ANALYSIS OF TWO STRUCTURES FOR COMBINED ACOUSTIC ECHO CANCELLATION AND NOISE REDUCTION

Yann Guelou ❶, Abdelkrim Benamar ❶❷, Pascal Scalart ❶
❶ France Telecom CNET LAA/TSS/CMC 2, avenue Pierre Marzin 22307 Lannion Cedex FRANCE
❷ Alcatel Mobile Phones 32, avenue Kléber 92707 Colombes Cedex FRANCE
Email : benamar@lannion.cnet.fr

ABSTRACT

This paper addresses the problem of speech enhancement in the context of GSM hands-free radiotelephony where the signal to be transmitted is corrupted by background noise and echo signals. We analyze possible schemes for combined acoustic echo cancellation (AEC) and noise reduction (NR) devices. Considering two AEC algorithms and one NR device, we show that the overall performances obtained by these schemes are greatly dependent on the intrinsic behaviour of the considered AEC algorithms. These results are confirmed by informal listening tests presented in that contribution.

1 INTRODUCTION

For better use and for safety reasons, the standard handset in cars is bound to be replaced by hands-free equipment. Such an equipment introduces specific technical difficulties steming from the high background noise level and reverberant environment encountered in moving vehicles, and from the coupling between the loudspeaker and the microphone(s) of this equipment. Thus, in order to provide satisfactory speech quality, this kind of equipment must include noise reduction (NR) and acoustic echo cancellation (AEC) devices.

To date, very few contributions [1] [2] have concerned combined NR and AEC devices. Nevertheless, global optimization of their performances must be conducted to the extent that the echo perception is greatly dependent on the background noise level. In that way, we propose in that contribution possible structures for combined devices introducing our analysis of their potential interaction.

As a conclusion, we raise the fact that the choice of the structure can be conditioned by the intrinsic performances of the AEC algorithm.

2 ACOUSTIC ECHO CANCELLATION AND NOISE REDUCTION

For our analysis, we consider two AEC algorithms and one NR system. The first AEC algorithm is the well known NLMS [3] where the identification filter is updated as follows :

$$H_t = H_{t-1} + \mu \left[\frac{y_t - H_{t-1}^T X_t}{X_t^T X_t} \right] X_t$$

where X_t is the vector of the last L input signal samples $[x_t, ..., x_{t-L+1}]$, L the filter length and y_t the microphone signal sample at time t.

The second one is based on the second order Affine Projection Algorithm (APA 2) [4] given by the following set of equations :

$$U_t = X_t - \frac{X_t^T X_{t-1}}{X_{t-1}^T X_{t-1}} X_{t-1}$$

$$H_t = H_{t-1} + \mu \left[\frac{y_t - H_{t-1}^T X_t}{U_t^T U_t} \right] U_t$$

The NLMS algorithm has been given a fixed value sufficiently low to reduce its sensitivity to noise. The APA 2 algorithm has been modified, using variable control parameters, in order to improve its robustness to noise and to double talk situations. The obtained algorithm is named Soft Decision APA 2 (SDAPA2).

Many algorithms have been proposed for speech enhancement. In our study, we emphasis on the single microphone approach based on the spectral substraction principle which provides a good compromise between system complexity and noise reduction. A simplified block-diagram of such speech enhancement system is depicted in Figure 1. The single microphone NR system considered for our analysis is based on the minimum mean squared approach proposed by Ephraim and Malah [5].

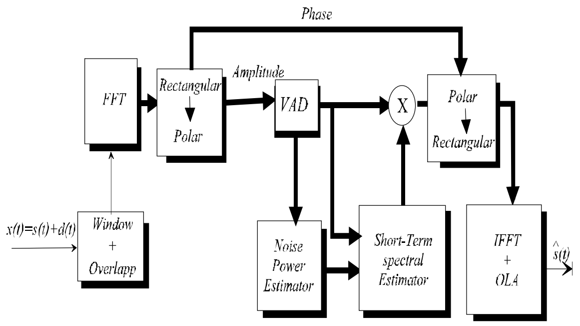

Figure 1 - Noise reduction system

3 ANALYZED STRUCTURES

The considered structures are depicted in the figures below. In the first one named structure A (see Figure 2), the analysis and the associated NR filter are both placed before the AEC algorithm.

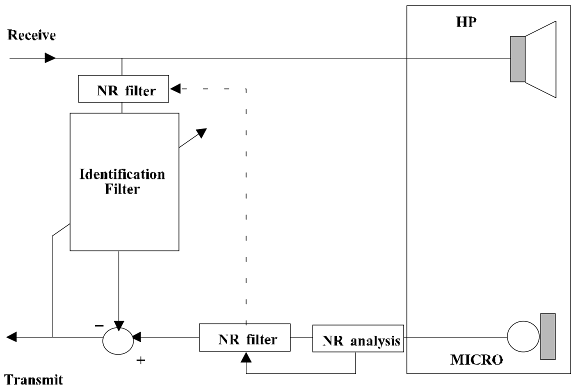

Figure 2 - Structure A

The NR operation enhances the Signal to Noise Ratio (SNR) which can improve the AEC behaviour, but it also introduces non-linear distortions on the echo signal which can disturb the identification operation. One should note that the copy of the NR filter in the identification branch is aimed at reducing this potential disturbance, even if the NR filter should be placed after the identification one in order to be in agreement with the real situation (the NR filter is placed after the acoustic path).

In the second structure named B (see Figure 3), the NR operation is placed after the AEC algorithm. In that configuration, the AEC algorithm doesn't take advantage of the NR operation, but it doesn't suffer from the distorsions mentioned before.

4 EXPERIMENTAL RESULTS

Experiments have been made with near end speech corrupted by echo signal and background noise. The considered microphone signal is composed of echo, noise and near end speech recorded separately in a real environment (car). The Echo to Noise Ratio is close to 0 dB and the Signal to Noise Ratio equals 5 dB. The stepped curve represented under the time-domain microphone signal indicates the vocal activity on the received path (see Figure 4).

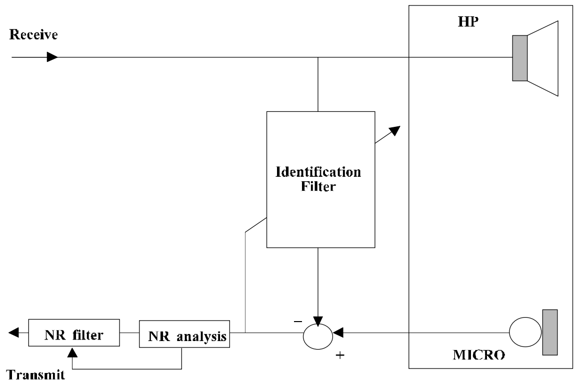

Figure 3 - Structure B

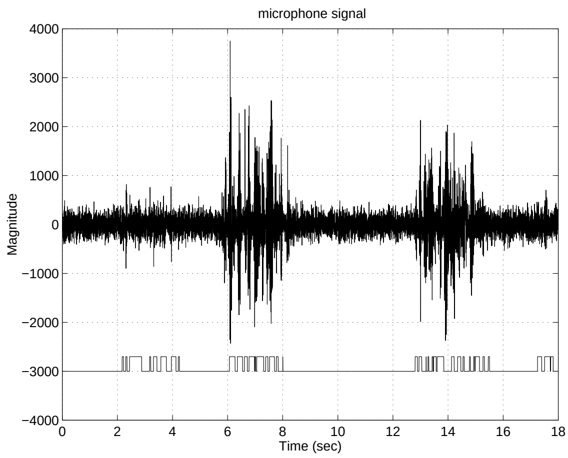

Figure 4 - Microphone signal

In order to provide information on the performances of the overall structures, we consider echo attenuation measured by the Echo Return Loss Enhancement parameter (ERLE) which is computed on blocks of 256 samples without overlapping for a sampling frequency equal to 8 kHz. The spectral and perceptual differences between the processed signals and the original ones are evaluated thanks to cepstral and basilar distances. The basilar distance is provided by the Perceptual Objective Measure system [6]. This measure allows monaural human ear modelization in order to provide the excitation pattern on the basilar membrane. This model is based on low pass filtering modelization of the external and middle ear, analysis in the Bark domain, convolution of the resulting signal with a model of the cochlea spreading function and addition of the internal excitation due to noises such as blood flow.

Our experimental results are represented on the figures below. In order to evaluate the influence of the NR device on the AEC operation, we compare for each structure and for each AEC algorithm the ERLE measures, and the cepstral and the basilar distances between the real echo and the echo estimated by the considered AEC algorithm. This later ones evaluate the quality of the identification operation.

Considering the Figures 5 and 6 which show the performances obtained with the NLMS algorithm, we notice that the identification operation is disturbed by the background noise (negative values for the ERLE parameter). Moreover, considering the distance measures, we can see that better results are obtained with the structure A in comparison with the structure B. This can be explained by the SNR improvement provided by the NR operation in the structure A. Such improvement reduces the influence on the overall performances of the wellknown lack of robustness to noise of the NLMS algorithm. One should also note that, when the NR operation is placed before the AEC one, the overall performances obtained with the NLMS algorithm are more sensitive to the SNR improvement than to the distorsions introduced by the NR device.

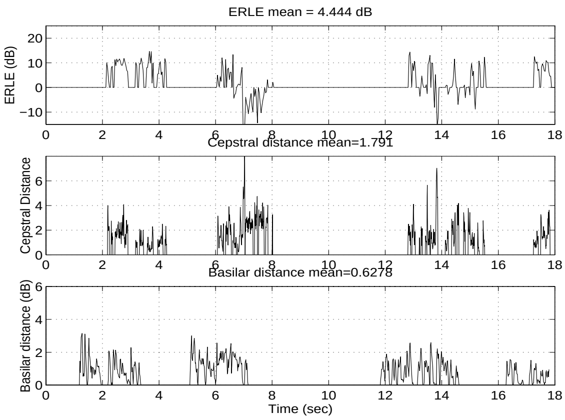

Figure 5 - Performances obtained with the NLMS algorithm considering the structure A

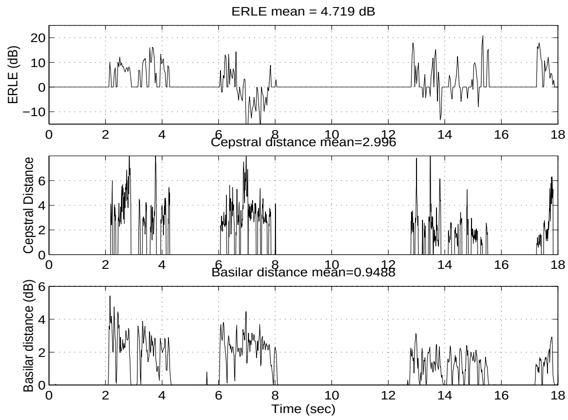

Figure 6 - Performances obtained with the NLMS algorithm considering the structure B

The objective criteria previously described are represented on the figures below when the SDAPA2 algorithm is considered.

According to Figures 7 and 8, we can notice that the SDAPA2 algorithm exhibits better results than those obtained with the NLMS one in terms of cepstral and basilar distances, and also in terms of echo attenuation (no negative value for the ERLE measure) for both structures. Moreover, we can easily see that the performances obtained with the structure B are better than those obtained by the first one. Such a behaviour can be first explained by the high robustness to noise of the considered AEC algorithm.

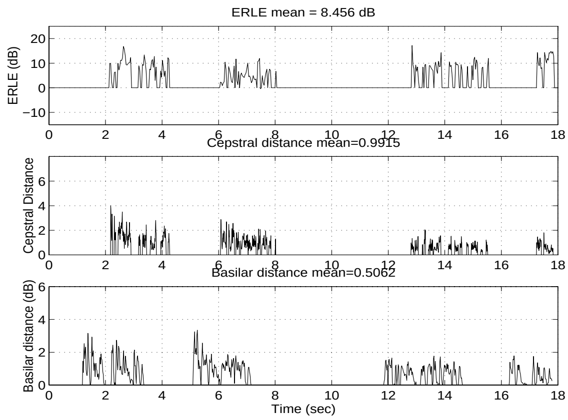

Figure 7 - Performances obtained with the SDAPA2 algorithm considering the structure A

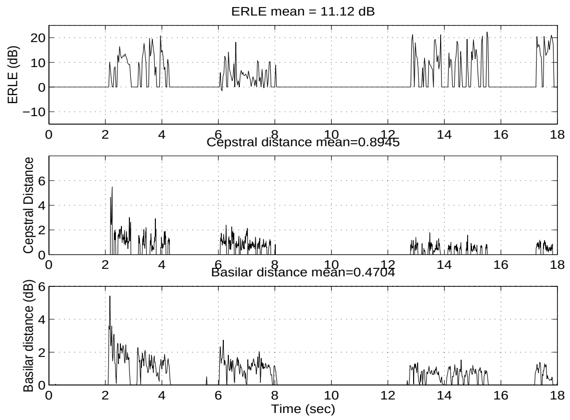

Figure 8 - Performances obtained with the SDAPA2 algorithm considering the structure B

Moreover, contrarily to the NLMS algorithm, the modified APA 2 one seems to be more sensitive to the distortions introduced by the NR operation on the echo signal in the structure A than to the SNR improvement.

5 INFORMAL LISTENING TESTS

Informal listening tests have been conducted in order to compare subjective results to the objective ones described before. The proposed structures have been evaluated through a comparison test. Input speech samples were short french sentences read by male and female speackers, and corrupted by background noise and echo recorded in real situations (stationnary and non-stationnary noise with two different SNR, two different conditions for the signal to echo ratio). These samples were processed by each structure considering the modified APA 2 algorithm for the AEC device. The ten subjects indicated which processed sentence was preferred in terms of noise reduction, echo attenuation and overall quality. The obtained results are summarized in the table below.

Structure A preferred	Structure B preferred	Equality between both structures
17.78%	70%	12.22%

These results confirm the previously reported conclusions drawn from objective criteria.

6 CONCLUDING REMARKS

In the previous parts, we introduced two structures for combined acoustic echo cancellation and noise reduction. Two AEC algorithms and one NR system were considered. It was shown both by objective criteria and informal listening tests that the performances obtained by these structures are greatly dependent on the intrinsic behaviour of the considered AEC algorithms.

Thus, if the AEC solution exhibits good robustness to background noise (like the modified APA2 algorithm), the structure B has to be chosen for combined devices. For the NLMS algorithm, both structures seem to be quite equivalent, and other aspects like the complexity of the implementation ..., etc have to be considered in order to make the final choice.

ACKNOWLEDGEMENTS

The authors would like to thank C. Colomes and M. Lever from CCETT (France) for providing the Perceptual Objective Measure (POM) system.

REFERENCES

[1] R. Martin, J. Altenhoner, "Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction", Proc. IEEE ICASSP'95, pp. 3043-3046, Detroit, USA, 1995.

[2] B. Ayad, G. Faucon, "Acoustic Echo ans Noise Cancelling for Hands-Free Communications Systems", Proc. of the International Workshop on Acoustic Echo and Noise Control, pp. 91-94, Roros, Norway, 1995.

[3] S. Haykin, "Adaptive Filter Theory", Englewood Cliffs, N.J., Prentice-Hall, 1991.

[4] K. Ozeki, T. Umeda, "An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and its Properties", Electronics and Communications in Japan, vol. 67-A, n°5, pp. 19-27, 1984.

[5] Y. Ephraim, D. Malah, "Speech Enhancement using a MMSE Short Time Spectral Amplitude Estimator", IEEE Trans. on ASSP-32, pp. 1109-1121, December 1984.

[6] C.Colomes and al, "A Perceptual Model Applied to Audio Bit-Rate Reduction", J. Audio Eng. Soc., vol. 43, No. 4, pp. 233-240, April 1995.