OPTIMIZATION OF A NOISE REDUCTION PREPROCESSING IN AN ACOUSTIC ECHO AND NOISE CONTROLLER

Beghdad Ayad, Gérard Faucon and Régine Le Bouquin-Jeannès

Laboratoire de Traitement du Signal et de l'Image - Université de Rennes 1
Bât. 22 - Campus de Beaulieu - 35042 RENNES CEDEX - FRANCE
Beghdad.Ayad@univ-rennes1.fr

ABSTRACT

In hands-free communications, the speech signal to be transmitted is disturbed by ambient noise and acoustic echo. So, a processing to reduce these disturbances must be performed before transmission. Classical solutions are cascaded structures where the acoustic echo canceller preceeds or follows the noise reduction system. Recently, we have proposed a new joint system where a noise reduction preprocessing allows to improve the performance of the acoustic echo canceller. This preprocessing reduces the noise but distorts the original echo. This paper presents an optimization of the preprocessing. Objective results in terms of Echo Return Loss Enhancement and gain are presented.

1. INTRODUCTION

For some applications such as teleconferencing and hands-free telephone, the near-end speech signal to be transmitted is disturbed by ambient noise and by an echo due to the coupling between the microphone and the loudspeaker. The dissemination of hands-free communication systems requires to provide users with some comfort. So, both problems have to be solved to obtain a good quality speech signal. If a number of studies have been done separately on noise reduction and echo cancellation, only a few studies concern joint systems including both processings [1,2]. Our objective is to optimize a joint structure to get a near-end speech signal only slightly distorted and low levels of echo and noise.

The microphone observation $x(t)$ is composed of the near-end speech signal $s(t)$, an echo $e(t)$, a noise $n(t)$ and the loudspeaker emits a signal $z(t)$ correlated with $e(t)$. The optimal structure in the sense of the minimum mean square error to be applied on $x(t)$ and $z(t)$ can be easily derived and is composed of two steps. In the first one, the echo is estimated by applying on the reference $z(t)$ a filtering whose transfer function is

$$\gamma_{xz}(f)/\gamma_{zz}(f),$$

where $\gamma_{xz}(f)$ is the cross power spectral density between signals x and z, $\gamma_{zz}(f)$ is the power spectral density of signal z. The filter output is subtracted from the microphone observation. For an ideal echo canceller, speech and noise are transmitted with no change and the echo is completely removed. Then, in the second step, noise is reduced by a Wiener filtering whose gain is

$$\gamma_{ss}(f)/(\gamma_{ss}(f)+\gamma_{nn}(f)).$$

So, the optimal structure is composed of the two cascaded optimal filters, where the Acoustic Echo Cancellation (AEC) system preceeds the Noise Reduction (NR) system. This structure is called AEC+NR (Figure 1).

In practice, the AEC system is adaptive. The coefficients of the AEC are disturbed by ambient noise which is omnipresent and it appears difficult to stop the AEC adaptation when speech occurs. To reduce the noise influence on the AEC system, the place of the NR system and the AEC system may be exchanged, so that the adaptation can be stopped in Double Talk (DT) mode (speech and echo present). Unfortunately, the disturbing noise is less reduced and this implementation differs from the optimal structure. Nevertheless, the echo estimated by the AEC is closer to the original echo when NR preceeds AEC. In [3], experiments reveal that, in spite of the distortion on the echo due to the NR system, it is better to reduce first the disturbing noise to obtain a more accurate echo estimate. Therefore, a new structure [3,4], called AEC+2NR (Figure 2), has been investigated. The noise influence on the AEC system is first reduced by applying a noise reduction filter H_2 on the microphone, then an AEC is performed. The echo $\hat{e}_2$ estimated by the AEC system is subtracted from the microphone observation $x(t)$ to get the signal $v(t)=s(t)+n(t)+e(t)-\hat{e}_2(t)$. Then, a second noise reduction filter H_3 is applied on $v(t)$ to get the final estimate. In this way, the AEC adaptation can be stopped in DT mode and $v(t)$ contains the unchanged speech signal.

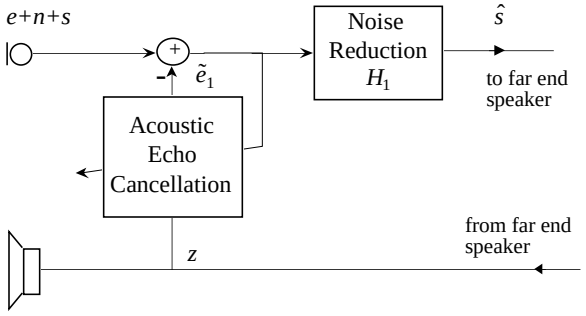


Figure 1. Structure AEC+NR

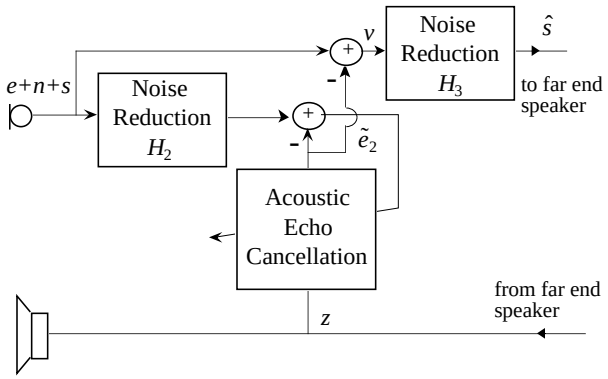


Figure 2. Structure AEC+2NR

2. AEC AND NR SYSTEMS

The Acoustic Echo Canceller is a Generalized Multi-Delay Filter (GMDF) algorithm [5]; it is based on a block frequency-domain adaptive filtering procedure. The two differences with the standard scheme lie in (i) the segmentation of the impulse response into segments, which allows to control the overall processing delay, and (ii) the introduction of a parameter controlling the overlap between the successive input blocks to modify the rate at which the filter coefficients are updated.

The noise reduction algorithm is derived from the Minimum Mean-Square Error Short-Time Spectral Amplitude (MMSE STSA) estimator proposed by Ephraim and Malah [6]. It is based on modeling speech and noise spectral components as stastically independent Gaussian random variables. This algorithm used as a preprocessing (filter H_2) in the structure AEC+2NR will be optimized and we give hereafter a more detailed description of this technique to understand where the optimization occurs. Let $Y(f)$ be the spectrum of a NR input $y(t)$ which is composed of a signal $w(t)$ and a noise $n(t)$. The signal estimate is given by

$$\hat{W}(f) = G_1(f) \cdot G_2(f) \cdot Y(f)$$

where $G_1(f)$ is a Wiener filtering and $G_2(f)$ represents the gain function taking the uncertainty of speech presence into account [7]. This estimator depends on the *a priori SNR* (Signal-to-Noise Ratio), R_{prio}, the *a posteriori SNR*, R_{post}, defined respectively as:

$$R_{prio}(f) = \frac{E\left[|W(f)|^2\right]}{E\left[|N(f)|^2\right]}, \qquad R_{post}(f) = \frac{|Y(f)|^2}{E\left[|N(f)|^2\right]}$$

and the probability of signal absence $q(f)$. $E\left[|N(f)|^2\right]$ is the noise power learned during speech pauses. Ephraim and Malah proposed to estimate the *a priori SNR* according to a decision-directed approach [6]:

$$\hat{R}_{prio}(f,n) = \lambda \frac{\hat{A}^2(f,n-1)}{E\left[|N(f)|^2\right]} + (1-\lambda) Q\left(R_{post}(f,n) - 1\right)$$

where n is the current block number, $\hat{A}(f,n-1)$ is the amplitude of the signal estimated on the block $(n-1)$, λ is a weighting factor, $Q(u)$ is an operator defined by $\mathrm{Max}(u,0)$. $R_{post}(f,n)$ is directly given by the ratio of the squared magnitude of the observation on the block n to the noise power $E\left[|N(f)|^2\right]$.

3. OPTIMIZATION OF THE NOISE REDUCTION PREPROCESSING

In the structure AEC+2NR, what is the best noise reduction filter H_2 to be applied on the microphone observation? A way to modify H_2 is to change the value of the weighting factor λ in the estimation of the *a priori SNR*. We compute a noise reduction factor *R* and a distortion factor *D* brought by the filter H_2:

$$R = \frac{1}{M}\sum_{k=1}^{M} 10\log\frac{P_n(k)}{P_{n_f}(k)}, \quad D = \frac{1}{M}\sum_{k=1}^{M} 10\log\frac{P_{e-e_f}(k)}{P_e(k)};$$

e_f and n_f represent the echo and the noise filtered by the noise reduction filter H_2 respectively, $P_u(k)$ is the power of u computed on the kth block of 256 samples, M is the number of blocks on which echo and noise are present together (Single Talk (ST) mode). Figure 3 shows the distortion *D* versus the noise reduction factor *R* for different values of λ and different Echo-to-Noise Ratios (*ENR*). The *ENR* is given by the ratio of the echo power to the noise power on the *M* blocks. When λ tends to 1, the noise reduction and the echo distortion increase, which corresponds to a lower gain of the filter H_2.

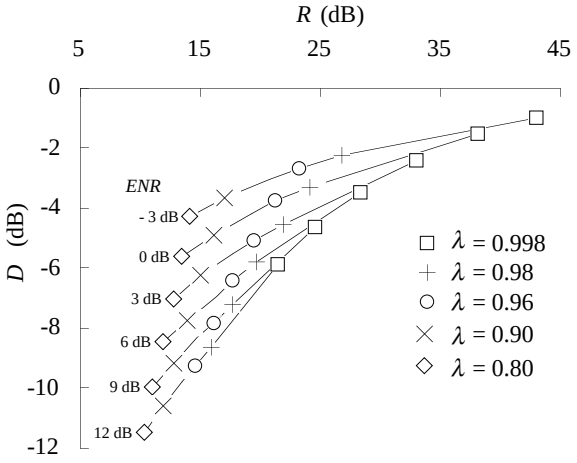


Figure 3. R versus D for different ENR

4. RESULTS

The influence of the noise reduction filter H_2 on the performance of the structure AEC+2NR is evaluated in terms of objective measures. For comparison, the structure AEC+NR is also studied.

a) Methodology of evaluation

The database is obtained by recording the speech signal, the echo and the disturbing noise separately to consider various

SNR and *ENR*. These signals are recorded in a car, and noise is due to the car moving at 130 km/h. From these recordings, we create files of composite signals (Figure 4), the first part is a noisy echo (ST mode) and the second part corresponds to speech added to a noisy echo (DT mode).

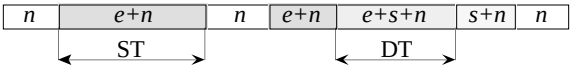


Figure 4. Composite signal

Only three objective measures [3,8] are presented:

- the similarity index *SIM* in ST mode

$$SIM = \frac{1}{N}\sum_{k=1}^{N} 10\log\frac{P_e(k)}{P_{e-\hat{s}_i}(k)},\ i=1,2$$

- the Echo Return Loss Enhancement *ERLE* in both modes

$$ERLE = \frac{1}{N}\sum_{k=1}^{N} 10\log\frac{P_e(k)}{P_{e_r}(k)}$$

- the gain *G* in DT mode

$$G = \frac{1}{N}\sum_{k=1}^{N} 10\log\frac{P_{e+n}(k)}{P_{s-\hat{s}}(k)}$$

k is a block index and N is the number of blocks corresponding to the estimation performed in ST or DT mode, $\hat{s}$ is the final estimate of the near-end speech signal, e_r represents a residual echo computed as follows:

- in the structure AEC+NR, e_r is obtained by filtering the difference $e-\hat{s}_1$ using H_1,

- in the structure AEC+2NR, e_r is obtained by filtering the difference $e-\hat{s}_2$ using H_3.

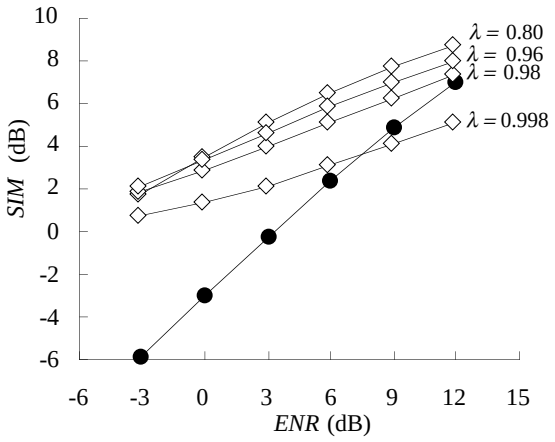
The optimization only concerns the noise reduction filter H_2 in the structure AEC+2NR. The parameters of the AEC system and filters H_1 and H_3 are fixed. We choose the following parameters: for the GMDF algorithm, the length of the impulse response is 256, it is divided into $L=2$ segments and the overlapping between successive blocks is $(256-32)$ samples, the adaptation step is 0.33; for the noise reduction filters, H_1 and H_3, the weighting factor λ is 0.98, the probability of signal absence $q(f)$ is 0.5. The noise power is learned on ten blocks of 256 samples where only noise is present.

Figures 5 to 10 present the objective measures, averaged on a set of ten files, where -●- corresponds to the structure AEC+NR (adaptation continued) and -◇- corresponds to the structure AEC+2NR.

In ST mode, the *ENR* varies from -3 dB to 12 dB and in DT mode, the *ENR* and *SNR* are identical and these ratios vary from -3 dB to 12 dB.

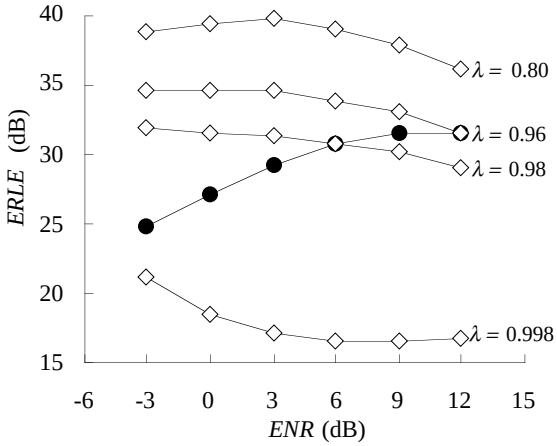
b) Influence of the weighting factor λ

In ST mode, the AEC is only disturbed by noise. Figures 5 and 6 represent *SIM* and *ERLE* for different values of λ. Values of λ in the range $[0.5; 0.8]$ lead to comparable results for high *ENR*; as for low *ENR*, performance falls because of the imperfect noise reduction and so these results are not presented. $\lambda = 0.80$ gives the best *SIM* and *ERLE*, which corresponds to a noise reduction by H_2 around 10 dB (Figure 3). When λ tends to 1, the echo is more distorted and the filtering is less efficient.



-●- *AEC+NR*, -◇- *AEC+2NR*

Figure 5. SIM in ST mode

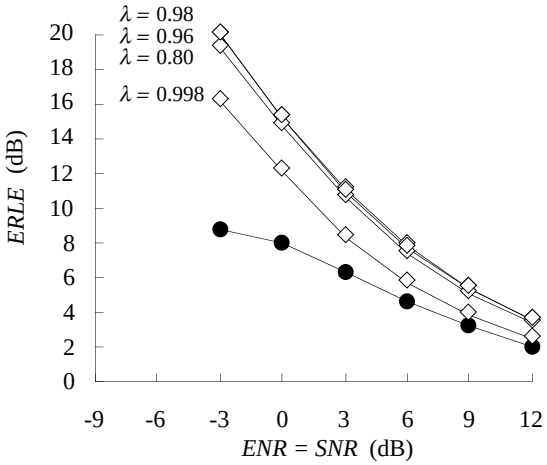


-●- *AEC+NR*, -◇- *AEC+2NR*

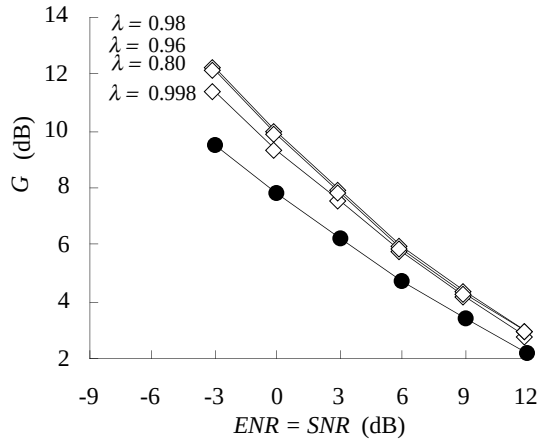
Figure 6. ERLE in ST mode

In DT mode, the question is: is there any advantage to stop the AEC adaptation in the structure AEC+2NR? To this end, two cases are considered:

1-adaptation continued. We can see that, for $0.80 \le \lambda \le 0.98$, the *ERLE* (Figure 7) and the gain *G* (Figure 8) are quite similar; for $\lambda < 0.80$, we note some degradation and results are not presented. Since we obtain interesting results in each mode for $\lambda = 0.80$, this value may be kept in both situations.



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation continued)*
Figure 7. ERLE in DT mode



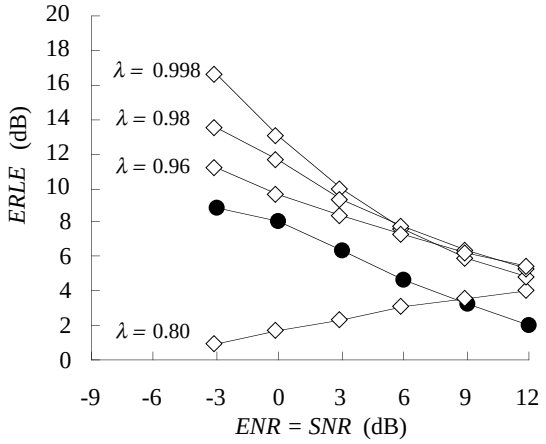
-●- *AEC+NR*, -◇- *AEC+2NR (adaptation continued)*
Figure 8. Gain in DT mode

2-adaptation stopped. We are sure that the speech signal *s* to be transmitted is not changed by the AEC. The *ERLE* (Figure 9) remains high for $0.96 \le \lambda \le 0.998$. These cases correspond to a noise reduction greater than 20 dB for $ENR \le 0$ dB. For these values, the gain of the structure AEC+2NR is greater than that of the structure AEC+NR (Figure 10). The value $\lambda = 0.96$ seems to be a good choice both in ST mode and in DT mode.

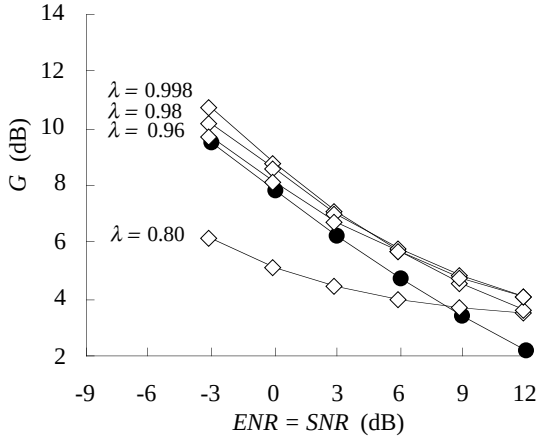
To conclude on these experiments, for low *ENR* (and *SNR*), it seems better to continue the adaptation and, for high *ENR* (and *SNR*), it is better to stop it. Informal listening tests confirm these remarks.

5. CONCLUSION

An optimization of a preprocessing included in a new acoustic echo and noise controller is proposed. Different weighting factors have been considered in the preprocessing in ST and DT modes when the AEC adaptation is continued or stopped. In this last case, the near-end speech signal is less distorted. A complete subjective evaluation has to be conducted to validate the objective measures.



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation stopped)*
Figure 9. ERLE in DT mode



-●- *AEC+NR*, -◇- *AEC+2NR (adaptation stopped)*
Figure 10. Gain in DT mode

REFERENCES

[1] R. MARTIN, J. ALTENHÖNER, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction", *ICASSP*, pp. 3043-3046, May 1995.
[2] H. YASUKAWA, "Acoustic Echo Canceller with Sub-band Noise Cancelling", *Electronics Letters*, vol. 28, n°15, pp. 1403-1404, July 1992.
[3] G. FAUCON, R. LE BOUQUIN JEANNÈS, "Joint System for Acoustic Echo Cancellation and Noise Reduction", *EUROSPEECH*, pp. 1525-1528, September 1995.
[4] R. LE BOUQUIN JEANNÈS, B. AYAD, "Systèmes Combinés d'Annulation d'Echo et de Réduction de Bruit", *GRETSI*, pp. 459-462, Septembre 1995.
[5] E. MOULINES *et al.*, "The Generalized Multidelay Adaptive Filters: Structures and Convergences Analysis", *IEEE Trans. on Signal Processing*, vol. 43, n°1, pp. 14-28, January 1995.
[6] Y. EPHRAIM, D. MALAH, "Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator", *IEEE Trans. on ASSP*, vol. ASSP-32, n°6, pp. 1109-1121, December 1984.
[7] A. AKBARI AZIRANI, "Rehaussement de la Parole en Ambiance Bruitée. Application aux Télécommunications Mains-Libres", *Thèse de l'Université de Rennes 1*, Novembre 1995.
[8] A. GILLOIRE, "Performance Evaluation of Acoustic Control: Required Values and Measurement Procedures", *Annals of Telecommunications*, 49, n°7-8, pp. 368-372, July-August 1994.