Acoustic Echo Cancellation and Noise Reduction in the Frequency-Domain: A Global Optimisation

F. Capman, J. Boudy, P. Lockwood
Matra Communication, Speech Processing Department
rue J.P. Timbaud, 78392 Bois d'Arcy Cedex, BP 26, France
phone: (+33-1) 34-60-76-84; fax: (+33-1) 34-60-88-32
e-mail: fcapman@matra-com.fr

ABSTRACT

The design of an efficient and robust hands-free system is now required by the growth of mobile radio and teleconference communications. The use of Frequency-Domain Adaptive Filters in the context of acoustic echo cancellation has been extensively studied in the literature. These algorithms are well-suited for long impulse response modeling and for correlated input signals like speech. A global optimisation of a frequency-domain acoustic echo cancellation algorithm with noise reduction is presented in this paper. This optimisation leads to both reduced complexity and improved performances when compared to classical cascaded structures.

1. INTRODUCTION

Adaptive transversal filters designed for acoustic echo cancellation schemes are confronted to several well-known difficulties. First, the number of taps needed to model the impulse response of an acoustic echo path can be very large. Moreover, speech signals are highly correlated and nonstationary signals, and finally the signal picked up by the microphone is corrupted by the near-end speech, and by the ambient noise. Dealing with long impulse responses, and designing adaptive filtering algorithms for highly correlated speech signals, has led to various solutions. Frequency-domain LMS algorithms are among the most efficient and investigated adaptive filters in acoustic echo cancellation. They are well-designed for handling the problem of long impulse responses, and high correlation degree of speech signals, when compared to classical time-domain LMS algorithm. Whereas some studies have been published on the improvement of interference-robustness of time-domain algorithms, little has been done on improving frequency-domain schemes.

In this paper, we propose to modify the adaptation process of the frequency-domain acoustic echo canceller by incorporating a spectral subtraction step aiming at better noise robustness properties. The possibility of designing an optimised and compact combination of an acoustic echo canceller with a noise reduction system for speech enhancement is also studied. Various double-talk detection methods are also proposed in this context. The complete system appears to be a valuable candidate to the realisation of a car hands-free function, characterised by a good trade-off between computational complexity, low transmission delay, echo cancellation and noise reduction properties.

2. FREQUENCY-DOMAIN BASELINE SOLUTIONS

In this section, the main properties of frequency-domain adaptive filters are briefly presented. The time-domain Block-LMS algorithm is based on a block updating procedure of the filter weights, instead of a sample-by-sample one. The gradient is therefore estimated on a block-by-block basis and becomes a cross-correlation between the input and error signals on the corresponding block. The convolution and the gradient estimation can be efficiently achieved in the frequency-domain using Fast Fourier Transform and a fast-convolution procedure like overlap-and-save or overlap-and-add. Whereas, convergence speed of time-domain gradient descent-based algorithms is driven by the eigenvalue spread of the input autocorrelation matrix, improved convergence speed is easily obtained with frequency-domain algorithms by normalising the input power spectrum. This normalisation is generally achieved with an exponentially weighted estimation procedure at each frequency bin. Because of overlap sectioning procedures, Fourier transforms are calculated on size equal to twice the length of the impulse response, 2N. The exact constrained implementation requires additional IFFT and FFT for applying the classical constraint to the time-domain gradient, while computational savings result from the unconstrained version proposed by Mansour & Gray in [1], and from the cosine-constrained version proposed by Sommen & al in [2], well-suited for modeling typical decreasing echo path impulse responses. Therefore, apart from all these attractive features, the input/output propagation delay introduced by block-by-block processing can be redhibitory for a practical implementation. A particular sectioning of the impulse response has been proposed by Soo & Pang in [3], and consists in dividing the impulse response into M sub-blocks, which results in both shortened delay and FFT size. The use of Fast Hartley Transform can also lead to some complexity reduction as described by Wong & Kwong in [4]. In order to further improve the performances and reduce the delay, we are using a modified algorithm (MDFO, Multi-Delay-Frequency-domain algorithm with Overlap) which processes overlapped input blocks by more than half the FFT size. The transmitted residual echo samples are then extracted from the last output residual error block. This implementation differs from [5], where a Weighted-OLA procedure is used to calculate the residual echo from successive filtered blocks corresponding to the same output sample.

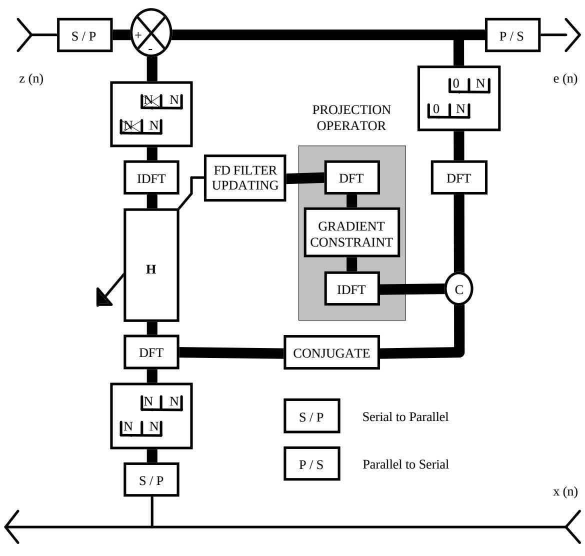

Fig. 1: OLS-based Frequency-Domain Adaptive Filter.

A large family of frequency-domain algorithms can be derived. Thanks to the relatively short impulse response length measured in a car cockpit, a good trade-off for a car hands-free application between low complexity, low delay and good performances can be achieved by frequency-domain adaptive algorithms.

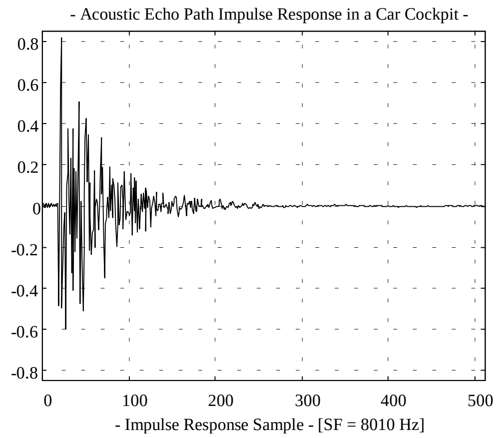

Fig. 2: Typical echo path impulse response in a car.

3. AN INTERFERENCE-ROBUST IMPLEMENTATION

3.1. Double-Talk Detection

Adaptive echo cancellers are greatly disturbed by simultaneous talk of near-end and far-end speakers, if the adaptive filter is continuously updated. When double-talk occurence is detected, several strategies can be applied. Nevertheless, a secure solution is to freeze the adaptation process, preventing the adaptive filter from identifying an erroneous acoustic echo path. Classical double-talk detection methods feature slow response properties and are often unable to distinguish a double-talk period from a variation of the acoustic echo path. Moreover, they are inadequate for hands-free systems in noisy environments. We propose two different methods for double-talk detection: the first one is based on spectral distortion measures such as an Itakura-Saito distance on two AR models or a cepstral distance, the second one results from the estimation of a coherence function. These methods can either be applied directly on the echo and loudspeaker signals, or on the raw echo and the estimated echo signals.

3.2. Modified Adaptation for Noise Robustness

A second source of interference is the acoustic background noise, which can induce low signal-to-noise ratios for the near-end speech in a car environment. Adaptive filters must incorporate some adaptation control to prevent them from diverging. Several modifications of the classical time-domain LMS-based algorithms have been proposed aiming at robustness to noise interference, [6], [7]. We propose to modify the adaptation process of frequency-domain adaptive algorithms, so as to achieve the required robustness to noise interference. This modification is based on spectral subtraction, which has been already investigated in speech enhancement and adverse-environment speech recognition. Spectral subtraction methods can easily be incorporated in frequency-domain algorithms, so as to adapt the adaptive filter weights with a noise-free residual echo signal. The frequency-domain weight vector updating relation is then given by:

$$\mathring{\mathbf{H}}(k+1) = \mathring{\mathbf{H}}(k) + 2.\mathbf{F}.\mathbf{G}.\mathbf{F}^{-1}.\mathbf{m}(k).\mathring{\mathbf{X}}^{H}(k).\mathbf{SpS}\left(\mathring{\mathbf{E}}(k)\right)$$

where **F** is the DFT matrix, **G** the gradient constraint matrix, **m** the diagonal power-normalised step-sizes matrix, and **SpS** stands for the Spectral Subtraction operator.
The interference noise is estimated in the frequency-domain during far-end and near-end speech pauses, and subtracted from the corrupted residual echo. The above modification of the adaptation process only slightly increases the computational complexity of the overall structure, since no additional discrete Fourier transform is needed. The modified adaptation part is depicted in fig. 3.

4. NOISE REDUCTION

The modified adaptation process can also be extended to noise reduction for speech enhancement, with only one extra inverse discrete Fourier transform to recover the enhanced near-end speech. Compared to classical cascaded structures (noise reduction followed by echo cancellation or echo cancellation followed by noise reduction), this new scheme achieves a better ERLE and better noise reduction.

We used the proposed non-linear spectral subtraction (NSS) in [8], which provides distorsion-free enhanced speech for an approximately 10 dB SNR improvement.

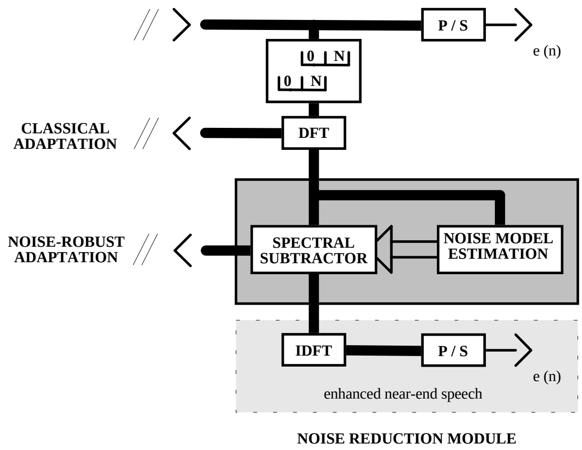

Fig. 3: Globally optimised echo cancellation and noise reduction.

5. SIMULATION RESULTS

This module has been tested on real speech files corrupted by various noise signals recorded within the European FREETEL project.
The acoustic front-end is a classical hands-free car kit, from which the impulse response shown on fig. 2 has been identified. The parameters of the adaptive filter were tuned to give quasi-optimal results on noise-free speech signals, and we obtained a maximum ERLE of approximately 35 dB. Speech signals were (1 male and 1 female speakers) corrupted with different noises at SNR from 20 dB to -5 dB. Some results are given on figures 4-7 in terms of SNR and deviation. The average deviation between the identified acoustic echo path at time n, and the optimal aoustic echo path is given by:

$$DEV_{dB} = 10.\log_{10}\left[\frac{\left\|H_{opt} - H\right\|^2}{\left\|H_{opt}\right\|^2}\right]$$

Unfortunately, for real echo signals the optimal impulse response is not available, and we supposed that the initial speech echo signal before adding noise is given by $H_{opt} * x$. We then obtained the approximated deviation by:

$$DEV_{dB} = 10.\log_{10}\left[\frac{E\left[\left(H_{opt} - H\right) * x\right]^2}{E\left[H_{opt} * x\right]^2}\right]$$

The different curves show an improved performance for the modified adaptation based echo canceller (dashed line) when compared to the classical one (solid line) , even for nonstationary noises.

6. CONCLUSION

In this paper, a new complete hands-free function based on frequency-domain processing, both for acoustic echo cancellation and noise reduction, has been presented. The adaptive echo canceller incorporates a double-talk detection and a modified robust to noise-interference updating relation, and is jointly optimised with a noise reduction technique based on non-linear spectral subtraction (NSS). Its performances have been assessed in a car environment context.

ACKNOWLEDGEMENTS

The speech database was recorded within the EC “ FREETEL ” project under contract number n° 6166. Part of this work was supported by the A.N.R.T. under grant n° 471/92.

REFERENCES

[1] MANSOUR D., and GRAY A.H.Jr., *"Unconstrained Frequency-Domain Adaptive Filter"*, in IEEE Trans. on ASSP, Vol. 30, No 5, pp 726-734, Oct. 1982.

[2] SOMMEN P.C.W., GERWEN P.J.V., KOTMANS H.J., and JANSSEN A.J.E.M., *"Convergence Analysis of a Frequency-Domain Adaptive Filter with Exponential Power Averaging and Generalized Window Function"*, in IEEE Trans. on CS, Vol.CS-34, No 7, pp 788-798, July 1987.

[3] SOO J.S., and PANG K.K., *"Multidelay Block Frequency-Domain Adaptive Filter"*, in IEEE Trans. on ASSP, Vol.38, No 2, pp 373-376, Feb. 1990.

[4] WONG T.W., and KWONG C.P., *"Adaptive Filtering Using Hartley Transform and Overlap-Save-Method"*, in IEEE Trans. on SP, Vol.SP-39, No 7, pp 1708-1711, July 1991.

[5] AMRANE O.A., MOULINES E., CHARBIT M., and GRENIER Y., *"Low-Delay Frequency-Domain LMS Algorithm"*, in Proc. IEEE ICASSP’92, pp IV-9-IV-12, 1992.

[6] SUGIYAMA A., *"An Interference-Robust Stochastic Gradient Algorithm With A Gradient-Adaptive Step-Size"*, in Proc. IEEE ICASSP’93, pp III-539-III-542, 1993.

[7] KINJO S., *"New Interference Robust Block Adaptive Filter with Correlated Signal Input"*, in Proc. IEEE ICASSP’94, pp III-85-III-88, 1994.

[8] LOCKWOOD P., and BOUDY J., *"Experiments with a Nonlinear Spectral Subtractor (NSS), Hidden Markov Models and the projection, for robust speech recognition in cars"*, in Speech Communication, Vol. 11, Nos 2-3, pp 215-228, June 1992.

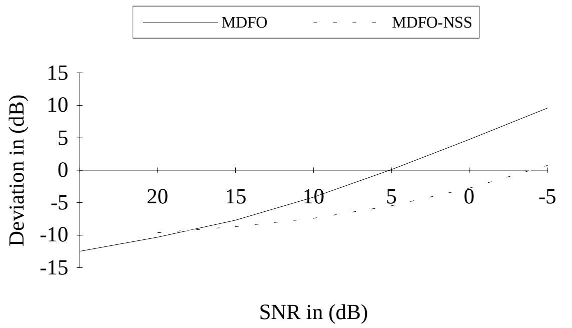

Fig. 4: Mean Deviation (dB) - Trafic Noise - Car stopped.

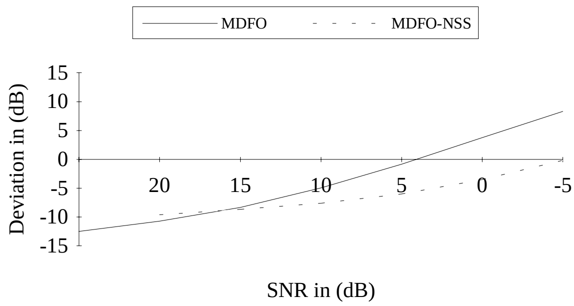

Fig. 5: Mean Deviation (dB) - Rain Noise - Car stopped.

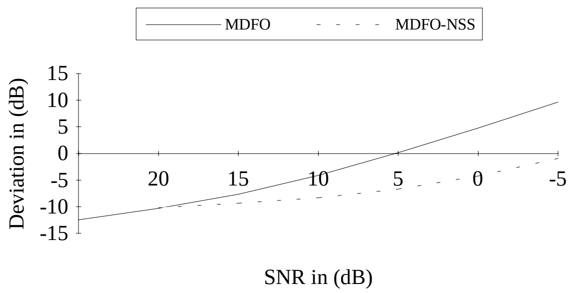

Fig. 6: Mean Deviation (dB) - Car Noise - 130 km/h.

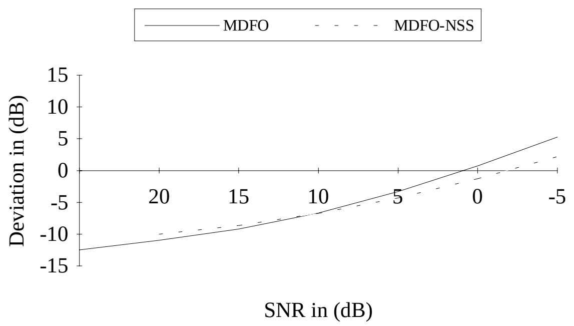

Fig. 7: Mean Deviation (dB) - Coktail Party Noise.