COMPARISON OF THREE POST-FILTERING ALGORITHMS FOR RESIDUAL ACOUSTIC ECHO REDUCTION

Valérie Turbin, André Gilloire, Pascal Scalart

CNET LAA/TSS/CMC
2, Avenue Pierre Marzin
22307 LANNION Cedex-FRANCE
e-mail : turbin@lannion.cnet.fr gilloire@lannion.cnet.fr scalart@lannion.cnet.fr

ABSTRACT

We consider an acoustic echo control system composed of a short conventional acoustic echo canceller combined with a post-filter in a teleconference context. The post-filter is implemented in an open-loop structure in the frequency domain, which provides good adaptive performance and flexibility for the choice of the post-filter length. Three post-filtering algorithms are compared in terms of residual echo attenuation and near-end speech distortion. The effect of the post-filter length is also examined. Our study confirms that the post-filtering approach provides high residual echo attenuation. Moreover, it appears that the distortion of the near-end speech can be controlled by choosing appropriately the post-filter length.

1. INTRODUCTION

Usual approaches for performing the acoustic echo control for telecommunications terminals operating in hands-free mode rely on the identification of the acoustic echo path with adaptive filters. The use of adaptive echo cancellers, which construct a replica of the acoustic echo path by means of an adaptive filtering method, is up to now considered as the best way to solve that problem. In teleconference contexts, long non-stationary impulse responses are to be identified by the echo canceller, which requires an adaptive filter of large size (several thousand coefficients). Much work has been dedicated to this problem, which led to various adaptive filtering algorithms and structures [1]. However, these algorithms generally exhibit heavy computational load and convergence and tracking problems which appear because of the ill-conditioning of speech and the variation in time of the acoustic echo path. Recently, a new approach combining adaptive and optimal filterings was proposed by Martin and Coll.[2] for hands-free mobile applications. An advantage of this method is to reduce the size of the echo canceller while preserving a high amount of echo attenuation, thanks to the post-filter. In the context of our interest -teleconference-, a similar approach was considered. The principle of this method is recalled Figure 1. A conventional echo canceller (C) of reduced size models the direct path from the loudspeaker to the microphone and the early reflections of the room impulse response, while the post-filter attenuates the residual echo corresponding to late reverberation.

In other words, the task of the post-filter is to achieve reduction of the residual echo that still exists at the output of the echo canceller. The echo canceller is a classical FIR filter with a limited number of filter taps that can be adapted with a standard algorithm like the NLMS.

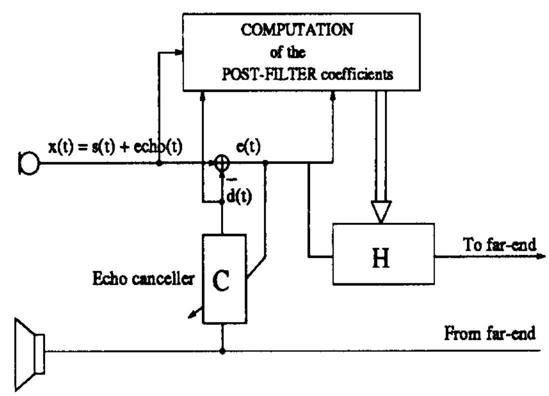

Fig.1 General structure of the combined acoustic echo cancellation system according to [2].

In this paper, we focus on the design and the performance of the post-filter. Thus it is assumed that no convergence problem occurs with the acoustic echo canceller, which is assumed to identify properly the beginning of the echo path impulse response and thus leads to significant echo attenuation, i.e. at least 10 dB. The first part of this paper is dedicated to the realization of the post-filter. A frequency domain implementation in an open-loop structure was preferred because of some advantages that will be discussed. In the second part, three different post-filtering algorithms are considered. The last part of this paper consists in comparing the performance of the three post-filters in terms of additional echo reduction and distortion of the near-end speech signal. Simulations results show that high overall echo attenuation can be obtained with a reasonable amount of signal distortion, although the size of the echo canceller (C) is much smaller than the size of the echo path impulse response.

2. IMPLEMENTATION OF THE POST-FILTER

The post-filter must be able to adapt as soon as talking conditions change : during single talk condition where only the far-end speaker is active, the post-filter is supposed to attenuate strongly the signal $e(t)$ whereas, when far-end and near-end speakers are simultaneously active, the post-filter still has to attenuate the residual echo while causing acceptable disturbances of the near-end speech. Basically the post-filter may be implemented either in an open-loop structure or in a closed-loop structure (e.g. NLMS adaptive filter). Closed-loop structures may face to convergence and tracking problems, especially with long filters. Open-loop structures are suitable for fast-tracking and are easily implemented in the frequency domain. We have chosen a frequency domain implementation of such open-loop structures. The major drawback of frequency domain implementations is the delay of the signal, usually higher than in time domain implementations; however, in teleconference context, delay issues are less critical than in other applications like mobile telephony. In our implementation, the adaptive behaviour of the post-filter is controlled by the smoothing constants of the spectral quantities involved. Moreover, the filter length can be easily modified: this possibility yields an additional degree of freedom for the design. Each block p of the three signals $x(t)$, $d(t)$ and $e(t)$ of length NFFT=512 (with overlapping factor=50%) is windowed (Hanning), then a FFT is performed, which produces the short term Fourier transforms $X(p,f)$, $D(p,f)$ and $E(p,f)$. The second order spectral densities are estimated according to the general formula:

$$\gamma_{xy}(p,f) = \alpha.\gamma_{xy}(p-1,f) + (1-\alpha).X(p,f)Y^*(p,f)$$

The forgetting factor α has been chosen equal to 0.7261, which corresponds to a time constant of 50 ms for a sampling frequency of 16 kHz. Those spectral quantities are used for the computation of the filter transfer function, according to the algorithms which will be described in the next section. Various lengths of the filter are obtained by going in the time domain (IFFT), then applying a rectangular window of appropriate size, and finally going back in the frequency domain using a FFT of size 2*NFFT (this choice makes possible the use of filter sizes up to NFFT while keeping linear convolution properties). The output of the filter is synthesized using an Overlap Add technique.

3. POST-FILTER COMPUTATION

Using similar guidelines as in [2], we have derived three different algorithms which obey the general structure shown Figure 1.

3.1. The Wiener Post-Filter

This algorithm is directly derived from [2], except that the time domain adaptive filter is replaced by a Wiener filter. The post-filter coefficients minimize the mean square error between $u(t)$ and $e(t)$, where $u(t)$ stands for the signal $x(t)$ filtered by H. The expression of the post-filter is:

$$H(p,f) = \frac{\gamma_{xe}(p,f)}{\gamma_{xx}(p,f)}$$

where γ_{xx} and γ_{xe} denote respectively the power spectral density of the microphone signal $x(t)$, and the cross spectral density of $x(t)$ and the compensated signal $e(t)$. During single talk, when the far-end speaker is active, the filter tries to approximate the residual echo by the echo signal. Assuming that the echo canceller delivers some attenuation (e.g. 10 dB), this leads to attenuate echo frequencies which have more power than the residual echo. During single talk when the near-end speaker is active, the signals $x(t)$ and $e(t)$ are identical, therefore the post-filter does not affect the signal $e(t)$. During double talk, frequencies which correspond mainly to the echo are attenuated while other frequencies are less affected.

3.2. The Overweighted Wiener Post-Filter

Good performance of the Wiener post-filter relies on an efficient echo canceller. The higher the attenuation provided by the echo canceller is, the more selective the post-filter is. In the case when the echo canceller is inefficient, the post-filter reduces to an all-pass filter. Since the echo canceller length is chosen small for complexity reasons, its efficiency is limited. One way of getting rid of this drawback is to overweight the influence of the echo in the filter input signal, according to:

$$H(p,f) = \frac{\gamma_{xe}(p,f)}{\gamma_{xx}(p,f) + A.\gamma_{dd}(p,f)}$$

where γ_{dd} stands for the power spectral density of the echo estimated by the echo canceller, and $A > 0$. This approach can be seen as a particular version of the solution proposed in [3]. Its behaviour is similar to the Wiener post-filter described above; it yields an additional degree of freedom to improve the efficiency of the post-filter.

3.3. Post-Filter using noise reduction techniques

This approach is based on a general concept of disturbance reduction, the residual echo being one of these disturbances. Therefore a large number of methods similar to noise reduction can be used to compute the post-filter. Our approach is based on a method of spectral subtraction used in noise reduction techniques [4]. The post-filter transfer function can be expressed as:

$$H(p,f) = \frac{SER(p,f)}{1 + SER(p,f)}$$

where SER stands for Signal to Echo Ratio. The signal at the microphone output $x(t)$ and the "direct" echo $d(t)$ estimated by the echo canceller are used to compute this ratio. For this algorithm, the post-filter phase is set to 0. In that way, the effect of the post-filter can be interpreted as a pure frequency weighting of the compensated signal $e(t)$.

4. EXPERIMENTAL RESULTS

We have carried out simulations with the three algorithms above. The experimental conditions were carefully controlled, in order to evaluate essentially the behaviour and the performance of the post-filters corresponding to these

algorithms. "Ideal" acoustic echo cancellation was performed by using directly the beginning of the echo path impulse response instead of the adaptive filter impulse response; thus, convergence problems linked with particular adaptive filtering algorithms were avoided (experiments showed that the use of a true adaptive filter impacted essentially on the beginning of the simulations, and that the overall performance was otherwise equivalent). We have used impulse responses measured in a teleconference room, with several values of the overall echo loss through the echo path. The length of the "ideal" acoustic echo canceller was fixed at 512 samples for all experiments (which corresponds to 32 ms with the sampling frequency Fs=16 kHz). The corresponding average ERLE was about 14 dB. The criteria used to evaluate the performance of the algorithms were:
- the Additional Echo Return Loss Enhancement (AERLE) provided by the post-filter (the AERLE does not include the usual ERLE provided by the echo canceller itself);
- the distortion of the near-end speech signal, measured by the cepstral distance between this signal and the same signal filtered by the post-filter.

4.1. Comparison of the three algorithms

All simulations considered in this subsection were carried out using a post-filter length N=20 in order to keep in agreement with similar choices made in previous works [2,3]. Figure 2 shows the AERLEs provided by the three algorithms.

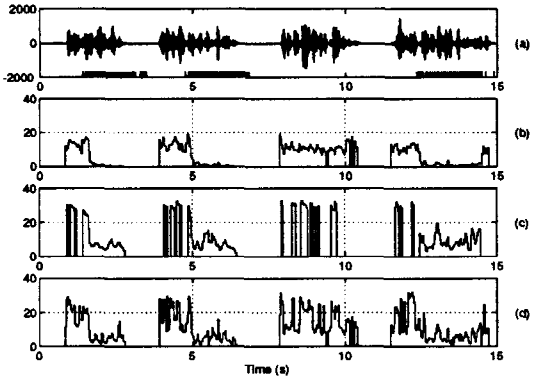

Fig.2 AERLE returned by the three algorithms.
(a) Echo signal (b) Wiener post-filter
(c) Overweighted Wiener post-filter
(d) Noise reduction post-filter

The figure 2a represents the echo signal at the microphone output; the double talk events are indicated by the stepped curve. Figures 2b, 2c and 2d show the AERLE obtained with the classical Wiener filter, the overweighted Wiener filter (A=30) and the noise reduction algorithm respectively. The high value of A used in the second algorithm was found necessary to get equivalent echo reduction as the one obtained with the noise reduction algorithm. It can be seen that the last two algorithms provide much larger echo attenuation than the Wiener filter. Informal listening tests confirmed that the residual echo is hardly audible for these two algorithms.
Figure 3 shows the near-end speech distortion provided by the three algorithms. The figure 3a represents the local speech. The figures 3b, 3c and 3d show the distortion obtained with the classical Wiener filter, the overweighted Wiener filter and the noise reduction algorithm, respectively, using the same parameter settings as above. The distortion provided by the noise reduction algorithm and the overweighted Wiener filter is larger than the distortion of the classical Wiener filter. Nevertheless, informal listening tests showed that the distortion had a moderate impact on speech quality (the local speech was fully understandable and the speaker's voice was fully recognizable).

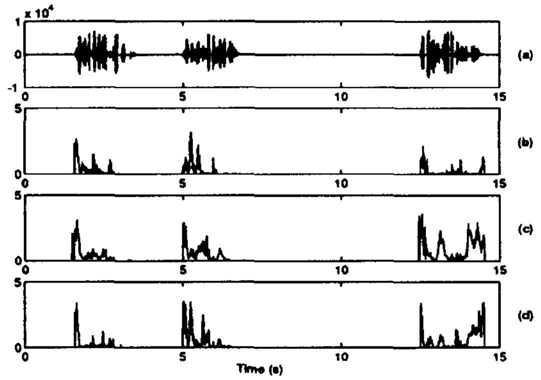

Fig.3 Cepstral distance measured for the three algorithms.
(a) Near-end signal (b) Wiener post-filter
(c) Overweighted Wiener post-filter
(d) Noise reduction post-filter

4.2. Effect of the Post-Filter length

The effect of different post-filter lengths was evaluated for the three algorithms. In this paper, we only present the results obtained with the third one (noise reduction); the other algorithms led to similar results. The use of different lengths for the post-filter leads to different smoothings of its transfer function; this effect can be seen in Figure 4. The upper part shows the short term spectra of the near-end speech and of the echo for a particular block p; the lower part shows the corresponding transfer functions of the filter for N=20 and N=512.

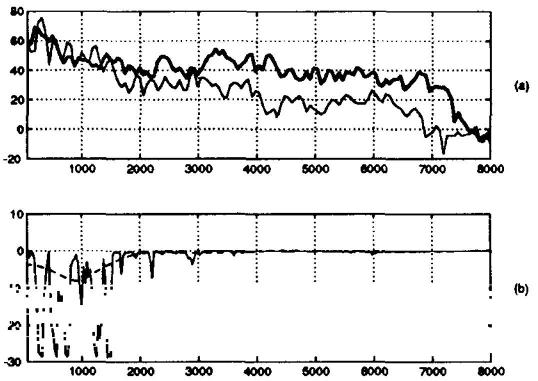

Fig.4 Postfilter transfer functions for different lengths.
(a) Echo (thin) and near-end signal (bold) psds
(b) Transfer function: N=512, N=20 (- -).

It appears clearly that, in the low frequencies where near-end speech and echo have similar levels, the shorter filter transfer function is smoothed and provides moderate amounts of attenuation (less than 10 dB). The longer filter provides much higher attenuation (up to 30 dB) and is more selective.

The effect on residual echo reduction is presented in Figure 5. During single talk (echo), the filter length does not influence the average attenuation. This is not the case in double talk situations where the longer filter is more effective: the shorter post-filter performs an average attenuation of 8.5 dB whereas the longer post-filter provides an average attenuation of 17.3 dB. This fact can be explained by the different attenuation characteristics of the transfer functions discussed above.

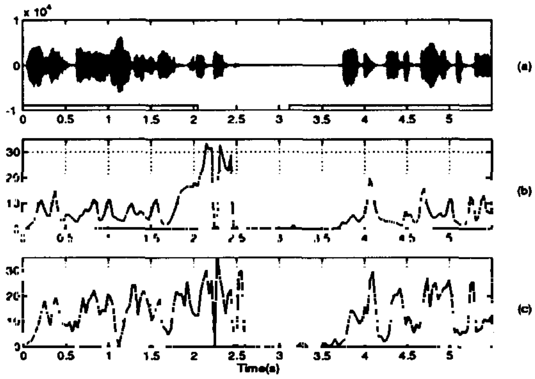

Fig.5 AERLE obtained for different post-filter lengths. (a) Echo signal (b) AERLE N=20 (c) AERLE N=512.

The effect on distortion of near-end speech is shown in Figure 6. The average distortion provided by the longer filter is somewhat smaller than the one produced by the shorter filter.

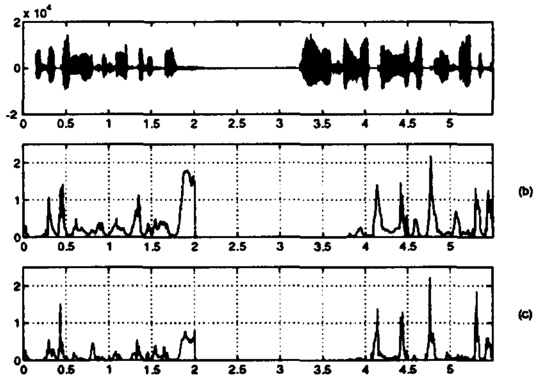

Fig.6 Cepstral distance for different post-filter lengths. (a) Near-end speech (b) Cepstral distance N=20 (c) Cepstral distance N=512.

The two filters, which have different behaviours, lead to different kinds of distortion which were subjectively observed during informal listenings. In double talk situations, the shorter filter produced a fairly stationary high-pass filtering of speech, which was found less annoying than non-stationary spectral variations of speech provided by the longer filter, although this latter one did not create noticeable high-pass filtering. These observations lead us to conclude that the use of the cepstral distance as a distortion measure may not be sufficient to explain the subjective performance.

5. CONCLUSION

Three post-filtering algorithms for residual acoustic echo reduction have been presented and compared. They have been implemented in an open-loop structure in the frequency domain which eliminates the potential problems experienced with adaptive methods. Those algorithms have different performance in terms of residual echo attenuation and near-end speech distortion. The noise reduction approach was found the most efficient. From our study, we can deduce that the post-filter length can be adjusted to get a trade-off between echo attenuation and spectral distortion, which can result in optimal sujective performance.

REFERENCES

[1] E. HÄNSLER, "The hands-free telephone problem: a second annotated bibliography update", International Workshop on Acoustic Echo and Noise Control, Roros, NORWAY, June 1995, pp. 107-114.

[2] R. MARTIN and J. ALTENHÖNER, "Coupled adaptive filters for acoustic echo control and noise reduction", ICASSP'95, DETROIT,USA, pp. 3043-3046.

[3] R. MARTIN and S. GUSTAFSSON, "The echo shaping approach to acoustic echo control", Speech Communication, special issue on acoustic echo and noise control, 20(3-4), January 1997.

[4] S. F. BOLL, "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. on ASSP, Vol. 27, No. 2, , April 1979, pp. 113-120.