PERFORMANCE IMPROVEMENT OF ACOUSTIC ECHO CANCELLERS IN THE PRESENCE OF EXTERNAL VOLUME CHANGE

Wilf Leblanc *Elias Nemer*

Broadcom Corporation
5300 California Avenue, Irvine CA 92617, USA
E-mail :{wilf, enemer}@ieee.org

ABSTRACT

This paper presents a new method for improving the performance of an acoustic echo canceller operating in a hands-free telephony context, where the terminal connects to an external amplifier whose volume can be changed arbitrarily. This change would normally results in the canceller loosing convergence, and temporarily causes an echo to be heard by the far side. The proposed method provides an effective way to quickly compensate, by continuously tracking key measures across the frequency bands, and adjusting the coefficients as well as the output of the canceller's filter, in order to minimize the level and duration of the residual echo.

A subband architecture consisting of two sets of cancellers is considered. Changes in external volume (or tone) are translated into a complex gain (or set of gains) and applied to the output or the coefficients of the foreground canceller filter, according to the optimality criterion of minimizing the mean square error (MSE) at the canceller output. The algorithm uses various metrics to ensure the validity of the computed gains and the overall canceller's stability. Listening tests and perceptual evaluation of speech quality (PESQ) score done in a real-life setting show the method results in a noted improvement of the overall perceptual performance of the canceller.

Index Terms— *AEC, External Volume Change*

1. INTRODUCTION

An acoustic echo canceller (AEC) is used in any duplex telephony system to eliminate the return echo to the far-end user that is caused by reflections or coupling between the speakers and the microphone in the near-end room [1]. The echo canceller typically consists of a linear (or nonlinear [2]) filter and an adaptation algorithm that adjusts the filter coefficients in a way to match the estimated echo path. The adaptation is based on an optimality criterion such that the outgoing signal to the far end contains a minimum level of residual echo. Changes in the echo path in the near-end room, such as positional shifts by the persons or the devices (the phone, the microphone, or the speaker) cause the adaptation to start a reconvergence process to readapt to the new path. Until the echo canceller has adapted to the new response, significant echo may be sent to the far-end user, depending on the level of change and the speed of convergence.

In the context of hands-free telephony, end-user devices, such as Voice over IP (VoIP) desk phones or car kits, can be connected to external speakers or amplifiers whose volume or tone settings (bass, treble) can be changed arbitrarily by the user (Figure 1). From the perspective of the AEC, this change yields an increase in the error output and causes the adaptation process to start reconverging to match the new echo path. In the meantime, the far-end user will hear an unexpected echo lasting several hundred milliseconds, which can prove annoying.

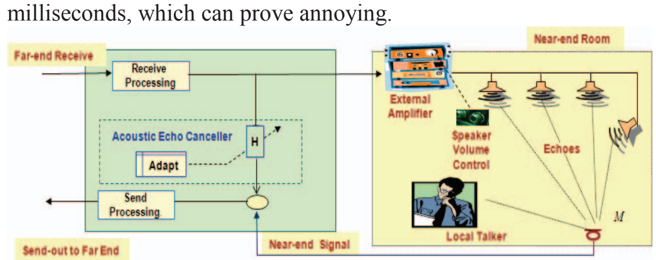

Figure 1. AEC in a Context of External Volume Change

Most practical echo cancelling systems include other components than the linear cancellers, such as an echo suppressor [3] or Nonlinear Processor (NLP) [4] that uses various methods to further attenuate residual echoes. Experiments however show that a change in external volume often goes largely untracked by such post processing components, and is likely due to the averaging of the metrics used in post processing that cannot detect quick changes such as external volume.

In this paper, we propose a new approach to improving the overall performance of an AEC with a changing external volume (or tone setting). The approach compensates for these changes by estimating a complex gain (or set of gains) that are applied to the coefficients or the output of the canceller filter and derived based on minimizing the mean square error at the canceller output.

In the proposed two-path, subband architecture, a single complex gain (or set thereof), which is derived by averaging entities across all subbands, is applied to the foreground filters while the background filters reconverge to the new 'echo path.' The method can be extended to accommodate a tone-setting change (instead of a volume change) by grouping frequency bands into M groups and (with proper averaging) computing a separate gain for each. The proposed system results in significantly reducing the level and duration of the residual echo, while ensuring overall canceller's stability.

The paper is organized as follows: Section 2 describes the derivations of the optimal gain, Section 3 describes the proposed

AEC architecture, Section 4 describes the various steps in the algorithm, Section 5 presents the simulation results, and Section 6 presents the conclusions.

2. TRACKING EXTERNAL VOLUME CHANGES

Given the system shown in Figure 1, with an external volume change, the following details the expression of the gain that could be applied at the output of the canceller's filter in order to minimize the residual echo.

2.1 Gain Computation

The optimal gain is found by minimizing the mean square error (MSE) at the output of the linear canceller (Figure 2):

$$MSE_{Gopt} \equiv E\left[\left|y - G_{opt}\hat{y}\right|^2\right] = E\left[(y - G_{opt}\hat{y})(y - G_{opt}\hat{y})^*\right] \quad (1)$$

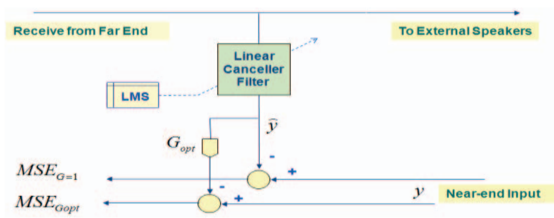

Figure 2. Optimal gain application

The gain –assumed real for now- is found by setting the MSE gradient to zero. The MSE is first rewritten as:

$$\begin{aligned} MSE_{Gopt} &= E\left[(y - G_{opt}\hat{y})(y^* - G_{opt}^*\hat{y}^*)\right] \\ &= E\left[yy^* - 4G_{opt}\operatorname{Re}\left\{y\hat{y}^*\right\} + 2G_{opt}^2\hat{y}\hat{y}^*\right] \end{aligned} \quad (2)$$

The gradient is therefore:

$$\frac{\partial MSE_{Gopt}}{\partial G} = -4E\left[\operatorname{Re}\left\{y\hat{y}^*\right\}\right] + 4GE\left[\hat{y}\hat{y}^*\right] \quad (3)$$

Setting the gradient to zero yields the gain:

$$G_{opt} = \frac{E\left(\operatorname{Re}\left\{y\hat{y}^*\right\}\right)}{E\left(\hat{y}\hat{y}^*\right)} \quad (4)$$

The above may be extended for the case of a complex gain, by setting the two partial derivatives to zero:

$$\frac{\partial MSE_{Gopt}}{\partial G_r} = 0 \qquad \frac{\partial MSE_{Gopt}}{\partial G_i} = 0 \quad (5)$$

With the MSE of a complex gain given by:

$$\begin{aligned} MSE_{Gopt} &= E\left[(y - G_{opt}\hat{y})(y^* - G_{opt}^*\hat{y}^*)\right] \\ &= E\left[yy^* - G_{opt}\hat{y}y^* - G_{opt}^*y\hat{y}^* + \left|G_{opt}\right|^2\hat{y}\hat{y}^*\right] \\ &= E\left[yy^* - (G_r + jG_i)\hat{y}y^* - (G_r - jG_i)y\hat{y}^* + \left(G_r^2 + G_i^2\right)\hat{y}\hat{y}^*\right] \end{aligned} \quad (6)$$

Setting both derivatives to zero yields the real and imaginary parts of the gain, which, after combining, is expressed as:

$$G_{opt} = G_r + jG_i = \frac{E\left(y\hat{y}^*\right)}{E\left(\hat{y}\hat{y}^*\right)} \quad (7)$$

Where $E\left(y\hat{y}^*\right)$ is the cross-correlation between the echo estimate ($\hat{y}$) and the near-end signal (y); and $E\left(\hat{y}\hat{y}^*\right)$ is the energy of the echo estimate.

2.2 Iterative updates

In a practical system, the processing is divided into frames of a given length (10 to 20 ms). A new gain may be computed every frame, and a smoothing scheme can be used to avoid abrupt changes. Alternatively, an iterative expression for updating the gain may be used to reduce divisions and provide more degrees of freedom in controlling the adaptation. A general gradient-based [5] expression for the gain can be defined as:

$$G_{opt}(n) = G_{opt}(n-1) + \mu \cdot \Delta(n) \quad (8)$$

Where the gradient is the derivative of the cost function (MSE) with respect to the gain. From the above derivation,

$$\begin{aligned} \Delta(n) &= \frac{\partial MSE_{G_{opt}}}{\partial G_r} + j\frac{\partial MSE_{G_{opt}}}{\partial G_i} = -E\left[\operatorname{Re}\left\{y\hat{y}^*\right\}\right] - E\left[\operatorname{Im}\left\{y\hat{y}^*\right\}\right] + (G_r + jG_i)E\left[\hat{y}\hat{y}^*\right] \\ \Delta(n) &= G_{opt}(n-1) \cdot E\left(\hat{y}\hat{y}^*\right) - E\left(y\hat{y}^*\right) \end{aligned} \quad (9)$$

Where statistical expectations may be estimated by time averaging over the past few frames. The expression may be further simplified by replacing the statistical averages with the instantaneous entities, thus:

$$\Delta(n) = \hat{y}^*(n)\left[G_{opt}(n-1) \cdot \hat{y}(n-1) - y(n-1)\right] \quad (10)$$

Which is similar to the approximation done in the general LMS adaptation. The step-size of the adaptation parameter may be a fixed, power of two ($\mu = 2^{-\alpha}$) step-size that results in a cost-effective implementation. It may also be variable and a function of the norm of the error (the full-band MSE) to speed up the adaptation when the error is large.

3. PROPOSED ARCHITECTURE

The gain derived earlier is applied in the context of a subband architecture, which is described in the following.

3.1 Two-path echo canceller

The two-path echo cancellation technique is a common approach for handling common acoustic echo cancellation problems such as double-talk or sudden echo path changes [6]. The method uses two filters: a background filter adapts its coefficients to the (changing) echo at all times, while a second foreground filter receives its coefficients from the background filter, when the latter is performing better than the former. Critical to good performance is the coefficient copy techniques of background-to-foreground [7], which often include various measures to assess the convergence quality of the two filters.

In the present context, a change in the external volume will trigger a reconvergence of the background canceller and an eventual copying of the coefficients to the foreground. While the background filter is reconverging, the foreground filter is ill-converged and results in significant residual echo until the new improved coefficients are copied over. Therefore, upon detecting a change in external volume, a gain is computed and applied to the output of the foreground filter and/or its coefficients (Figure 3), thus reducing the level and duration of the residual echo. Upon reconvergence, the coefficients of the background canceller may override those of the foreground, should they yield a better MSE than the scaled foreground filter.

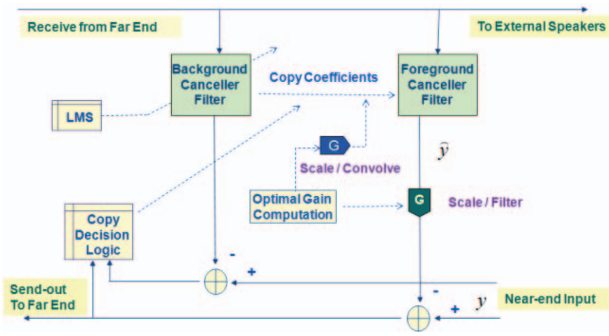

Figure 3. Foreground-Background AEC Structure

3.2 Subband-based Processing

In subband architecture, the spectrum is divided into J bands using analysis and synthesis filter banks. A two-path AEC is used in each band, as described earlier and shown in Figure 4.

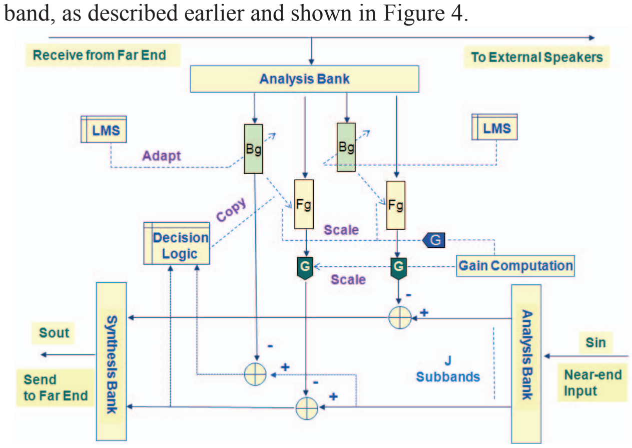

Figure 4. Subband-based 2-path Architecture

The derivation of the gain is obtained by minimizing the total MSE across all subbands. Considering -for instance the case of two subbands and a real-valued gain, the MSE is given by the following expression:

$$MSE_{Fullband\_Gopt} \equiv E\left(\left|y_1 - G_{opt}\hat{y}_1\right|^2\right) + E\left(\left|y_2 - G_{opt}\hat{y}_2\right|^2\right) \qquad (11)$$

$$MSE = E[y_1y_1^* - 2G_{opt}y_1\hat{y}_1^* + G_{Gopt}^2\hat{y}_1^*\hat{y}_1] + E[y_2y_2^* - 2G_{opt}y_2\hat{y}_2^* + G_{Gopt}^2\hat{y}_2^*\hat{y}_2]$$

The gain that minimizes the full-band MSE is obtained by setting the derivative to zero:

$$\frac{\partial}{\partial G}MSE = 0 \quad \Rightarrow -2E\left[y_1\hat{y}_1^*\right] + 2G_{opt}E\left[\hat{y}_1^*\hat{y}\right] - 2E\left[y_2\hat{y}_2^*\right] + 2G_{opt}E\left[\hat{y}_2^*\hat{y}_2\right] = 0$$

(12)

From this, it may be generalized for the case of 'j' subbands and a complex gain to:

$$G_{opt} = \frac{E\left(y_1\hat{y}_1^*\right) + E\left(y_2\hat{y}_2^*\right) + \cdots + E\left(y_j\hat{y}_j^*\right)}{E\left(\hat{y}_1^*\hat{y}\right) + E\left(\hat{y}_2^*\hat{y}_2\right) + \cdots + E\left(\hat{y}_j^*\hat{y}_j\right)} \qquad (13)$$

The expression for the gain thus involves averaging the cross-correlations as well as the echo-estimate energies across all subbands. Experimental results show that, for the case of volume change compensation, a single complex gain is sufficient across all bands. For a tone-setting change (treble and bass), the spectrum is divided into L groups, with a separate gain computed for each group, using the averaging expression (Eq 13) across the bands in that group.

The top-level operational sequence of the algorithm is illustrated in Figure 5. The linear canceller is run and an estimate of the echo is generated. The gains are computed by averaging the cross-correlations between the echo estimates and the near-end signals across all subbands (Eq 13).

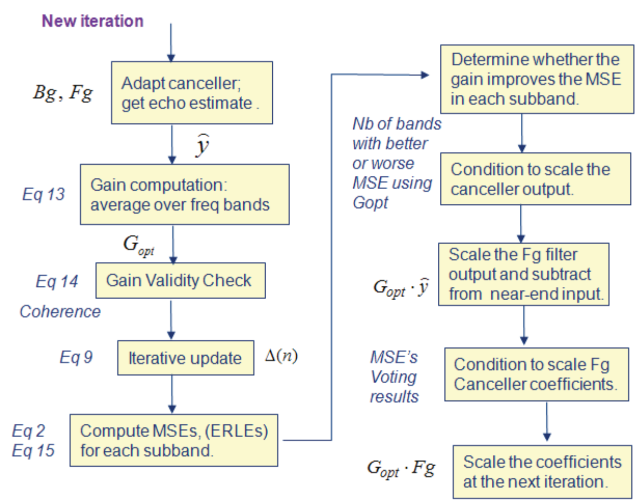

Figure 5. Algorithm Flowchart

An initial validity check is done to dismiss the estimated gain values that are likely wrong, such as during double-talk or during an echo path change, before reconvergence. The coherence measure between the echo estimate and the actual return signal is used as a necessary condition, and is defined as:

$$EchoCoherence = \frac{E\left(y\hat{y}^*\right)}{\sqrt{E\left(\hat{y}\hat{y}^*\right)E\left(yy^*\right)}} \qquad (14)$$

In the case of a properly convergent canceller, the coherence will be real and near unity, irrespective of external gain changes. This is true since any such change will affect both numerator and denominator equally in the above expression. During double talk and echo path changes, the echo estimate will not be properly correlated to the actual return, and thus the coherence will take on small and possibly complex values. Just like in the case of the gain computation, the coherence may be computed by averaging across all subbands, thus yielding a single measure per frame.

Once the gain is validated, the a priori and a posteriori MSEs are computed in each subband to determine the performance improvement (if any) from applying the gains. The a posteriori MSE is given by (Eq 2), whereas the a priori MSE is:

$$MSE_{G=1} = E\left[yy^*\right] + E\left[\hat{y}\hat{y}^*\right] - 2\,\mathrm{Re}\left\{E\left[y\hat{y}^*\right]\right\} \qquad (15)$$

Once all MSEs are computed and compared, certain metrics are tracked. Such metrics include the number of subbands for which the MSE would significantly improve by using the computed gain, as well as those for which the MSE would marginally or adversely improve. The voting results are used to determine whether to enable the scaling of the canceller filter outputs as well as the scaling of the filter coefficients.

In the subband architecture described above, when the scaling of the filter outputs is enabled, the decision to actually scale the output of the filter in a given band is based on whether the MSE (or ERLE) improves by carrying out this scaling.

Finally, a sanity-checking mechanism is used to restart all the averaging operations when it is deemed that an irrecoverable divergence has occurred. For instance, when the number of subbands that perform worse with the new gain is greater than the number that performs better, the gain may be deemed to have diverged and is reset.

4. IMPLEMENTATION RESULTS

The system shown in Figure 1 is set up in a lab environment, with the far-end receive signal simulated by playing phonetically balanced sentences, with and without background noise. An external home stereo amplifier is used with two loudspeakers placed about a meter away on either side of the near-end microphone *M*. A sound meter is used to measure the level of the signal, and the external volume is varied within a range of +/−10 dB from the nominal level. The send-out signal at the output of the canceller, and prior to NLP or any other post processing, is recorded with and without volume tracking enabled and is analyzed offline.

PESQ analysis is used to assess the improvement in both single and double talk scenarios. In the single (far-end) talker case, the score is evaluated using as reference the far-end speech, and as processed input, the sum of the far-end speech and the delayed residual echo (send out signal), with the delay being set to a typical network delay. In double talk cases, the score is evaluated using as reference the combination of the far end speech and the delayed near end, and as processed input the sum of the far-end speech and the delayed send-out signal. The results, averaged across all test cases, are summarized in table 1 below.

Single Talker Tests	PESQ Improv.
Volume change in receive	0.3
Volume change in silence	0.45
Obstacle in front of terminal in receive	0.17
Obstacle in front of terminal in silence	0.19
Double Talk Tests	**PESQ Improv.**
Volume change in receive	0.05
Volume change in silence	0.1

Table 1: PESQ improvements using volume tracking

Figure 6 illustrates a case in which volume is changed in a single-talker scenario during the quiet periods between utterances. The improvement in the residual echo energy varies and can be up to 30 dB.

In double-talk, the algorithm is fairly transparent, in that the correlation estimates are not reliable and, therefore, the gain adaptation is frozen. Consequently, the gain is unlikely to pass the criteria for improving the MSE, so it is not applied. Some score improvement was however noted during double talk, depending on where in the speech utterances the changes occurred.

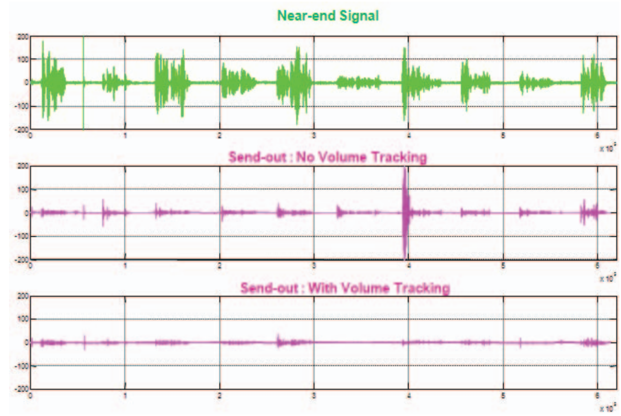

Figure 6. Residual Echo With and Without Vol Tracking

5. CONCLUSIONS

The paper described a new method for modeling and compensating for the changes in external volume for an acoustic echo canceller operating in a context where volume can be changed arbitrarily. The algorithm consists of estimating a set of complex gain factors based on a minimum MSE optimality criterion, and applying them to the output and the coefficients of the linear filter, in order to reduce the residual echo resulting from loss of convergence. A two-path subband architecture is used whereby scaling is done on the foreground filter, while the background set of cancellers adapts independently. Using recording in a real-life setting where the volume is changed within a range of −10 to +10 dB, the algorithm significantly reduces residual echo in single-talker scenarios, without any distortion to the local speech in double-talk. The proposed solution is expandable to suit external amplifiers that alter tone, or that have output equalization.

6. REFERENCES

[1] I. Benesty, T. Gaensler, D. R. Morgan, M. M. Sondhi and S. L. Gay, Advances in Network and Acoustic Echo Cancellation, Springer-Verlag, Berlin, 2001.

[2] A. Strenger *et al.* “Nonlinear Acoustic Echo Cancellation With 2nd Order Adaptive Volterra Filter”. *ICASSP* 1999.

[3] O. Hoshuyama & A. Sugiyama. “An Acoustic Echo Suppressor Based on a Frequency-Domain Model Highly Nonlinear Residual Echo. *ICASSP* 2006.

[4] ITU-T Recommendation G.168. Digital Network Echo Cancellers, April 2000.

[5] S. Haykin. Adaptive Filter Theory. Prentice-Hall 1991.

[6] T. A. K. Ochiai and T. Ogihara, “Echo Canceller with Two-Echo Path Models,” *IEEE Transactions on Communications*, vol. 25, pp. 589–595, June 1977.

[7] M. Iqbal and S. Grant. “Novel and Efficient Download Test for Two Path Echo Canceller”. *IEEE WASPAA*. Oct. 2007.