Combined Acoustic Echo Control and Noise Reduction for Hands-Free Telephony - State of the Art and Perspectives

Rainer Martin and *Peter Vary*
IND, 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 paper we summarize and discuss recent results in acoustic echo cancellation and noise reduction with emphasis on methods which combine both aspects. It is shown that echo control and noise reduction can support each other in a true synergy. The paper discusses fundamental issues of algorithm design and suggests that a frequency domain multi-microphone solution might be best suited to achieve the desired performance.

1 Introduction

With the advent of mobile communication and multimedia systems hands-free telephony and hands-free voice input systems (e.g. for automatic speech recognition) are in high demand.

The realization of a hands-free communication system requires solutions to two fundamental problems. First, due to the feedback of the far end speech signal via the loudspeaker, the room, and the microphone (the "LRM system") an echo control device is necessary to guarantee the stability of the electro-acoustic loop and to supply sufficient echo reduction. We would like to emphasize that the stability of the electro-acoustic loop is a physical condition (more precisely a control problem) while the echo reduction is concerned with the audibility of the disturbing echo and is as such linked with psychoacoustics.

The second problem to be solved is the reduction of noise which becomes necessary due to the relative large distance from the microphone of the hands-free terminal to the mouth of the speaker. Especially, in mobile communication environments the signal-to-noise ratios can be very low. Quality and intelligibility improvements of the noisy speech are thus highly desirable.

The above problems were addressed independently for many years (see e.g. [1, 2] and [3, 4] for reviews of these methods). Although the algorithms which were developed are computationally demanding, they do not always deliver the desired performance. In the last years it has been recognized, however, that the echo control and noise reduction problem can be tackled in a combined approach [5, 6, 7, 8, 9, 10, 11]. It has been shown that the combined treatment yields algorithms which deliver better performance at less computational costs than systems based on separate algorithms [8, 12]. It is the objective of this paper to summarize some of the fundamental considerations underlying the combined treatment and to discuss algorithms which make use of this knowledge.

The remainder of this paper is organized as follows: In the next Section we will discuss fundamental considerations concerning the design of combined systems as well as practical realizations as they are presented in the literature. In Sections 3 and 4 we will outline limitations inherent in current systems and our perspectives with respect to successful algorithms.

2 Combined Systems

In a hands-free telephone the far end signal supplies a reference which is highly correlated with the disturbing echo at the microphone input. Therefore, the echo reduction problem is in principal ideally solved by means of an echo canceller [13]. For this reason combined echo and noise reduction systems almost always include an echo canceller. Beside the canceller, combined systems utilize some additional filter or device to reduce the residual echo which arises from insufficient canceller convergence or insufficient compensator order. The objective of this section is to show how this additional filter might be adapted to reduce echoes as well as additive ambient noise.

2.1 Fundamental Design Issues

2.1.1 Processing Order

To combine acoustic echo cancellation with echo and noise reduction it must be asked in which order these two processing operations should be performed. Figures 1 A and B depict two principal cases: the configuration EC/ENR where the acoustic echo cancellation (EC) preceedes a combined echo and noise reduction filter (ENR), and vice versa, the configuration ENR/EC.

Although there are good arguments in favour of processing first the noise reduction, our considerations and

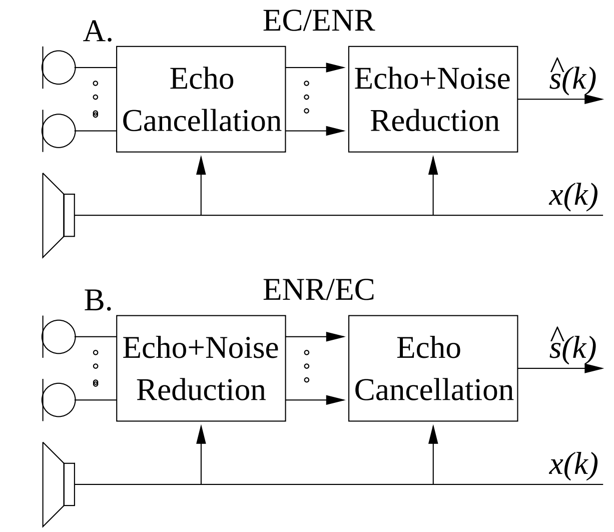

Figure 1: Processing configurations for combined systems: A) EC preceedes ENR, B) ENR preceedes EC.

experiments clearly show that the configuration of Figure 1A, where the echo compensation preceeds the noise reduction, is preferable.

The main advantage of the EC/ENR configuration is that the noise reduction has not to cope with the disturbing echo signal as it is present in the microphone signal and that there is no time varying noise reduction filter in the echo path. Besides that, if the echo canceller does not deliver sufficient echo attenuation, the residual echo can be treated similar to the background noise signal and can be further attenuated by the noise reduction filter. This idea is successfully exploited in a frequency selective echo reduction technique, called "echo shaping" [14, 15], which does not require complete cancellation of the echo by the echo canceller and is easily combined with a noise reduction filter. A disadvantage of the EC/ENR configuration is that the echo canceller has to process signals which are not noise reduced. As a result, algorithms have been proposed where besides the noise reduction filter in the sending path, noise reduced signals are used to adapt the echo canceller [7, 11].

It was pointed out by *Faucon et al.* [11, 10] that the minimum mean square solution to the combined problem also indicates the EC/ENR processing order. Indeed, assume that an estimate $\hat{s}(k)$ of the clean near end speech signal $s(k)$ is obtained by a linear filtering operation of the microphone signal $y(k)$ and the far end speech signal $x(k)$

$$\hat{s}(k) = \boldsymbol{Y}^T(k)\boldsymbol{W}_1 + \boldsymbol{X}^T(k)\boldsymbol{W}_2 \qquad (1)$$

with $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ denoting coefficient vectors of order N and M of two FIR filters and $\boldsymbol{Y}(k)$ and $\boldsymbol{X}(k)$ denoting vectors which contain the most recent N samples of $y(k)$ and the most recent M samples of $x(k)$, respectively. It is then straightforward to show that the minimization of $E\{(s(k) - \hat{s}(k))^2\}$ yields the estimated speech signal

$$\hat{s}(k) = (\boldsymbol{Y}^T(k) - \boldsymbol{X}^T(k)\boldsymbol{R}_{xx}^{-1}\boldsymbol{R}_{xy})\boldsymbol{W}_{ENR} \qquad (2)$$

and the echo and noise reduction filter

$$\boldsymbol{W}_{ENR} = (\boldsymbol{R}_{yy} - \boldsymbol{R}_{yx}\boldsymbol{R}_{xx}^{-1}\boldsymbol{R}_{xy})^{-1}\boldsymbol{R}_{ss} \qquad (3)$$

where $\boldsymbol{R}_{xx}$, $\boldsymbol{R}_{ss}$, $\boldsymbol{R}_{xy}$, $\boldsymbol{R}_{yx}$, and $\boldsymbol{R}_{yy}$ denote correlation matrices of the signals in the subscripts. From equation (2) it can be concluded that echo cancellation should preceed noise reduction. If the echo canceller achieves perfect cancellation the echo and noise reduction filter $\boldsymbol{W}_{ENR}$ depends only on the autocorrelation matrices of the noise $n(k)$ and the clean speech $s(k)$

$$\boldsymbol{W}_{ENR} = (\boldsymbol{R}_{ss} + \boldsymbol{R}_{nn})^{-1}\boldsymbol{R}_{ss} \quad , \qquad (4)$$

otherwise the echo and noise reduction filter also takes the residual echo $d(k) - \hat{d}(k)$ into account.

2.1.2 Single Microphone vs. Multiple Microphone Approach

As far as the echo compensation part of the combined system is concerned, there is no need to use more than one microphone. Indeed, besides the additional hardware the echo compensation task becomes more complicated when multiple microphones are used. However, it can be of great advantage to incorporate a multi-microphone noise reduction algorithm into the combined system. Besides the single microphone procedures, multiple microphone algorithms can exploit the spatial coherence of sound fields. This leads to noise reduction algorithms which are more robust in the presence of non-stationary noise sources and to intelligibility improvements. It should be also noted that late echoes and reverberation exhibits similar spatial correlation properties as diffuse noise fields. Multi-microphone techniques are therefore suited to reduce echoes as well as the potentially annoying reverberation of the near end speech.

2.1.3 Time Domain vs. Frequency Domain

Frequency domain solutions provide the algorithm designer with more degrees of freedom and lead to computational efficient implementations. They tend, however, to have a larger processing delay which might be critical, especially in mobile communication applications. If the hands-free device is designed to be an "add-on" device for arbitrary telecommunication terminals, a low delay solution is certainly desirable. On the other hand, it is generally agreed that in contrast to the echo compensation problem the noise reduction task, as it is discussed here, cannot be solved without some processing delay. Since some delay is inevitable, the frequency domain provides more flexibility to solve the combined problem.

2.2 Algorithms

We will now discuss some of the combined algorithms as they have been proposed in the literature.

The first attempts to develop a combined echo and noise reduction system can be attributed to *Grenier et. al.* [6, 16] and to *Yasukawa* [5]. Both employ more than one microphone. While *Grenier et. al.* use an adaptive beamforming approach [17] the algorithm by *Yasukawa* is based on the noise cancelling idea [13] and uses a noise reference microphone. It may be remarked that in a reverberant environment the noise reduction component of both techniques cannot be very effective. The beamforming technique [16], however, will also dereverberate the near end speech signal to some extent, and thus has the potential for improved speech intelligibility.

A two microphone system based on the spatial coherence of the near end speech and noise in reverberant rooms was investigated in [8, 9, 12, 18]. It has been shown that combining conventional echo cancellers (possibly of reduced order) with a noise reduction technique suitable for diffuse noise fields a true synergy of echo and noise reduction can be obtained. In this algorithm the same adaptive filter is used to suppress noise, to reduce residual echoes, and to dereverberate the near end speech signal. Since this adaptive filter achieves a significant echo reduction, the order of the echo compensator can be reduced with the advantage of faster convergence and higher robustness with respect to background noise.

Faucon et. al. [11, 10] proposed a single microphone frequency domain system. They investigated and optimized spectral weighting rules based on the Minimum Mean Square Short Time Spectral Amplitude estimator [19].

3 Limitations - "The Speech Enhancement Trilemma"

Despite all the progress during the recent years there are certain restrictions which are not easily overcome. These apply both to the echo cancellation and the noise reduction components of the combined system and will be discussed below.

Besides the complexity, the main problems of echo cancellation are background noise, near end speech activity, and a time varying LRM system. Furthermore, the exitation signal, i.e. the far speech signal, is nonstationary and thus not well suited for the identification of a time-varying system with thousand or more coefficients. Whenever the near end speaker is active (and moves) and the far end signal is zero, there is no way to identify the time varying LRM-system. As a consequence, when the far end speaker becomes active again, the deviation of the estimated impulse response with respect to the true impulse of the LRM-system can be large. In these situations a compensator of reduced length, as it can be used in combination with an echo shaping filter, is of advantage since it will converge faster.

Also, it is well known that the amount of noise reduction which can be achieved by a noise reduction filter in the sending path is limited by the admissable distortion of the near end speech signal.

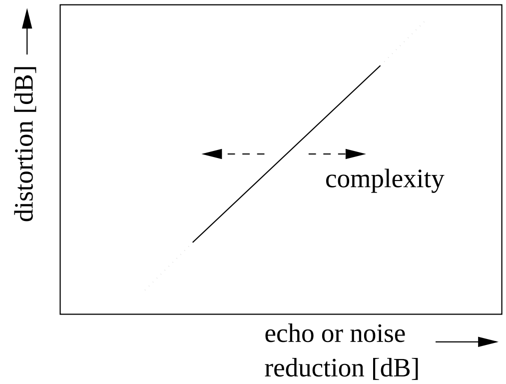

Figure 2: Qualitative plot of the near end speech distortion vs. the echo or noise reduction characteristics for combined systems.

We are thus faced with a trilemma which requires a compromise between the echo and noise reduction, the distortion of the near end speech signal, and the complexity of the overall system. This "Speech Enhancement Trilemma" is illustrated in Figure 2 for a hypothetical combined system [20, 21]. The system is characterized by a curve in the "distortion of the near end speech signal" vs. "echo or noise reduction" plane. The resulting curve which is not necessarily a straight line is typical for a given algorithm and can be shifted (or tilted) by means of increasing or reducing the complexity of the algorithm or choosing a different kind of algorithm. E.g. adding more microphones or more compensator coefficients will result in improved noise or echo reduction at the price of increased complexity. On the other hand, keeping the complexity constant, we can trade echo or noise reduction for improved quality of the near end speech signal. It depends on the application how much distortion of the near end speech signal can be accepted, and on the performance of a given algorithm how much echo and noise reduction can be achieved. The plot in Figure 2 is in any case well suited to compare different systems.

4 Perspectives

As we consider the processing order EC/ENR as the best configuration it remains an open question how to improve the compensator performance in the presence of strong ambient noise. The echo shaping technique is here helpful since it allows to reduce the order of the compensator and adjusts the residual echo power to the background noise level. Since background noise can mask much of the residual echo, the echo needs only to be suppressed when it is actually audible. Optimal estimators eventually incorporating psychoacoustic criteria as outlined e.g. in [22] are certainly desirable.

Also, the noise reduction component can be improved using more effective two microphone noise reduction techniques [23]. One of the questions linked to multi-microphone techniques is how to compensate the microphone signals efficiently. One approach, outlined in [8, 9] is to combine the signal of several microphones and to compensate the combined signal. Due to the rapid progress in signal processing technology a single DSP implementation of a two microphone combined system seems to be feasible, at least for environments with little reverberation. An improvement with respect to convergence can be expected as efficient fast converging echo cancellation algorithms emerge and their sensitivity to background noise is reduced [24].

5 Acknowledgements

We would like to thank our colleagues at the Institute of Communication Systems, Aachen University of Technology, at the Institute for Network and Signal Theory, University of Darmstadt, and at the Technology Centre of German Telekom, Berlin, for the continued scientific exchange and stimulation.

References

[1] E. Hänsler, "The Hands-Free Telephone Problem - An Annotated Bibliography", Signal Processing, Vol. 27, pp. 259-271, 1992.

[2] E. Hänsler, "The Hands-Free Telephone Problem - An Annotated Bibliography Update", Annales des Télécommunication, Vol. 49, No. 7-8, pp. 360-367, 1994.

[3] Y. Ephraim, "Statistical-Model-Based Speech Enhancement Systems", Proc. IEEE, Vol.80, No. 10, pp. 1526-1555, October 1992.

[4] J.R. Deller, J.G. Proakis, and J.H.L. Hansen, "Discrete-Time Processing of Speech Signals", Macmillan, 1993.

[5] H. Yasukawa, "Acoustic Echo Canceller with Sub-Band Noise Cancelling", Electr. Lett., Vol. 28, No. 15, pp. 1403-1404, July 1992.

[6] Y. Grenier, M. Xu, J. Prado, and D. Liebenguth, "Real-Time Implementation of an Acoustic Antenna for Audioconferencing", 1st Workshop on Acoustic Echo Control, Berlin, 1989.

[7] T. Huhn and H.-J. Jentschel, "Kombination von Geräuschreduktion und Echokompensation beim Freisprechen", Nachrichtentechnik, Elektronik, Vol. 43, No. 6, pp. 274-280, December 1993 (in German).

[8] 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.

[9] 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.

[10] B. Ayad and G. Faucon, "Acoustic Echo and Noise Cancelling for Hands-Free Communication Systems", Proc. Fourth Int. Workshop on Acoustic Echo and Noise Control, pp. 91-94, Røros, Norway, June 1995.

[11] G. Faucon and R. Le Bouquin Jeannes, "Joint System for Acoustic Echo Cancellation and Noise Reduction", Proc. EUROSPEECH '95, Madrid, pp. 1525-1528, 18-21 September, 1995.

[12] R. Martin and J. Altenhöner, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction", Proc. Int. Conf. Acoustics, Speech, Signal Processing, Detroit, pp. 3043-3046, May 1995.

[13] B. Widrow, J.R. Glover, J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, and R.C. Goodlin, "Adaptive Noise Cancelling: Principles and Applications", Proc. IEEE, Vol. 63, No. 12, pp. 1692-1716, December 1975.

[14] R. Martin, "Combined Acoustic Echo Cancellation, Spectral Echo Shaping, and Noise Reduction", Proc. Fourth Int. Workshop on Acoustic Echo and Noise Control, pp. 48-51, Røros, Norway, June 1995.

[15] R. Martin and S. Gustafsson, "An Improved Echo Shaping Algorithm for Acoustic Echo Control", Proc. *EUSIPCO, 1996*, Triest, 1996.

[16] M. Xu and Y. Grenier, "Acoustic Echo Cancellation by Adaptive Antenna", 1st Workshop on Acoustic Echo Control, Berlin, 1989.

[17] O.L. Frost, "An Algorithm for Linearly Constrained Adaptive Array Processing", Proc. IEEE, Vol. 60, No. 8, pp. 926-935, August 1972.

[18] R. Martin, "Hands-free Telephones based on Multi-Microphone Echo Cancellation and Noise Reduction", Dissertation (in German), Aachen University of Technology, June 1995.

[19] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator", IEEE Trans. Acoustics, Speech, Signal Processing, Vol. 32, No. 6, pp. 1109-1121, December 1984.

[20] S. Gustafsson, R. Martin, and P. Vary, "On the Optimization of Speech Enhancement Systems Using Instrumental Measures", Proc. Workshop on Quality Assessment in Speech, Audio and Image Communication, Darmstadt, Germany, March 11-13, 1996.

[21] R. Le Bouquin, G. Faucon, and A. Akbari Azirani, "Proposal of a Composite Measure for the Evaluation of Noise Cancelling Methods in Speech Processing", Proc. EUROSPEECH '93, pp. 227-230, Berlin, September 1993.

[22] R. Martin and S. Gustafsson, "The Echo Shaping Approach to Acoustic Echo Control", Submitted to Speech Communication.

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

[24] A. Gilloire, "Recent Advances in Adaptive Filtering Algorithms for Acoustic Echo Cancellation", Proc. Fourth Int. Workshop on Acoustic Echo and Noise Control, pp. 115-134, Røros, Norway, June 1995.