COMBINED SYSTEMS FOR NOISE REDUCTION AND ECHO CANCELLATION

Christophe Beaugeant, Pascal Scalart
CNET DIH/DIPS, 2, Av. Pierre Marzin, 22307 Lannion Cedex, FRANCE
Tel: +30 96051481; Fax: +33 2 96053530
e-mail: {Christophe.Beaugeant, Pascal.Scalart}@cnet.francetelecom.fr

ABSTRACT

The performance of three different combined systems for noise reduction and acoustic echo cancellation is compared in this paper. This comparison is made in the specific context of hands-free radiotelephony in cars, where both noise and acoustic echo highly degrade speech quality. This article shows that the best system is based on a cascaded combination of a conventional echo canceller with a filter reducing noise and residual acoustic echo. Nevertheless, recent novel approaches based on the use of a single Wiener filter are to be efficient enough in the specific context of hands-free sound pick-up in cars.

1 INTRODUCTION

Development of hands-free telephony for mobile radio-communications increases the need of efficient front end speech processing techniques. In the last few years, techniques combining both acoustic echo cancellation and noise reduction have been proposed and implemented in hands-free mobile kits ([1], [2]). However, through effectively reducing both noise and echo. Those techniques are not the best solution to the problem of speech enhancement. Indeed, they are based on the cascade of two locally optimized filters (adaptive echo canceller and noise reducer) which do not lead to a unique global solution. New approaches have recently been proposed ([3]) in order to obtain a unique global filter simultaneously reducing echo and noise.

The aim of this paper is to compare such global algorithms with previous cascaded systems in car environment. In section 2, the three solutions tested will be described. The comparison of performance is proposed in section 3 regarding objective measurements and subjective tests.

2 DESCRIPTION OF COMBINED SYSTEMS STUDIED

Three different systems are compared echo cancellation and noise reduction.

2.1 Cascaded combination.

The first system is a 'conventional' cascaded system of an independent echo canceller and a noise reduction filter, as illustrated in Fig. 1 and proposed in [2]. It was chosen as it proved to be the optimal theoretical solution of echo and noise reduction. However, in practical applications, the echo cancellation algorithm should be robust to noisy conditions if we choose this cascaded solution. This is the case of the 2nd order Affine Projection Algorithm (APA2), as shown in [2].

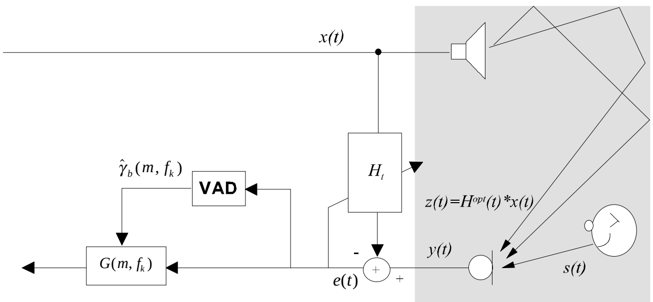

Fig.1. 'Conventional' cascaded system

The specific algorithms we used are the APA2 for the acoustic echo canceller [4] and the open-loop Wiener filter implemented in the frequency domain [5] for noise reduction. With the notation of Fig. 1, The APA2 filter coefficients are updated according to

$$e(t) = y(t) - H_{t-1}^T X_t \qquad U_t = X_t - \frac{X_t^T X_{t-1}}{X_{t-1}^T X_{t-1}} X_{t-1};$$

$$H_t = H_{t-1} + \mu \frac{e(t)}{U_t^T U_t} U_t$$

The input sequence vector is represented by $X_t = \left[x(t)\ x(t-1)\cdots x(t-L+1)\right]^T$ where L is the filter length and $H_t = \left[h_t^0\ h_t^1 \cdots h_t^{L-1}\right]^T$ stands for the filter coefficients vector.

The Wiener filter expressed in the frequency domain is given by

$$G(m, f_k) = \frac{SNR(m, f_k)}{1 + SNR(m, f_k)} \quad \text{for} \quad k = 0,1,\cdots, NFFT-1$$

where the subscript m stands for the number of the current block, and f_k for the k^{th} beam of frequency. Parameter *SNR*, standing for the Signal to Noise Ratio is computed as proposed in [5]:

$$SNR(m, f_k) = \beta \cdot \frac{\left|E(m, f_k)\right|^2 - \hat{\gamma}_b(f_k)}{\hat{\gamma}_b(f_k)} + (1-\beta)\frac{G^2(m-1, f_k)\left|E(m-1, f_k)\right|^2}{\hat{\gamma}_b(f_k)}$$

where $E(m, f_k)$ is the Short Time Fourier Transform (STFT) at the frequency f_k of the frame *m* of *e(t)*, and $\hat{\gamma}_b(f_k)$ the estimation of the noise power spectrum density (psd) computed during non vocal activity periods.

From now on, this solution will be called either cascaded system or CASC.

2.2 Global filtering.

The second system corresponds to Wiener global filtering of all disturbances [3] (see Fig. 2). This filter, which will be referred to GLO as global, is implemented in the frequency domain. The expression of its frequency response for frame *m* at frequency f_k is given by :

$$G(m, f_k) = \frac{SNR(m, f_k)}{1 + SNR(m, f_k)} \cdot \frac{SER(m, f_k)}{1 + SER(m, f_k)} \text{ for } k = 0,1,\cdots, NFFT-1$$

where parameters $SNR(m, f_k)$ and $SER(m, f_k)$ stand respectively for the Signal to Noise Ratio and the Signal to Echo Ratio. These ratios are computed as in [5]:

$$SNR(m, f_k) = \beta_1 \cdot \frac{\left|Y(m, f_k)\right|^2 - \hat{\gamma}_p(m, f_k)}{\hat{\gamma}_b(f_k)} + (1-\beta_1) \cdot \frac{G^2(m-1, f_k)\left|Y(m-1, f_k)\right|^2}{\hat{\gamma}_b(f_k)}$$

$$SER(m, f_k) = \beta_2 \cdot \frac{\left|Y(m, f_k)\right|^2 - \hat{\gamma}_p(m, f_k)}{\hat{\gamma}_z(m, f_k)} + (1-\beta_2) \cdot \frac{G^2(m-1, f_k)\left|Y(m-1, f_k)\right|^2}{\hat{\gamma}_z(m, f_k)}$$

In this last expression, $Y(m, f_k)$ is the STFT of y(t), the microphone signal, and $\hat{\gamma}_p(m, f_k)$ is defined by:

$$\hat{\gamma}_p(m, f_k) = \hat{\gamma}_b(f_k) + \hat{\gamma}_z(m, f_k)$$

$\hat{\gamma}_z(m, f_k)$ is the estimated acoustic echo psd, calculated in the frequency domain, using :

$$\hat{\gamma}_z(m, f_k) = \frac{\gamma_{xy}(m, f_k).\gamma_{yx}(m, f_k)}{\gamma_x(m, f_k)}$$

where $\gamma_x(m, f_k)$ and $\gamma_{xy}(m, f_k)$ are the loudspeaker signal psd and the cross power spectrum between the loudspeaker and the microphone signals, respectively. Both are calculated by first order IIR filtering of *x(t)* and *y(t)*.

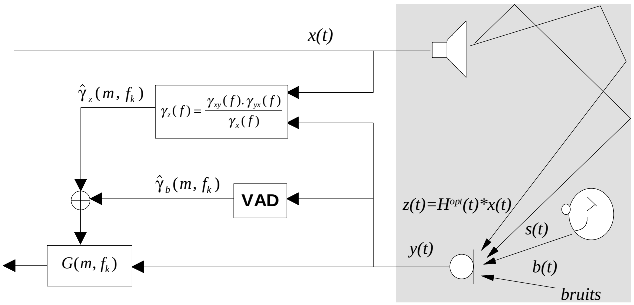

Fig. 2. Wiener global filtering

It may be noted that this kind of solution leads to using a single filter in the frequency domain, which greatly simplifies the computation load regarding solutions with two filters.

2.3 Optimized combination.

The last solution (Fig. 3) is an optimized system in which the APA2 algorithm is associated with the previous Wiener global filter described in section 2.2. In this case, the Wiener filter is used to reduce noise and the residual echo remaining after echo cancellation. Such a system is similar to those proposed in [6]. However, our solution is based on an open-loop implementation of a global post-filter. Filter G acts as a frequency selective adapter of a conventional voice echo suppresser based on variable loss which is usually needed in practical implementation. In the following, this solution will be called iether optimized system or APA+GLO.

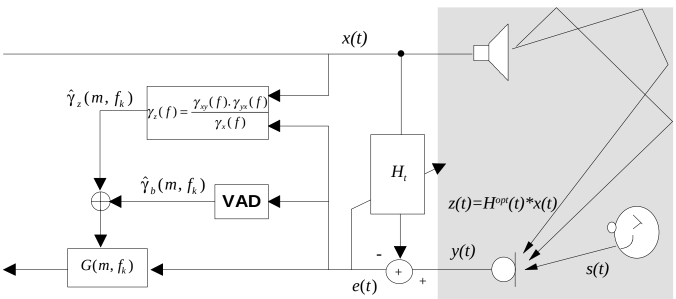

Fig. 3. Optimized system

3 COMPARED PERFORMANCE LEVEL.

To compare the previously described systems, several separate recordings of speech, echo and noise signals at a sampling frequency of 8 kHz were made in different car environments (different speeds, roads and cars). These recordings were filtered by the different systems presented in section 2, to compare the performance of the algorithms, using objective measurements and subjective tests.

The choice of the different algorithms tends to limit the difference induced by implementation. As a result, all the filters in the frequency domain analyse frames of 16 ms Hanning windowed signals with an overlapping of 50%. The STFT are also computed with *NFFT*=256. The length of the APA filters is chosen as *L*= 256.

It may be seen that with such parameters the computer load of the cascaded solution and the optimized solution are almost equivalent, whereas the computer load of GLO is about 80 times lower.

3.1 Objective measurements.

Objective measurements were made on the separate recordings described above. A typical sequence of noisy speech with echo is given in Fig. 4 for an averaged segmental SNR of 6 dB and a segmental SER evaluated during double-talk periods of 4 dB. The stepped curve represented under the time domain signal indicates the presence of acoustic echo in the microphone signal.

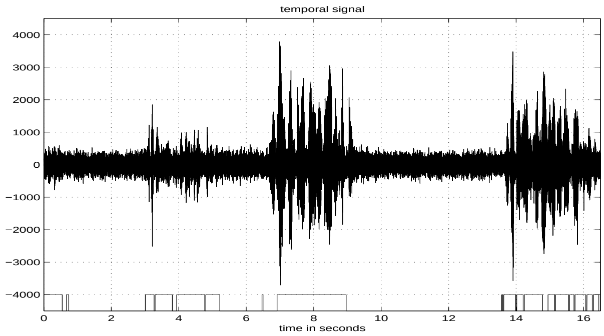

Fig. 4. Temporal speech signal

Two main criteria were used to compare the different systems, namely :

➾ the echo return loss enhancement (ERLE) which expresses the echo attenuation introduced by each device;

➾ the basilar distance (in dB) computed between the original near-end speech signal and the same signal filtered by each device. The basilar distance is evaluated using a Perceptual Objective Measurement system [7] which transforms the signal in the basilar domain according to a mono aural human ear modeling.

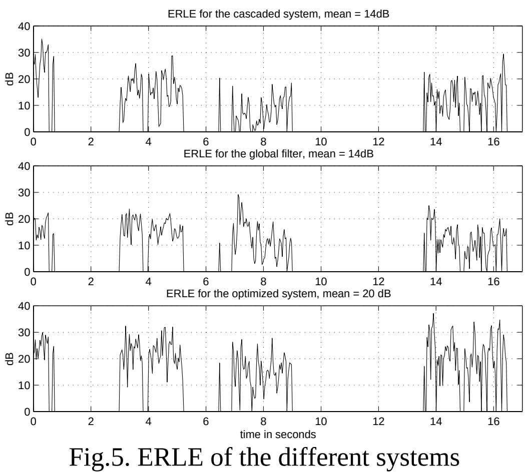

Fig.5. ERLE of the different systems

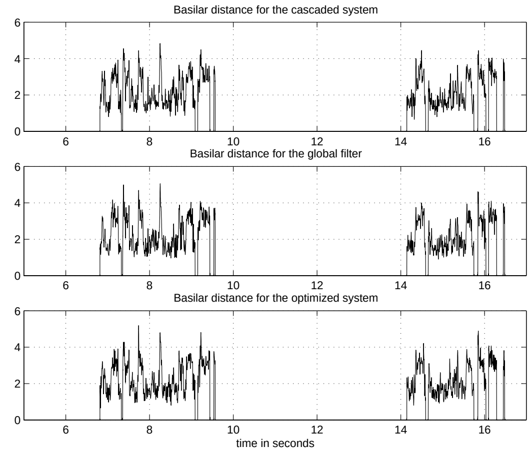

Fig. 6. Near-end speech distortions

Observing Fig. 6 shows that the distortion of the near-end speech signal is about the same independently of the system. The distortion is mainly due to the noise reduction method which is the same in all three approaches. In addition, informal listening tests show that this distortion is slightly audible but remains quite limited. For the three systems, the corresponding output signal sounds good with respect to our context. The main differences can be noted on the ERLE curves. The attenuation of the acoustic echo obtained using the cascaded system is greater during non vocal activity of the near-end speech than that obtained using the global filter. Conversely, during double talk, as the adaptive step of the APA2 algorithm is frozen, the echo attenuation is greater using the global system than with cascaded one. The optimized device yields a better attenuation in all cases (double talk, acoustic echo alone). It appears to be the most efficient solution with high echo reduction and limited distortion.

3.2 Subjective tests.

3.2.1 Experiment.

A subjective test campaign was conducted to confirm the previous results obtained through informal listening and objective measurements.

The test was a Comparison Category Rating test (CCR) [8], which allows the comparison of each processed signal with a reference signal. A series of recordings in different environmental conditions (SNR and SER) was chosen and filtered by each system. The resulting signals were then compared using the CCR test procedure:

As the system described is section 2.3 is supposed to be the most efficient one according to informal listening tests and objective measurements, the reference chosen is the signal *s(t)* filtered by the APA+GLO combination, $s_{ref}(t)$. The signal *s(t)* filtered by the GLO filter $s_{GLO}(t)$ and the cascaded system filter $s_{CASC}(t)$ are then compared by the subjects with $s_{ref}(t)$. The evaluation integer scale ranges from -3 to +3. The mark of -3 means that $s_{GLO}(t)$ (or $s_{CASC}(t)$) is far worse than $s_{ref}(t)$, 0 means that they are almost equivalent and 3 means that they $s_{GLO}(t)$ (or

$s_{CASC}(t)$) is far better than $s_{ref}(t)$. Sixteen listeners participated in this subjective test using telephone handset and 12 different configurations were submitted to the auditors (which means 24 different pairs of filtered signals to compare).

3.2.2 Results.

An analysis of variance (ANOVA, [9]) was made from the comparison marks obtained. It appears that the relative performance of the three algorithms was highly dependent on SER and SNR. The graph in Fig. 7 shows this influence through a third order of interactivity between the algorithm type, segmental SER and segmental SNR. For different segmental SNRs and SERs computed during double-talk and single-talk periods respectively, these graphs give in the *y* axis the average comparative notes of the cascaded solution and of the global one to the optimized algorithm.

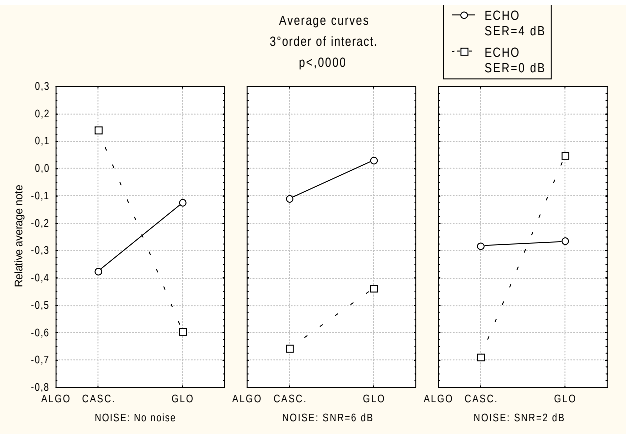

Fig. 7. Influence of SER and SNR.

The first result shown in these graphs is the relatively low difference between the algorithms (An average mark of -0.5 only shows that an algorithm is slightly worse than the optimized system). The algorithms are even almost equivalent in several cases.

The performance of the cascaded system decreases as the noise power increases. This can be explained by the fact that the echo cancellation convergence is disturbed by background noise. In this case, the global post-filter in the optimized system is useful to reduce the residual echo due to the poor performance of echo cancellation in a noisy environment.

These curves also show that the Global filter performance is not much different from that of the optimized system when SER and SNR are almost equal (See the graphs for the couples (SER,SNR)=(4 dB, 6 dB) and (0 dB, 2dB)). This shows that when ENR is low enough, the global filter is efficient. In the case of high ENR, the pre-filtering proposed in the optimized system permits to reduce the echo before the global filter. Hence the ENR obtained is low enough for optimal performance of the Wiener global filter.

4 CONCLUSION.

Three different solutions proposed for echo cancellation and noise reduction proved their relatively good performance. The different measurement campaigns show that using a global filter or a global post-filter allows the improvement of echo cancellation and noise reduction. Indeed, the performance of the cascaded system is lower than that of the GLO and APA+GLO systems, mainly regarding echo reduction in a noisy environment. One the other hand, these last two solutions are almost equivalent, with a slight preference for the APA+GLO in contexts where ENR is low. The choice between GLO and APA+GLO will mainly depend on the computer load required for a specific application, as global filtering is much cheaper than the optimized solution.

REFERENCES :

[1] F. Capman, J. Boudy, P. lockwood, ‘Acoustic Echo Cancellation and Noise Reduction in the Frequency Domain : a Global Optimisation.’ Proc. of EUSIPCO-96, pp. 29-32.

[2] Y. Guélou, A. Benamar, P. Scalart, ‘Analysis of two Structures for Combined Acoustic Echo Cancellation and Noise Reduction.’ Proc. of ICCASP’96, pp. 637-640.

[3] C. Beaugeant, V. Turbin, P. Scalart, A. Gilloire, ‘New Optimal Filtering Approaches for Hands-free Telecommunication Terminals.’ Signal Processing, vol. 64, issue 1.

[4] K . Ozeki, T. Umeda, ‘An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Subspace and its properties.’ Elec. and Com. in Japan, vol. J67-A, n°5, pp. 126-132, 1984.

[5] Y. Ephraïm, D. Malah, ‘Speech Enhancement using Optimal non-linear Spectral Amplitude Estimation.’ ICCASP’83, Boston, pp. 1118-1121.

[6] R. Martin, J. Altenhoner, ‘Coupled Adaptive Filters for Acoustic Echo Control and Noise Reduction.’ ICCASP’95, Detroit, USA, pp. 3043-3046.

[7] C. Clomes, M. Lever, Y.F. Dehery, 'A Perceptual Objective Measurment Sytem (POM) for the Quality Assessment of Perceptual Codecs', AES, 96th Convention, February 1994.

[8] Annex C of ITU-T Rec. P.800, 'Method for Subjective Determination of Transmission Quality', Geneva, November 1996.

[9] R.D. Bock, 'Multirate Statistical Methods in Behavioral Research.' 1975, New-York : McGraw-Hill.