end speaker and the background noise.

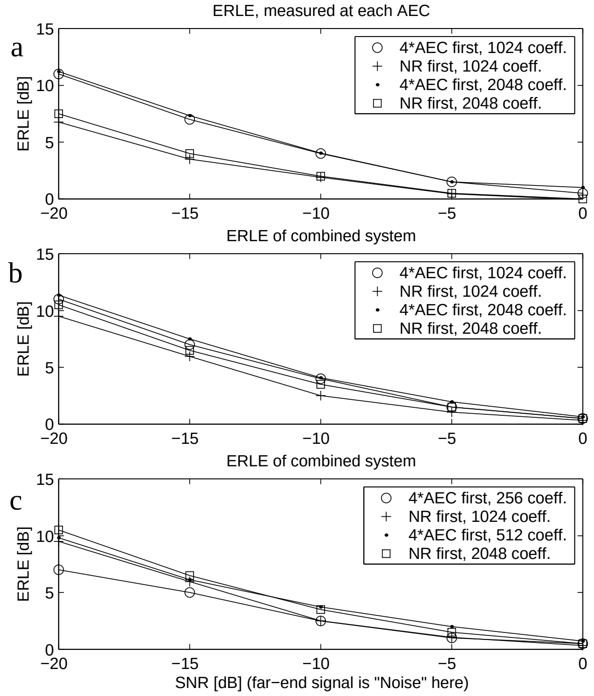

Figure 5: Maximum ERLE measured at different SNR-values using white noise.

is too small can be found in [10].

In figure 5b we examined the ERLE of the whole combined system. The differences between setup 1 and setup 2 become smaller, as the NRU additionally suppresses the far-end signal. We can still see a slight advantage of setup 1.

In figure 5c we have reduced the length of each AEC in setup 1 to a quarter compared to the single AEC in setup 2. Now, there are no distinct differences between the two setups anymore, except for one measurement using setup 1 with a length of 256 for each AEC. Here, the residual error is to high, especially at SNRs between $-20dB$ and $-10dB$. For setup 1 using a filter length of 512 we can see that the filter length can be greatly reduced without decreasing the ERLE.

Finally, we examined the behaviour with speech signals. Figure 6 shows the ERLE measures, when we use the NLMS- or the APA10-algorithm. Again, there are no explicit differences between setup 1 and setup 2. We can see, that the four APA10 AECs (with a length of 512 each) in setup 1 converge even faster than the single AEC in setup 2.

4.1. Step-size control

In order to show that the computational complexity is not increased by the usage of four AECs we examined the lapse of the step-size control parameter $\tilde{\alpha}(k)$ at all four AECs (see figure 7). No distinct differences can be seen, and therefore, only one estimator for the step-size can be used to control the whole AEC-unit. We recommend to use one of the middle channels of the linear array to get the best results.

5. CONCLUSION

In this contribution we have presented a study on multi-channel AECs and NRUs and their interaction when combining them in different setups. Our results indicate that a preceding multi-channel AEC will not increase the computational complexity as the filter lengths can be reduced and one control unit is sufficient. Therefore, both setups are comparable when non-adaptive multi-channel NRUs were used. However, combining AECs with adaptive NRUs a preceding AEC-unit is preferable, since a preceding adaptive multi-channel NRU will disturb the AEC-unit, as in the single-channel case.

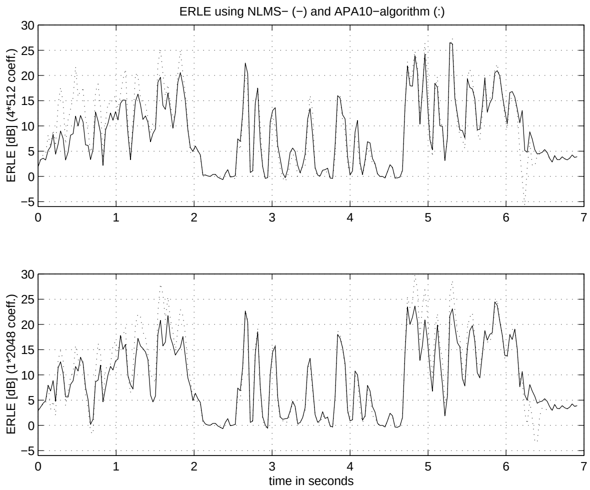

Figure 6: ERLE of whole system with AEC-unit and delay and sum beamformer. Filter length of 512 in each AEC in **setup** 1, 2048 in **setup** 2.

Acknowledgement

The authors would like to thank Y. El-Hajame for his first studies on the theme of this paper.

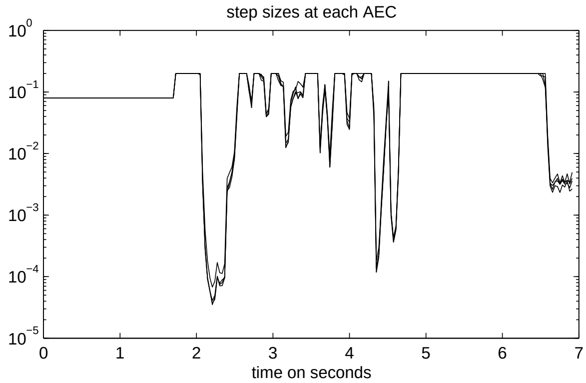

Figure 7: Lapses of the step-sizes $\tilde{\alpha}(k)$ at each AEC for the double talk situation given in figure 4. Parameters: **setup** 1; array in endfire steering; each AEC's length is 512; SNR= $20dB$

6. REFERENCES

[1] S. Gustafsson and R. Martin, "Combined Acoustic Echo Control and Noise Reduction Based on Residual Echo Estimation," in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, (London, Great Britain), pp. 160–163, Sep 1997.

[2] R. Martin and P. Vary, "Combined Acoustic Echo Cancellation, Dereverberation and Noise Reduction: A Two Channel Approach," in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, (Lannion, France), pp. 429–438, Sep. 1993.

[3] P. Scalart and A. Benamar, "A System for Speech Enhancement in the Context of Hands–Free radiotelephony with Combined Noise Reduction and Acoustic Echo Cancellation," *Speech Communication*, vol. 20, pp. 203–214, Dec 1996.

[4] W. Kellermann, "Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, vol. 1, (Munich, Germany), pp. 219–222, April 1997.

[5] R. Martin, *Hands-free Telephones based on Multi-Microphone Echo Cancellation and Noise Reduction*. PhD thesis, Aachen University of Technology, Verlag der Augustinus Buchhandlung, June 1995. Aachener Beiträge zu digitalen Nachrichtensystemen, Band 3 (in German language).

[6] K. Ozeki and T. Umeda, "An Adaptive Filtering Algorithm Using Orthonormal Projection to an Affine Subspace and it's Properties," *Electronic and Communications in Japan*, vol. 67-A, pp. 126–132, Feb 1984.

[7] E. Haensler, "From Algorithms to Systems – It's a Rocky Road," in *Proc. Int. Workshop on Acoustic Echo and Noise Control*, pp. K1–K8, Sep 1997.

[8] C. Antweiler, J. Grunwald, and H. Quack, "Approximation of Optimal Step Size Control for Acoustic Echo Cancellation," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Proc. (ICASSP)*, vol. 1, (Munich, Germany), pp. 295–298, April 1997.

[9] J. B. Allen and D. A. Berkley, "Image Method for Efficiently Simulating Small–Room Acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, 1979.

[10] R. Poltmann, "Stochastic gradient algorithm for system identification using adaptive FIR-filters with too low number of coefficients," *IEEE Trans. on Circuits and Systems*, vol. 35, pp. 247–250, Feb 1988.