ECHO AND NOISE REDUCTION FOR HANDS-FREE TERMINALS - STATE OF THE ART -

Gérard FAUCON, Régine LE BOUQUIN-JEANNÈS

Laboratoire du Traitement du Signal et de l'Image - Université de Rennes 1
Bât. 22 - Campus de Beaulieu - 35042 RENNES CEDEX - FRANCE

ABSTRACT

This paper deals with speech enhancement in hands-free telecommunication systems. We summarize and discuss recent results on methods combining the two major problems encountered in such systems - acoustic echo cancellation and noise reduction -. Single microphone and two-microphone approaches are addressed. Finally, we outline the limitations of the different techniques and propose some prospects.

1. INTRODUCTION

The increasing demand for communication systems including hands-free technology stimulates effort to develop efficient and compact joint systems for noise reduction (NR) and acoustic echo cancellation (AEC). As a matter of fact, the realization of a hands-free communication system requires solutions to these two fundamental problems. First of all, an echo control device is necessary to suppress the feedback of the far end speaker, to guarantee the stability of the electro-acoustic loop and to supply sufficient echo reduction. The second problem concerns the reduction of the ambient noise which becomes necessary due to the relative large distance from the microphone to the speaker's mouth. Moreover, in mobile environments the signal to noise ratio is often very low and as a consequence we cannot overrule the noise reduction problem. Increased efforts have been made independently in the development of AEC and NR systems for many years [1,2]; now, in modern applications involving noisy environments, the requirements for acoustic echo cancellers are more stringent. In the last few years, it has been recognized that the two problems can be tackled in a combined approach to recover a near-end speech signal only slightly distorted for a sufficient attenuation of echo and noise [3,4,5]. Of course, the double talk detection has to be taken into account to provide a near-end speech which is more pleasant to listen to. This contribution summarizes papers dealing with the combined treatment. It is organized as follows: in the next section we recall techniques relative to AEC including some improvement. In sections 3 and 4, we present combined systems developed for one and two microphone(s) respectively. Finally, we outline the crucial point of the objective evaluation of such systems and we draw some conclusions and prospects.

2. ACOUSTIC ECHO CANCELLATION

A great number of solutions has been proposed these last years regarding echo cancellation [6]. A new approach where a post-filtering is applied to the AEC output to reduce residual echo was recently proposed by Martin [7]. Following the same approach, Turbin proposes and compares three post-filtering algorithms [8]. The echo canceller is a classical adaptive FIR filter, such as an NLMS (Normalized Least Mean Squares) algorithm, with a limited number of filter taps. Then the post-filter coefficients are derived from the observation and the AEC output. The first algorithm is a Wiener filtering. Its performance relies on an efficient echo canceller. If its length is chosen to be small for complexity reasons, the efficiency of the post-processing is limited. The second algorithm outweights the influence of the echo by adding a quantity proportional to the psd (power spectral density) of the estimated echo to the observation psd. In the third method the post-filtering is based on spectral subtraction and uses the signal to echo ratio. The last technique was found to be the most efficient. The post-filter length can be adjusted to get a trade-off between echo attenuation and spectral distortion. Other approaches to improve AEC consist in optimising the convergence factor of the AEC in a noisy environment [9].

3. SINGLE MICROPHONE APPROACHES

We assume that only one microphone and a loudspeaker are available. The microphone observation $x(t)$ is composed of the near-end speech signal $s(t)$ to be transmitted, an echo $e(t)$ and a noise $n(t)$: $x(t)=s(t)+e(t)+n(t)$. The loudspeaker emits a signal $z(t)$ which is correlated with $e(t)$ and used as a reference to cancel this echo. No noise reference is available and noise characteristics must be learned during speech pauses for further use in noise reduction.

3.1. Optimal filtering

In the frequency domain, the optimal filter applied on the observation vector $Y(f)=\left[X(f)\quad Z(f)\right]^T$ in the sense of the minimum mean-square error is [10]:

$$W(f)=\frac{\gamma_{ss}(f)}{\gamma_{ss}(f)+\gamma_{nn}(f)}\begin{bmatrix}1 & -\dfrac{\gamma_{xz}(f)}{\gamma_{zz}(f)}\end{bmatrix}^T \qquad (1)$$

such that $\hat{S}(f)=W^T(f)Y(f)$. $\hat{S}(f)$, $X(f)$, $Z(f)$ represent the spectra of the signals $\hat{s}(t)$, $x(t)$ and $z(t)$ respectively, $\gamma_{ss}(f)$, $\gamma_{nn}(f)$ and $\gamma_{zz}(f)$ are the psd of $s(t)$, $n(t)$ and $z(t)$. $\gamma_{xz}(f)$ is the cross psd between the observations $x(t)$ and $z(t)$. There are two steps involved in the optimal structure: in the first step the echo is estimated by filtering the reference $z(t)$. For an optimal echo canceller, speech and noise are transmitted with no change and the echo is completely cancelled. The echo canceller output is ideally

$s(t)+n(t)$. In the second step the noise is reduced by a Wiener filtering whose gain is $\gamma_{ss}(f)/(\gamma_{ss}(f)+\gamma_{nn}(f))$. The optimal structure is composed of the two optimal filters relevant to each topic.

3.2. Combined systems

The basic combined structure corresponds to the implementation of the optimal filtering where the AEC system preceeds the NR system (Figure 1) [10,11].

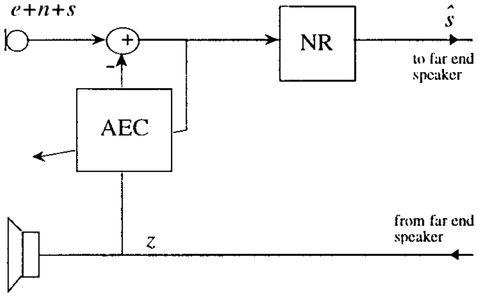

Figure 1. Implementation of the optimal filtering for a single microphone

In practice, the AEC system is disturbed by the additive noise and by the near-end speech signal present on the microphone. The noise is omnipresent and the adaptation is necessarily performed in the presence of noise.

To reduce the noise influence on the AEC system, an analysis and an associated NR filter can be placed before the AEC algorithm (Figure 2) [12].

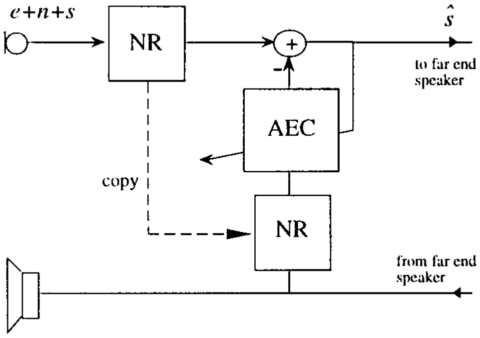

Figure 2. Structure developed in [12]

The NR operation enhances the signal to noise ratio, but it also introduces non-linear distortions on the echo signal which can disturb the identification operation. The copy of the NR filter in the identification branch is aimed at reducing this potential disturbance [12]. For this analysis, two AEC algorithms and one NR system have been considered. The first AEC algorithm is the well-known NLMS and the second one is based on the second order Affine Projection Algorithm, named Soft Decision APA2 (SDAPA2) [13].

The previous structure allows to reduce the noise influence on the AEC. An experimental study shows that, in spite of the distortion brought by the NR system, it is better to first reduce the disturbing noise to obtain a more accurate echo estimate. In the structure proposed in [10] (Figure 3), the noise influence on the AEC system is first reduced by the introduction of a noise reduction filter H_1 as in the previous structure. The echo $\hat{e}(t)$ estimated by the AEC system is subtracted from the observation $x(t)$ to get $v(t)=s(t)+n(t)+e(t)-\hat{e}(t)$. Then a second noise reduction filter H_2 is applied to $v(t)$ to give the final estimate. For the practical implementation, the AEC system is the GMDF (Generalized Multi-Delay Filter) algorithm [14], and for the noise reduction, the algorithms are based on the WI method [15] derived from the Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE STSA) estimator proposed by Ephraim and Malah [16].

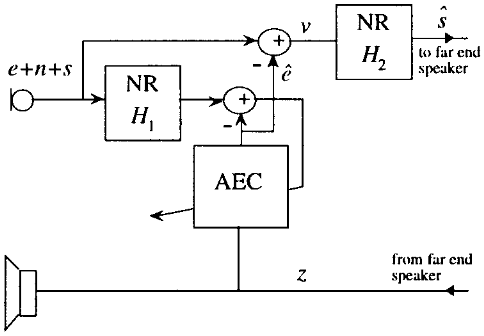

Figure 3. Structure developed in [10]

Another structure is proposed in [17,18] to decrease the distortion on the near-end speech signal. In the system given in Figure 1, the AEC is disturbed by the ambient noise and the near-end speech signal in double talk mode. At the AEC output, this speech signal may be distorted and it seems better to derive the psd of the speech signal from the observation to obtain a less distorted signal. Therefore, the observation $x(t)$ is used to derive the noise filtering. The NR filter and the AEC system are estimated simultaneously. The practical implementation is the same as in the previous structure.

Following the objective of reducing the noise influence on the echo canceller, Capman proposed to modify the adaptation process of a frequency-domain acoustic echo canceller by incorporating a spectral subtraction step aiming at better noise reduction properties [19]. First of all, to improve the performance and reduce the delay, he uses an MDFO algorithm (Multi-Delay Frequency domain algorithm with Overlap) which processes overlapped input blocks by more than half the FFT size. Then he adapts the adaptive filter weights with a noise-free residual echo signal. This modified adaptation process can be extended to noise reduction for speech enhancement with an extra inverse discrete Fourier transform to recover the enhanced near-end speech. The noise reduction is performed using the NSS algorithm [20], which provides distortion-free enhanced speech for an SNR improvement of approximately 10 dB.

In the approach developed by Martin, the conventional echo canceller is followed by a second filter whose aim is to attenuate the noise and the residual echo [7,21,22]. At first, a signal combining the microphone signal and the AEC output is created: $g(t)=a(t).x(t)+(1-a(t)).y(t)$ where $a(t)$ is an adaptive mixing factor in the range [0-1]. This signal is used as a reference to an adaptive filter (NLMS) whose primary channel is the delayed AEC output $y(t)$ to get uncorrelated noise components. The coefficients of this filter are copied in a second filter which processes the compensated signal $y(t)$ prior to transmitting it to the far-end speaker. Martin indicates that under simplifying assumptions, for $a(t)=1$, the filter H doubles the echo attenuation given by the AEC. During single talk, $a(t)=1$ and so the filter tends to attenuate the noise and the residual echo. To avoid distortions of near end speech during double talk, $a(t)$ is set to 0.3.

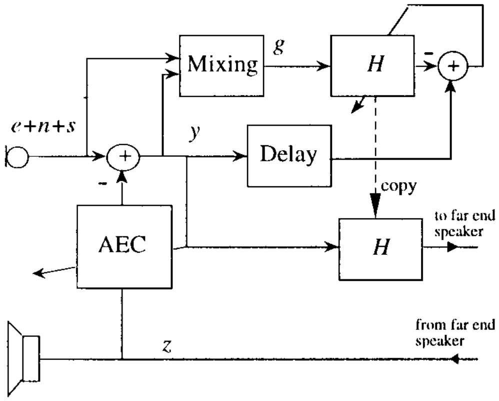

Figure 4. Structure developed in [7]

4. TWO-MICROPHONE APPROACHES

In this section two microphones are used, each receiving a useful near end speech signal $s_i(t)$, an echo $e_i(t)$ and a disturbing noise $n_i(t)$: $x_i(t)=s_i(t)+e_i(t)+n_i(t)$ $(i = 1{,}2)$. For each channel i, $s_i(t)$, $e_i(t)$ and $n_i(t)$ are additive and independent. The echoes $e_i(t)$ are related to the signal z emitted by the loudspeaker and the speech signals $s_i(t)$ are correlated.

4.1. Optimal filtering

The distance between the microphones is such that noises are assumed to be spatially decorrelated. Theoretically, the optimal filtering in the sense of the minimum mean square error leads to the following expression of the signal estimate in the frequency domain, $\hat{S}_1(f)$:

$$\begin{aligned}\hat{S}_1(f)=&\left(X_1(f)-\frac{\gamma_{x_1z}(f)}{\gamma_{zz}(f)}Z(f)\right)\frac{\gamma_{s_1s_1}(f)\gamma_{n_2n_2}(f)}{\Delta}\\&+\left(X_2(f)-\frac{\gamma_{x_2z}(f)}{\gamma_{zz}(f)}Z(f)\right)\frac{\gamma_{s_1s_2}(f)\gamma_{n_1n_1}(f)}{\Delta}\end{aligned}\qquad(2)$$

where $\Delta=\gamma_{s_1s_1}(f)\gamma_{n_2n_2}(f)+\gamma_{s_2s_2}(f)\gamma_{n_1n_1}(f)+\gamma_{n_1n_1}(f)\gamma_{n_2n_2}(f)$. $U(f)$ is the Fourier transform of the signal $u(t)$ and $\gamma_{uv}(f)$ is the power spectral density between $u(t)$ and $v(t)$. Equation (2) is equivalent to first cancelling the echo on each channel and then applying a vectorial Wiener filtering at the AEC outputs to reduce the disturbing noises.

4.2. Combined systems

In the same way as in the single microphone approaches, the first structure (Figure 5) consists in the implementation of the optimal filtering where the acoustic echo cancellers come before the noise reduction system [23]. In this structure, the distance between the microphones is about 40 cm so that noises are decorrelated.

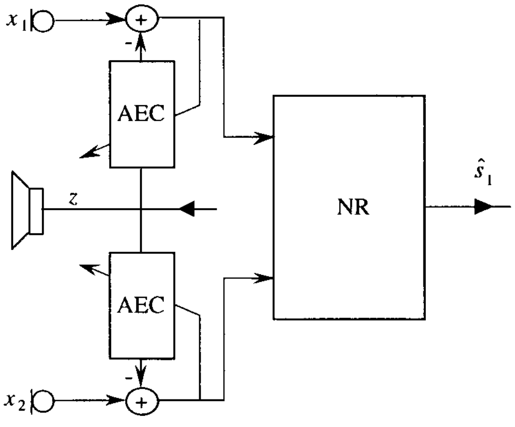

Figure 5. Implementation of the optimal filtering for two microphones

In this structure, it appears that the noises dramatically disturb the AEC systems. Another structure which minimizes the noise influence on the AEC systems by first performing a noise reduction on each channel is proposed in [23]. The estimated echoes are subtracted from the observations to give residual echoes which are less audible. Then a two-channel noise reduction system is applied to get the estimated signal. For the practical implementation, the AEC systems are GMDF algorithms. As for the two-channel noise reduction filter, the vectorial Wiener filtering has been dropped due to its complexity and the estimation errors in the computation of the gain. The two-channel noise reduction technique is a PSI (Preprocessing + Signal Identification) method [15]. In the noise reduction preprocessing, the algorithm is the WI technique [15].

In the same way, the concept proposed by Martin for one microphone is extended to a two-microphone system [24]. At first, an echo cancellation is performed on each channel. The distance is chosen such that the noise components received by the microphones are mutually uncorrelated on both microphones. A Wiener filtering H is applied to the half-sum of the AEC outputs to attenuate all uncorrelated components. The filter H is the mean of two adaptive filters H_1 and H_2 computed as in the single microphone approach but each one now using the other AEC output as reference. This approach reduces noise, reverberation and far end echoes.

Contrary to the previous structures, Yasukawa considered the case of spatially correlated noises [25]. In the presence of noise only, an identification of the transfer function between noises is performed to reduce the noise on a microphone when the near-end speech or the far-end speech is present. Because noise mainly exists in the lower frequency band, the noise cancelling is only implemented in a low band. It is assumed that the near-end speech is not distorted. Then, when the far-end speech signal is present, an AEC is carried out on the NR output.

5. EVALUATION, LIMITS AND PROSPECTS

The basic objective of speech enhancement systems when designing a hands-free terminal is to provide the human user with satisfactory quality. The problem is to specify the performance of the systems so that they reasonably satisfy the users [6,26]. It is obvious that objective measurements are more tractable than subjective tests. Among quantities specified for echo cancelling, the most common measures are the ERLE (Echo Return Loss Enhancement) in ST (single talk) and DT (double talk) modes, the attenuation and the distortion in DT mode (such as cepstral distance and basilar measure). Concerning the noise reduction, the criteria are often based on noise attenuation factor and speech distortion. However, it is likely that more refined criteria would be better correlated to overall speech quality. Little work has been done to correlate objective criteria and subjective tests. The directions given by Gilloire for AEC also stand for the echo and noise reduction [26]; they are to improve objective performance towards better correspondence with subjective quality and to specify objective test conditions. As regards the problem of echo and noise reduction, the evaluation is even more complicated since the audible echo level depends on the residual noise. The choice of method is not easy since, for a given application which requires some performance, we must take into account the complexity and the improvement brought by implemented systems. Moreover, the acoustic channel is a time-varying system and the far-end speech signal is non stationary. This one is intermittent and the deviation of the estimated impulse response with respect to the true impulse response can be large. It is well known that the amount of noise reduction achieved by a noise reduction filter is limited by the admissible distortion. It is necessary to obtain a compromise between the echo and noise reduction and the distortion on the near-end speech due to the AEC and NR systems. Each combined system is tested under varying conditions. The AEC and NR systems are different and it is obvious that the performance of a structure depends on the NR and AEC systems used. These remarks increase the difficulty of comparison.

Adding more microphones or supplementary steps may result in higher performance at the price of increased complexity. For the echo cancellation, there is no need to use more than one microphone. What is more, the task to have additional steps to get a more robust AEC becomes more complicated in a multimicrophone system. However, it appears interesting to incorporate multiple microphone algorithms which can lead to a more efficient noise reduction in the presence of non stationary noises and simultaneously reduce reverberation and late echoes. Optimal estimators incorporating psychoacoustic criteria must be considered.

References

[1] E. HÄNSLER, "*The Hands-Free Telephone Problem: an Annotated Bibliography Update*", Signal Processing, vol. 27, pp. 259-271, 1992.

[2] E. HÄNSLER, "*The Hands-Free Telephone Problem: an Annotated Bibliography Update*", Annals of Telecommunications, vol. 49, n°7-8, pp. 360-367, 1994.

[3] H.J. MATT, M. WALKER, "*Handsfree Speaking for Communication Terminals*", EUSIPCO, Trieste, pp. 1139-1142, Sept. 1996.

[4] R. MARTIN, P. VARY, "*Combined Acoustic Echo Control and Noise Reduction for Hands-Free Telephony - State of the Art and Perspectives*", EUSIPCO, Trieste, pp. 1107-1110, Sept. 1996.

[5] W. STAMMLER, M. SCHULZ, F. SCHEPPACH, "*Echo Compensation and Noise Suppression for Speech Recognition Applications*", EUSIPCO, Trieste, pp. 1135, 1138, Sept. 1996.

[6] P. NAYLOR, J. ALCAZAR, J. BOUDY, Y. GRENIER, "*Enhancement of Hands-Free Telecommunications*", Annals of Telecommunications, vol. 49, n°7-8, pp. 373-379, 1994.

[7] R. MARTIN, "*Combined Acoustic Echo Cancellation, Spectral Echo Shaping, and Noise Reduction*", Fourth International Workshop on Acoustic Echo and Noise Control, Roros, pp. 48-51, June 1995.

[8] V. TURBIN, A. GILLOIRE, P. SCALART, "*Comparison of Three Post-Filtering Algorithms for Residual Acoustic Echo Reduction*", ICASSP, Munich, pp. 307-310, Apr. 1997.

[9] J. MARX, "*Estimation of the Optimal Convergence Factor for Acoustic Echo Cancellation in a Noisy Environment*", EUSIPCO, Trieste, pp. 1740-1743, Sept. 1996.

[10] G. FAUCON, R. LE BOUQUIN-JEANNÈS, "*Joint System for Acoustic Echo Cancellation and Noise Reduction*", EUROSPEECH, Madrid, pp. 1525-1528, Sept. 1995.

[11] P. SCALART, 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, 1996.

[12] Y. GUÉLOU, A. BENAMAR, P. SCALART, "*Analysis of Two Structures for Combined Acoustic Echo Cancellation and Noise Reduction*", EUSIPCO, Trieste, pp. 1123-1126, Sept. 1996.

[13] A. BENAMAR, "*Étude et Implémentation de la fonction de Contrôle de l'Écho Acoustique pour la Radiotéléphonie Mains-Libres*", Thèse de l'Université de Paris-Sud, Orsay, 1996.

[14] J. PRADO and E. MOULINES, "*Frequency-Domain Adaptive Filters with Applications to Acoustic Echo Cancellation*", Third International Workshop on Acoustic Echo Control, Lannion, pp. 249-258, Sept. 1993.

[15] A. AKBARI AZIRANI, "*Rehaussement de la Parole en Ambiance Bruitée. Application aux Télécommunications Mains-Libres*", Thèse de l'Université de Rennes 1, Nov. 1995.

[16] Y. EPHRAIM and D. MALAH, "*Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator*", IEEE Trans. on Acoustics, Speech and Signal Processing, vol. ASSP-32, n°6, pp. 1109-1121, Dec. 1994.

[17] B. AYAD, G. FAUCON, "*Acoustic Echo and Noise Cancelling for Hands-Free Communication Systems*", Fourth International Workshop on Acoustic Echo and Noise Control, Roros, pp. 48-51, June 1995.

[18] R. LE BOUQUIN-JEANNÈS, G. FAUCON, B. AYAD, "*How to Improve Acoustic Echo and Noise Cancelling Using a Single Talk Detector*", Speech Communication, vol. 20, pp. 191-202, 1996.

[19] F. CAPMAN, J. BOUDY, P. LOCKWOOD, "*Acoustic Echo and Noise Reduction in the Frequency-Domain: a Global Optimisation*", EUSIPCO, Trieste, pp. 29-32, Sept. 1996.

[20] P. LOCKWOOD, J. BOUDY, "*Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the Projection, for Robust Speech Recognition in Cars*", Speech Communication, vol. 11, n°2-3, pp. 215-228, 1992.

[21] R. MARTIN and J. ALTENHÖNER, "*Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction*", ICASSP, Detroit, pp. 3043-3046, May 1995.

[22] R. MARTIN, S. GUSTAFSSON, "*The Echo Shaping Approach to Acoustic Echo Control*", Speech Communication, vol. 20, pp. 181-190, 1996.

[23] R. LE BOUQUIN-JEANNÈS, G. FAUCON, B. AYAD, "*A Two-Microphone Approach for Speech Enhancement in Hands-Free Communications*", International Conference on Communication Technology, Beijing, pp. 13.03.1-13.03.4, May 1996.

[24] R. MARTIN, P. VARY, "*Combined Acoustic Echo Cancellation, Dereverberation and Noise Reduction: a Two Microphone Approach*", Annals of Telecommunications, vol. 49, n°7-8, pp. 429-438, 1994.

[25] H. YASUKAWA, "*Acoustic Echo Canceller with Sub-Band Noise Cancelling*", Electronics Letters, vol. 28, n°15, pp. 1403-1404, July 1992.

[26] A. GILLOIRE, "*Performance Evaluation of Acoustic Echo Control: Required Values and Measurement Procedures*", Annals of Telecommunications, vol. 49, n°7-8, pp. 368-372, 1994.