STUDIES OF A WIDEBAND STEREOPHONIC ACOUSTIC ECHO CANCELER

Peter Eneroth  
Tomas Gänsler

Dept. of Applied Electronics,  
Lund University,  
P.O. Box 118 S-221 00 Lund, Sweden  
{peter.eneroth,tomas.gansler}@te.lth.se

Steven Gay  
Jacob Benesty

Bell Laboratories, Lucent Technologies  
600 Mountain Avenue  
Murray Hill, NJ 07974-0636, USA  
{slg,jbenesty}@bell-labs.com

ABSTRACT

In this paper a wideband stereophonic acoustic echo canceler is presented. The fundamental difficulty of stereophonic acoustic echo cancellation (SAEC) is described and an echo canceler based on a fast recursive least squares algorithm in a subband structure is proposed. This structure has been used in a real-time implementation, on which experiments have been performed. In the paper, simulation results of this implementation on real life recordings, with 8 kHz bandwidth, are studied. The results clearly verify that the theoretic fundamental problem of SAEC also applies in real-life situations. They also show that more sophisticated adaptive algorithms are needed in the lower frequency regions than in the higher regions.

1. INTRODUCTION

Even though echo cancellation belongs to the traditional problems of signal processing since its introduction in 1967, [1], it is still an active field of research. New attention has been brought to the subject during the 1990s mainly for two reasons: First, the echo problem became more severe when digital wireless communication systems were introduced. The cause of the problem was longer signal path delays. These larger delays were mainly produced by more sophisticated speech coding algorithms than had previously been used in the ordinary PSTN (Public Switched Telephone Network). Secondly, the increasing use of teleconferencing systems and desktop conferencing where the acoustic echo canceler (AEC) plays a central role, has led to the requirement of faster and better performing algorithms. In these applications there is a desire to have far better sound quality and sound localization than what has been provided previously. These quality improvements can be achieved by increasing the signal bandwidth and also by adding more audio channels to the system. This last fact spurred the need for multi-channel acoustic echo cancelers of which the two channel (stereo) AEC is the most interesting since only complexity issues differ in the more general multi-channel case. Figure 1 illustrates the concept of stereophonic echo cancellation between a transmission room and a receiving room. The transmission room is sometimes referred to as the far-end and the receiving room as the near-end.

Stereophonic acoustic echo cancellation is fundamentally different from traditional mono echo cancellation. Four mono echo cancelers straightforwardly implemented in the stereo case not only would have to track changing echo paths in the receiving room but also in the transmission room! For example, the canceler has to reconverge if one talker stops talking and another starts talking at a different location in the transmission room. There is no adaptive algorithm that can track such a change sufficiently fast and this scheme therefore results in poor echo suppression. Thus, a generalization of the mono AEC in the stereo case does not result in satisfactory performance.

The theory explaining the problem of SAEC was first described in an early paper, [2] and later on in [3]. The fundamental problem is that the two channels may carry linearly related signals which in turn may make the normal equations to be solved by the adaptive algorithm singular. This implies that there is no unique solution to the equations but an infinite number of solutions and it can be shown that all (but the physically true) solutions depend on the transmission room. As a result, intensive studies of how to handle this properly have been made.

A complete theory of non-uniqueness and characterization of the SAEC solution was presented in [4] (a short version can be found in [5]). It is shown that the only solution to the non-uniqueness problem is to reduce the correlation between the stereo signals and an efficient low complexity method for this purpose was also given, [4].

Lately, attention has been focused on the investigation of other methods that decrease the cross-correlation between the channels in order to get well behaved estimates of the echo paths, [6, 7, 8, 9, 10]. The main problem is how to reduce the correlation sufficiently without affecting the stereo perception and the sound quality.

The performance of the SAEC is also more severely affected by the choice of adaptive algorithm than the monophonic counterpart. This is easily recognized since the performance of most adaptive algorithms depends on the condition number of the input signal. In the stereo case, the condition number is very high, hence standard Least Mean Square (LMS) or Normalized LMS (NLMS), converge very slowly to the true solution. More sophisticated algorithms such as the APA (Affine Projection Algorithm) or the RLS (Recursive Least Squares), that are less affected by a high condition number, handle the stereo case much better.

2. DESCRIPTION OF THE SAEC

As has been described in the previous section, a stereophonic acoustic echo canceler needs four adaptive filters, each adapted with a sophisticated algorithm. This will of course lead to a system with high computational complexity, but by applying the echo canceler in a subband scheme, it is possible to reduce this complexity. The subband structure, decomposes the signals into $M$ downsampled subband signals, where each subband corresponds to a given fre-

are needed to decompose the transmission room signals, \( x(n) \) and receiving room signals \( y(n) \). The adaptive filters then operate on these subband signals, estimating the corresponding subband residual echo signals \( e_{m}(k) \), i.e. the echo is removed in each subband separately. Finally two synthesis filter banks reconstruct the two fullband residual echo signals, to be transmitted back to the transmission room.

In the implementation, the computationally efficient polyphase FFT filter bank was chosen [11, 12]. This filter bank, due to the Fourier transform, generates complex valued subband signals, and therefore complex versions of the adaptive filters are needed. This does not necessarily increase the computational complexity significantly, since only the \( M/2 + 1 \) lower subbands need to be processed, and the upper subbands are formed by complex conjugate symmetry. Proper filter design in the filter banks is crucial in order to reduce the distortion in the reconstruction, and a flexible filter design method is described in [13]. Aliasing due to downsampling in the filter bank effectively decreases the performance of the adaptive filters [14]. Therefore the downsampling rate \( r \) is chosen to be non-critical, i.e. \( r < M \). Increasing \( r \) reduces the complexity of the adaptive filters, but increases the complexity of the filter bank, since longer filters are needed for sufficient attenuation in the stop-band. In addition to decreasing the computational complexity, it should be noted that a subband structure also increases robustness of the adaptive filters compared to a fullband echo canceler implementation, since the number of taps in each adaptive filter is reduced. It also increases the ability of efficient implementation on parallel execution units. The disadvantage is the signal path delay introduced by the filter bank.

Two different adaptive filter algorithms were implemented and used in the evaluation. As the high performance contender, the two channel fast recursive least squares (FRLS) algorithm [15, 16] was chosen. It has proven to perform better than other algorithms in SAEC [16], but has a relatively high calculation complexity and needs extra units to monitor and adjust the state of the algorithm in order to guarantee stability. As the second algorithm, the reference algorithm, a properly regularized NLMS algorithm was chosen. Known for its robustness and simplicity, but it has slower convergence on signals with high condition number.

As described in the introduction, the fundamental problem in SAEC lies in the correlation between the two channels. In [4], a half-wave rectifier, described by

\[
\begin{align*}
x_1'(n) &= x_1(n) + \alpha \frac{|x_1(n)| + |x_2(n)|}{2}, \\
x_2'(n) &= x_2(n) + \alpha \frac{|x_2(n)| - |x_2(n)|}{2},
\end{align*}
\]

was presented as a simple and effective non-linear method to decrease the correlation between the channels. Informal listening tests have verified that \( \alpha \leq 0.5 \) insignificantly affects the audio quality.

A real-time implementation of the system was made on a parallel floating point DSP board. Devices for residual echo suppression were implemented and used during the real-time experiments, but were deactivated in the simulations, presented in the following section.

3. A REAL LIFE EXPERIMENT

The purpose of this experiment is to show the phenomenon of convergence of the echo canceler when one talker stops and another starts talking at a different location in the transmission room. This is caused by the inherent non-uniqueness problem of SAEC. Perceptually transparent nonlinear processing, [4], has theoretically been proven to solve this problem. We show how one such nonlinear processor, the half-wave rectifier (1, 2), decreases the correlation between the two channels, and improves the performance of the SAEC. A comparison between the FRLS and the NLMS algorithms in lower and higher subbands is also performed.

The recordings were made in actual rooms at the Bell labs facility (HuMaNet I, [17]). The receiving room SNR is fairly high, \( \approx 40 \text{ d}B \), and the sampling rate is 16 kHz. The echo canceler uses a 64 band filter bank, and the downsampling rate is 48. The residual echo signals in the lower 33 subbands are estimated by the adaptive filters, whereas the upper subbands are reconstructed as described in Section 2. In each subband, the length of the adaptive filter is 66 taps. This corresponds to 3168 fullband taps by which it is possible to cancel an echo path of length 198 ms.

As shown in Section 1, the echo canceler problem has an infinite number of solutions when the two input signals, \( x_1 \) and \( x_2 \) in
Fig. 1. are highly correlated. The coherence function,

\[ \gamma(f) = \frac{S_{x_1x_2}(f)}{\sqrt{S_{x_1x_1}(f)S_{x_2x_2}(f)}} \]  

is a measure of how correlated the two channels are [4], where \( \gamma(f) = 1 \) shows that the two channels are completely linearly related to each other. In Fig. 3, this relation is shown for the recorded speech signals described above. The power spectral estimates, \( S_x(f,x) \), were calculated using the Welch method [18], with a 8196 taps long Hanning window. Figure 3a shows the coherence between the left and right channels for unprocessed transmission room speech, and Fig. 3b, the coherence for the same signal, but processed with the half-wave rectifiers, \( \alpha = 0.5 \).

One way to show the effectiveness of the decorrelation is to study how the performance of the echo canceler decreases after a speaker position change in the transmission room. If the physically true solution is found by the adaptive filter, the performance is independent of impulse response changes in the transmission room, \( g_1 \) and \( g_2 \) in Fig. 1. As performance index, the mean square error (MSE) energy of the residual is used. The MSE is given by,

\[ \text{MSE} = \frac{P_x}{P_y}, \quad P_y = \text{LPF}\{e^2(n)\}, \]  

where LPF denotes a low-pass filter. \( P_y \) is calculated analogously. Figures 4b, c show the MSE for unprocessed transmission room speech and speech processed with the half-wave rectifier (\( \alpha = 0 \), and 0.5 respectively, where \( \alpha \) is the degree of nonlinearity).

As shown in Fig. 3, the correlation between the channels is higher in the lower frequencies than in the higher frequencies. Therefore, the stability problems, due to changes of the transmission room signal path \( (g_1,g_2) \), should be severer in the lower sub-bands. In Fig. 5a, the MSE before the transmission room speaker position change for an unprocessed signal (solid line) versus the MSE directly after the position change (dashed line) is plotted as a function of the subband number. As can be seen, the echo canceler performance is decreased more in the lower subbands. In other speech signals, where the transmission room background noise is higher (SNR < 30 dB), the stability problem is only apparent in the lower half of the subbands. In Fig. 5b, it is shown that the half-wave rectifier effectively eliminates the stability problems in all subbands.

In the previous examples, the two-channel FRLS algorithm is used in all subbands. In order to reduce the calculation complexity, it is possible to switch to an NLMS algorithm in the upper subbands without significantly reducing the performance of the echo canceler. In Fig. 6 the MSE performance difference between the FRLS and the NLMS algorithm is shown for one typical lower and one typical higher subband. The impressive performance gain of the FRLS algorithm only applies for the lower subbands.

4. CONCLUSIONS

With the real-time stereophonic acoustic echo canceler presented in this paper, we have been able to confirm that the use of two channels significantly enhances the ability to separate speech signals in video conferencing systems. Therefore a listener has an improved ability to distinguish one speaker in the other room, when another speaker, situated at another position, is talking at the same time.

It has also been confirmed that decorrelation is crucial for the stability of the system, both in real-time experiments and in offline simulations on the real-life recorded signals, presented in the previous section. The studies confirm also that without a decorrelator, it is unlikely that the echo canceler converges to the correct solution. This is especially the case in the lower subbands, and in situations when the far-end background noise is low. Finally it is shown that the two-channel FRLS adaptive algorithm is superior
Figure 5: Subband mean square error at two time instances, solid line at 5.1 s and dotted line at 5.4 s, where the latter is directly after a speaker position change in the transmission room. (a) Unprocessed signal, shown in Fig. 4b. (b) Signal processed with the half-wave rectifier, $\alpha = 0.5$, shown in Fig. 4c.

to the NLMS in the lower subbands, but the performance gain is less obvious in the upper subbands.

5. REFERENCES


