ACOUSTIC ECHO CANCELLATION IN FREQUENCY DOMAIN
USING COMBINATIONS OF FILTERS

PACS: 43.60U

Azpicueta-Ruiz, Luis Antonio; Figueiras-Vidal, Aníbal; Arenas-García, Jerónimo
Deparment of Signal Theory and Communications, Universidad Carlos III de Madrid,
28911 Leganés-Madrid, Spain; lazpicueta@tsc.uc3m.es

ABSTRACT
In hands-free telephone or teleconference systems acoustic echo cancellation is important due
to the acoustic feedback coupling between the loudspeaker and the microphone. Subband
techniques have been developed for adaptive systems as a way to reduce their computational
requirements, and to accelerate their convergence when colored inputs -such as speech- are
present. However, since the filters in every subband are subject to a compromise between
speed of convergence, adaptability and residual noise, so it is the overall canceller. Variable
step-size adaptive filters have been proposed to alleviate this problem, but they are complex
and require some ‘a priori’ statistical knowledge about the filtering scenario for an efficient use.
In this paper, we study the applicability of a recent approach for adaptively combining adaptive
filters. The use of a convex combination of two NLMS filters with different step-sizes in every
subband, not only provides simultaneously fast convergence and low steady-state
misadjustment, but also allows to improve the overall performance when changes are only
present in some of the subbands. We present several experiments where these advantages are
illustrated in realistic echo cancellation scenarios.

INTRODUCTION
Acoustic echo cancellation (AEC) is a key component of hands free telephony, teleconference
and video conference systems. In these applications, there is an acoustic feedback coupling
between the loudspeaker and the microphone that degrades the communication, and AEC can
be used to minimize this effect.

In contrast to full-band adaptive filtering, subband adaptive techniques [1], [2] enjoy some
advantages that make them attractive for AEC. First, the convergence of a least mean-square
(LMS) or a normalized least mean-square (NLMS) filter can be very slow if the spectrum of the
reference signal has a large dynamic range (i.e., the input signal is highly colored), as it is the
case of speech. Second, usual impulse responses require very long filters for their identification,
thus filter adaptation in AEC is very computationally demanding. Subband adaptive techniques
reduce the high computational cost by efficient frequency operation, and improve the
convergence speed because the dynamic range is reduced in every subband.

Traditional subband adaptive filters present an important drawback: a delay is introduced in the
signal path as a consequence of the block processing necessary to transform the signal into
frequency domain. The delayless subband adaptive filter architecture suggested by Morgan and
Thi [3] avoids this problem by adapting the weights in subbands, and then transforming them to
an equivalent full-band filter which is used to cancel the acoustic echo without delay (see also
[4]). In [3] several delayless architectures are presented; in this paper we use the one that
achieves the best results, which is referred to as the closed-loop configuration.

Independently of the particular subband architecture being used, the LMS-type filters included in
every subband are subject to a compromise regarding their speed of convergence, adaptability
and residual misadjustment. Variable step-size adaptive filters can be used to alleviate this
problem, but their effective use requires some knowledge about the statistics of the filtering
scenario which is rarely available. In [5], it was shown that the adaptive convex combination of
two transversal filters with different step-sizes can easily be applied in this situation to obtain adaptive filters with improved performance. Under this approach, the individual filters are adapted using their own error signals, while the combination filter is adapted by means of a stochastic gradient algorithm in order to minimize the global error. The filter combination supplies fast convergence and low steady-state misadjustment simultaneously.

In this paper, we propose to extend the application of combination filters to subband AEC using the delayless structure, considering a different combination in every subband. In this way, not only a better convergence speed vs steady-state misadjustment tradeoff can be obtained, but it is also possible to more efficiently track changes that occur only in certain subbands.

The rest of the paper is organized as follows. First, we explain in detail the general scheme for AEC using combination of filters. Then, the results of several experiments carried out in realistic scenarios are provided to illustrate the effectiveness of our approach. Finally, the conclusions of our work and further research lines are discussed.

GENERAL SCHEME

Fig. 1 shows the general scheme of the proposed acoustic echo canceller. The far-end signal is denoted by x(n), e0(n) is the near-end disturbance, considered to be an independent and identically distributed (i.i.d) noise, h(n) is the room impulse response and

\[ d(n) = x(n)^*h(n) + e_0(n) \quad \text{(Eq.1)} \]

is the microphone signal. The objective of AEC is to produce an estimation of \( \hat{d}_c(n) \), which is then subtracted from \( d(n) \) to get the error signal, \( e(n) \). If the cancellation is perfect, we have \( e(n) = e_0(n) \).

![Figure 1. Block diagram for the AEC architecture using combination filters in every subband.](image)

When using subband AEC, the input and error signals have to be transformed to frequency domain and weight filters are independently adapted for every subband. Additionally, when considering a different combination of one fast and one slow adaptive filters for every subband, the error signals corresponding to both filter components have also to be frequency-transformed and split into the different subbands (see Fig. 1).
In the next two subsections we will first focus on the algorithm for combining adaptive filters, explaining how the adaptation of weights is carried out in every subband. Then, we will explain the implications that these combination filters have on the overall delayless structure.

Filter combination
In every subband we use an adaptive convex combination of two NLMS filters (CNLMS algorithm). The use of NLMS is advantageous with respect to standard LMS because speech signals usually alternate between activity and silence periods.

Following [5], both NLMS components have to be independently updated using their own error signal, and using a different step-size. Without loss of generality, we assume that \( \mu_1 > \mu_2 \), so that the first filter adapts faster. Then, the adaptation rule for the weights of the \( i \)th subband and the \( f \)th filter (\( f = 1, 2 \)) component is given by:

\[
\mathbf{w}_{ij}(n+1) = \mathbf{w}_{ij}(n) + \mu_i \frac{\mathbf{x}_i(n)}{\|\mathbf{x}_i(n)\|_2^2} \mathbf{e}_{ij}^*(n)
\]  
(2)

where \( \mathbf{x}_i(n) \) is the input signal corresponding to the \( i \)th subband, \( \mathbf{e}_{ij}(n) \) stands for the error incurred by the \( j \)th filter (\( j = 1, 2 \)) in that subband, and the star superscript denotes complex conjugation.

In order to obtain the overall weights for every subband, the weights of the component filters are linearly combined using an adaptive convex combination determined by

\[
\mathbf{w}_i(n) = \eta_i(n)\mathbf{w}_{i,1}(n) + [1 - \eta_i(n)]\mathbf{w}_{i,2}(n)
\]  
(3)

where the mixing parameter, \( \eta_i(n) \), is adapted using a gradient descent method to minimize the overall quadratic error of the combination. According to [5], it is convenient to introduce a non-linear activation for the mixing parameter, defining

\[
\eta_i(n) = \text{sgm} \left[ a_i(n) \right] = \left[ 1 + \exp \left( - a_i(n) \right) \right]^{-1}
\]  
(4)

and update \( a_i(n) \) instead of \( \eta_i(n) \) (note that the relation between these parameters is one to one). The update equation for \( a_i(n) \) when the subband errors \( \mathbf{e}_{i,1}(n) \) and \( \mathbf{e}_{i,2}(n) \) are complex, becomes:

\[
a_i(n+1) = a_i(n) - \frac{\mu_a}{2} \frac{\mathbf{e}_{i,1}(n) \mathbf{\bar{e}}_{i,1}(n)}{\|\mathbf{e}_{i,1}(n)\|^2} - a_i(n) - \frac{\mu_a}{2} \frac{\mathbf{e}_{i,2}(n) \mathbf{\bar{e}}_{i,2}(n)}{\|\mathbf{e}_{i,2}(n)\|^2} \frac{\eta(n)}{\|\mathbf{e}_{i,2}(n)\|^2} \\
= a_i(n) + \mu_a \Re \left[ \mathbf{e}_{i,1}(n)^* (\mathbf{e}_{i,2}(n) - \mathbf{e}_{i,1}(n)) \frac{\eta_i(n)}{1 - \eta_i(n)} \right]
\]  
(5)

Where \( \mathbf{e}_{i}(n) = \eta_i(n) \mathbf{e}_{i,1}(n) + [1 - \eta_i(n)] \mathbf{e}_{i,2}(n) \) is the error achieved by the overall combination in the subband.

Note that the update of \( a_i(n) \) could stop whenever \( \eta_i(n) \) is very close to 0 or 1. In order to overcome this problem, we constrain \( a_i(n) \) to the range \([-4,4]\).

As it was shown in [5], this adaptive filtering scheme is able to put together the convergence speed of its first component and the low steady-state misadjustment of the filter with \( \mu_2 \).

Delayless subband architecture
Since we are implementing the close loop configuration proposed in [3], three wideband filters (for the fast and slow components and for the overall filter), each one with \( N \) coefficients, are needed. As we can see in Fig. 1, the output of wideband filters \( \mathbf{w}_1 \) and \( \mathbf{w}_2 \), which are derived from the subband weights associated to the fast and slow filters, respectively, are used to calculate error signals \( \mathbf{e}_{1}(n) \) and \( \mathbf{e}_{2}(n) \), which are then fed back for CNLMS adaptation. On the other hand, the overall filter response, \( \mathbf{w} \), is used to model the room impulse response, and to estimate the value of the target signal, \( \mathbf{d}(n) \).
The polyphase FFT technique [6] can be used to generate the subband signals \((x_i(n), e_{1i}(n), e_{2i}(n))\). Thus we can make \(M\) contiguous single-sideband bandpass filters whose output are downsampled by a factor \(D = M/2\), producing \(M\) subband signals steamed from every wideband signal \((x(n), e_{1}(n), e_{2}(n))\). For real signals, as with speech, wideband coefficients are real too and only half of the subbands need to be processed.

The filter weights in every frequency subband are transformed back by properly stacking, complex conjugation and inverse transformation (FFT\(^{-1}\)) to produce the wideband filter [3]. This process has to be repeated three times, both for the two component filters, and for the overall filter coming out from the combination in each subband. In this paper we use the stacking scheme described in [3], but more efficient stacking configurations can be used [7].

**EXPERIMENTAL RESULTS**

In this section we illustrate the performance of the proposed AEC structure in several realistic scenarios. The sampling frequency is 8 kHz and the signal-to-noise-ratio (SNR) is 30 dB. Two input signals with different spectral properties are used: white noise and noise with a speech-like spectrum (USASI noise, [8]). The impulse response, \(h(n)\), is obtained through realistic acoustic simulation of an office with the software ODEON Room Acoustics Program, courtesy of Brüel & Kjaer (www.bksves.com). The room response \(h(n)\) is truncated to 1024 samples for the experiments.

The parameters used in the simulations were: \(N = 1024\) wideband taps, \(M = 32\) subbands, \(D = 16\) (down-sampling factor), \(N/D = 64\) taps per subband, \(\mu_1 = 0.5\), \(\mu_2 = 0.05\) and \(\mu_a = 500\). All results have been averaged over 1000 runs. We have used excess mean-square error (EMSE) as the figure of merit, defined as the excess over the minimum mean-square error achievable by any filter, namely \(\text{EMSE}(n) = E[(e(n) - e_0(n))^2]\).

Two different sets of experiments are described in this section: first, we show the initial convergence of the algorithm in a situation in which no changes are produced in the room response, exciting the echo channel with white noise as the input signal. Fig. 2 represents CNLMS behavior in this situation, together with the EMSE achieved by the component filters. The lower panel shows the evolution of the mixing parameter associated to the third subband (375 Hz - 625 Hz). As we can see, the combined filter keeps the fast convergence of the NLMS with \(\mu_1 = 0.5\) and the lower residual error of the slow filter. This behavior is also seen from the evolution of the mixing parameter which changes from 1 to 0 when the EMSE of the slow filter descends below the EMSE of the fast one. This behavior is observed in every subband and, therefore, inherited by the overall echo canceller (Fig. 3).

For the second experiment, we have introduced an artificial change in the room response in the frequency range from 375 to 525 Hz. The input signal in this case is USASI noise. Fig. 4 represents the behavior of the third subband of the echo canceller for this case. We can see that CNLMS follows the fast component not only following the initial transition, but also at \(t = 20\) s, when the change in the channel occurs. In steady state, the residual misadjustment of CNLMS is as low as that of the slow NLMS component. Adjacent subbands also show small perturbations at \(t = 20\) s, because the single-sideband bandpass filters are not sharp enough. Fig. 5 depicts the behavior of the filters in the 15th subband (3625-3875 Hz). There is no change in this frequency band and the mixing parameter maintains its value after the initial convergence has taken place. If we compare Figs. 4 y 5 we realize that a slower transition from the fast to the slower component occurs for the higher subbands. This is a consequence of the slower adaptation of (Eq. 5) resulting from the lower power of the input vectors in those subbands.

Finally, Fig. 6 shows the overall echo canceller behavior. We can see that the use of a CNLMS filter in every subband, not only provides simultaneously fast convergence and low steady-state misadjustment in every subband, but also allows enhancing the overall performance when changes in the room response affect every subband in a different way.
Figure 2. Cancellation EMSE in the third subband. a) EMSE of CNLMS and component filters. 
b) Evolution of the mixing parameter.

Figure 3. Overall echo canceller behavior. EMSE of filters $w_1$, $w_2$ and $w$.

Figure 4. Third subband behavior with rapid change in the room response. 
a) EMSE of CNLMS and component filter. b) Evolution of the mixing parameter.
Figure 5. Fifteenth subband behavior with rapid change in the room response. EMSE of CNLMS and component filters.

Figure 6. Overall echo canceller behavior with rapid change in the room response. EMSE of filters $w_1$, $w_2$ and $w$.

CONCLUSIONS AND FUTURE WORK

In this paper we present a new algorithm for adaptive AEC using a delayless architecture and implementing the adaptive filter of each subband using a combination of a fast and a slow component. The delayless scheme avoids the delay introduced in the signal path in classic subband schemes, while the use of a convex combination of two NLMS with different step-sizes simultaneously provides fast convergence and low residual error. We have illustrated in realistic scenarios that using CNLMS in every subband also improves the overall performance when changes appear only in a reduced number of subbands.

Further investigation is currently being carried out to minimize the effects of the different delays that appear in each subband when dealing with colored signals.

Acknowledgements

This work has been partly supported by the Spanish Ministry of Education and Science under grant CICYT TEC-2005-00992 and by Madrid Community grant S-505/TIC/0223.

References: