PERCEPTUAL QUALITY OF HYBRID ECHO CANCELER/SUPPRESSOR

Fredrik Wallin[1] *and Christof Faller*[2]

[1]Audiovisual Laboratory, EPFL Lausanne, Switzerland
[2]Mobile Terminals Division, Agere Systems, Allentown, PA, USA

ABSTRACT

Acoustic echoes arise whenever there is acoustic coupling between a loudspeaker and a microphone. A traditional solution for eliminating the undesired echo signal is an acoustic echo canceler (AEC), which identifies the echo path between a loudspeaker and a microphone by means of an adaptive filter. The echo signal can be canceled successfully when the modeling filter approaches the true echo path. In practice, however, a modeling filter often differs from the true echo path due to complicated reasons such as environment changes, the lack of knowledge about the length of the echo path, and so on, resulting in residual echo signals. Another way to mitigate the echo effect is through echo suppression. Unlike an AEC, an acoustic echo suppressor (AES) achieves echo attenuation by means of spectral modification. This approach usually has a much lower complexity, and is robust against minor echo path changes. However, it sometimes introduces audible distortions to the processed signal. Many practical systems combine an AEC and an AES in a sequential way such that the former achieves major echo cancellation, and the latter attenuates the residual echoes. In this paper, we propose a novel hybrid approach for addressing the echo problem. The full-band signal is split into two frequency bands, with AEC being applied to the lower band, and an AES to the upper band, respectively. Through subjective tests we demonstrate the superior quality and robustness of this new method, compared to a full-band AEC or a full-band AES.

1. INTRODUCTION

In hands-free telecommunication systems, such as teleconferencing or telephony with loudspeaker playback, the loudspeaker signal feeds back to the microphone. When this happens, the listener on the far-end side hears a delayed version of his own voice, an echo, which is very annoying. Therefore, there is a need for algorithms which are able to eliminate the undesired echo effect.

Ideally, an algorithm should be able to eliminate the echo signal while maintaining other signal components. For example, if the near-end talker is active at the same time as the far-end talker, the near-end signal needs to be let through, while the echo needs to be eliminated. The most commonly used method to accomplish this task is an acoustic echo canceler (AEC) [1], which obtains an echo estimate by adaptively identifying the echo path between a loudspeaker and a microphone, and then subtracting the estimate from the microphone signal. More recently, schemes based on spectral modification were investigated [2, 3]. As opposed to an AEC these schemes, denoted acoustic echo suppressors (AES), acquire echo estimates in the frequency domain, and then attenuate echo components through the parametric Wiener filtering technique.

Both AEC and AES have their advantages and disadvantages. The AEC is a well-defined technique. When the modeling filter approaches the true echo path, an AEC can eliminate echo signal successfully without introducing much distortion to the outgoing signal. However, in reality the modeling filter often differs from the true echo path owing to complicated reasons. For example, the modeling filter may be shorter than the true echo path, the echo path may change or the adaptive algorithm may not have enough time to update the modeling filter, and so on. As a result, some residual echoes may still remain. In addition, an AEC is often computationally expensive. In comparison, an AES is a more computationally efficient and robust algorithm. Furthermore, the AES has been shown to be more resilient to minor echo path changes [2]. But this technique sometimes introduces audible distortions to the outgoing signal.

To gain maximum echo attenuation, while maintaining affordable complexity, AEC and AES are often combined in real applications [4]. In many practical systems AEC performs major echo cancellation, followed by an AES to attenuate the residual echo signals.

Through experimental investigation and subjective listening tests, we have found that improvements can be achieved from subband processing. By splitting the full-band signal into two subbands, we observed that applying an AEC in the lower frequency band and an AES in the upper frequency band is a better way of combining these two techniques. Such an observation motivates us to design a hybrid system, which is shown to have several advantages over a full-band AEC or a full-band AES. The proposed hybrid system is more robust with respect to echo path changes compared to a full-band AEC and the results of the subjective tests indicate that the proposed scheme also provides significantly improved perceived quality compared to a full-band AES.

2. ACOUSTIC ECHO CANCELER (AEC)

The traditional solution to the acoustic echo problem is the acoustic echo canceler (AEC), which was invented in the 1960s at Bell Labs by Kelly, Logan, and Sondhi [1]. The acoustic coupling between a loudspeaker and a microphone is modeled with a linear filter, $\mathbf{h} = [h_1, h_2, \cdots, h_L]^T$, where L is the length of the echo path, and $(\cdot)^T$ denotes the transpose of a matrix or a vector. The microphone signal is then expressed as

$$y(n) = \mathbf{h}^T \mathbf{x}(n) + z(n)\,, \tag{1}$$

where $\mathbf{x}(n) = [x(n-L+1), x(n-L+2), \cdots, x(n)]^T$ is a vector of the loudspeaker signal, and $z(n) = v(n) + w(n)$ is the sum of the near-end speech and ambient noise, respectively. A modeling filter $\hat{\mathbf{h}} = [\hat{h}_1, \hat{h}_2, \cdots, \hat{h}_N]^T$ is used to approximate the

0-7803-8484-9/04/$20.00 ©2004 IEEE

true echo path $\mathbf{h}$, where N is the length of the modeling filter. An echo estimate is then obtained as

$$\hat{y}(n) = \hat{\mathbf{h}}^T \mathbf{x}(n)\,. \qquad (2)$$

Adaptive algorithms are used to search the optimum $\hat{\mathbf{h}}$, which is the best approximation of $\mathbf{h}$ in the least square sense. This is achieved by minimizing the mean square error (MSE), $E\{e^2(n)\}$, where $e(n) = y(n) - \hat{y}(n)$. Once the adaptive filter converges, it can be easily shown that the error signal $e(n)$ is in fact the echo-cancelled, outgoing signal.

3. ACOUSTIC ECHO SUPPRESSOR (AES)

Unlike AEC, an acoustic echo suppressor achieves echo attenuation through manipulating the magnitude spectrum of the microphone signal in the frequency domain, while leaving the phase spectrum untouched. A widely adopted spectral manipulation algorithm is the parametric Wiener filter (or sometimes called spectral subtraction [5, 6]). If $|\hat{U}_k(j\omega)|$ denotes an estimate of the magnitude spectrum of the echo signal at time instant k, the parametric Wiener filter based echo suppression algorithm can be expressed as

$$e(n) = F^{-1}[G(\omega)Y_k(j\omega)],$$

where e(n) is the echo-suppressed outgoing signal, $Y_k(j\omega)$ is the short-term spectrum of the microphone signal at time instant k, $F^{-1}[\cdot]$ denotes the inverse Fourier transform, and

$$G(\omega) = \left[\frac{|Y_k(j\omega)|^\alpha - \eta|\hat{U}_k(j\omega)|^\alpha}{|Y_k(j\omega)|^\alpha}\right]^\beta \qquad (3)$$

is a Wiener gain filter, where α, β, and η are design parameters to control the echo suppression performance. If the echo is under-estimated, $\eta > 1$ is used, and $\eta < 1$ if it is over-estimated.

As seen, the paramount issue in the above echo suppression algorithm is a good estimate of the magnitude spectrum of the echo signal, i.e., $|\hat{U}_k(j\omega)|$. There are different ways to obtain such an estimate. One is to simply estimate the echo signal in the time domain similarly to an AEC, based on which $|\hat{U}_k(j\omega)|$ is computed. Another is to obtain an estimate of $U_k(j\omega)$ in each frequency bin using an adaptive filter as proposed in [2]. Recently, a scheme was proposed for obtaining a smoothed version of $|\hat{U}_k(j\omega)|$ directly with low complexity [7].

4. HYBRID CANCELER/SUPPRESSOR

Under favorable conditions, when the modeling filter approaches the true echo path, AEC successfully eliminates the echo signal while other signal components are let through. Under less favorable conditions, e.g. when the echo path is changing, AEC often lets through residual echoes. AES has a higher degree of robustness, but its echo elimination performance is sub-optimal due to distortions caused by the spectral magnitude modification. The goal is to combine AEC and AES, resulting in a hybrid system exploring the advantages of both, AEC and AES.

On one hand, the main drawback of AEC is its low robustness when echo path changes occur. Experimental investigation indicates that residual echoes contain no or little energy at low frequencies. Thus, minor echo path changes seem to affect the frequency response of the echo path less at low frequencies than at high frequencies. Thus, an AEC applied only at low frequencies is more robust than an AEC applied full-band.

On the other hand, the main drawback of AES is that the phase of the residual signal is corrupted [3]. The auditory system can be considered to be "phase deaf" at frequencies above $1 - 2$ kHz [8]. Thus, by applying AES only at higher frequencies we expect that the degradations due to phase corruption are less perceptible than if AES were applied full-band.

Motivated by the fact that the strength of AEC lies at low frequencies and the strength of AES at high frequencies, the proposed hybrid acoustic echo canceler (HAEC) decomposes the signal into two subbands and applies AEC to the low frequency subband and AES to the high frequency subband, as illustrated in Fig. 1. The loudspeaker signal $x(k)$ and the microphone signal $y(k)$ is high-pass and lowpass filtered, at a cut-off frequency of f_c. The low-pass signals, $x_l(k)$ and $y_l(k)$, are processed by the AEC, while the highpass signals, $x_h(k)$ and $y_h(k)$, are processed by the AES. After being processed, the output signals from the respective systems are combined into one signal, which is transmitted to the far-end side.

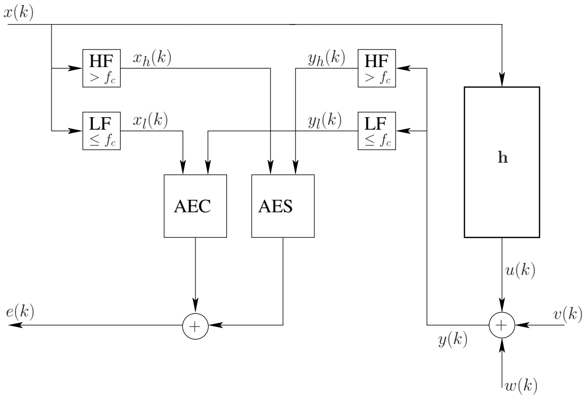

Fig. 1. A schematic illustration of HAEC. LF and HF stand for lowpass filter and highpass filter, respectively. The cut-off frequencies of the respective filters are f_c.

5. SUBJECTIVE EVALUATION

A number of simulations were performed to analyze HAEC. The perceived quality of HAEC was evaluated with subjective listening tests. As opposed to adaptively estimating the system parameters, ideal and non-ideal system parameters were used. Male and female speech signals with a length of 10 seconds and a sampling frequency of 16 kHz were used. Two male signals and two female signals were used to generate two microphone signals simulating two doubletalk situations. For all simulations, echo path responses measured in the Varechoic Chamber at Bell Labs were used [9, 10].

Doubletalk simulations were used for all evaluations. This is the most difficult situation, since HAEC needs to suppress the echo component while letting through the near-end talker signal. All simulations were carried out with SNR $= \infty$ (no noise was added to the microphone signal).

Impairment:	Grade:
Imperceptible	5.0
Perceptible, but not annoying	4.0
Slightly annoying	3.0
Annoying	2.0
Very annoying	1.0

Table 1. The five-grade impairment scale, used for the subjective listening tests.

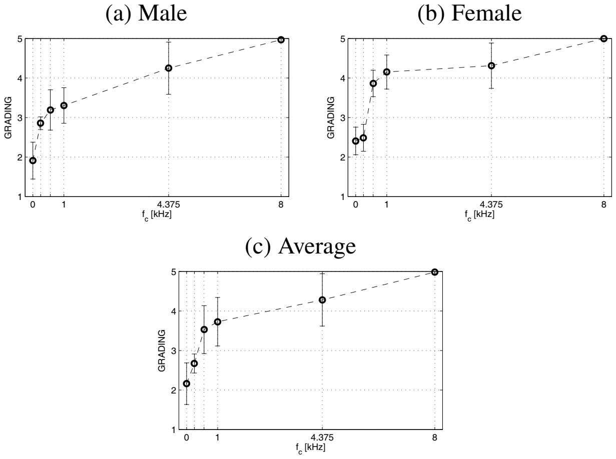

Fig. 2. The subjective test results for HAEC under ideal conditions, for (a) male speech, (b) female speech, and (c) the average test results.

The listening tests were carried out in a sound insulated room and high-quality headphones (Sennheiser HD 600) were used. The test method was the hidden reference method used according to the ITU-R recommendation BS.1116 [11]. With this test method, the listener is presented with an R-A-B triple of sound signals, where either A or B is identical to the reference signal R. The listener has to decide whether A or B is degraded and grade the degree of impairment relative to R. For this purpose the 5-grade, continuous, comparison scale shown in Table 1 was used.

The AES system parameters were optimized through informal listening. The gain filter parameters α and β were set to 1 (spectral magnitude subtraction). Additionally, the gain filter was smoothed over frequency, further reducing artifacts.

5.1. Simulations Assuming Ideal Conditions

The system was first simulated under ideal conditions by using "perfect" estimates of the system parameters. For AES this means that the exact magnitude spectrum of the echo signal was given to the system and for AEC that the true echo path response was given. Note that such a comparison is in favor of AEC, since it does not address the problem of residual echoes of AEC.

By performing the tests assuming these ideal conditions, the upper performance bound of HAEC is obtained, i.e. how good HAEC performs compared to perfect echo cancellation.

Given male and female input audio signals, HAEC was used with different cut-off frequencies, $f_c = 0, 250, 560, 1000, 4375, 8000$ Hz.

The listening test consisted of a training session of 5 items, followed by a test session of 12 items. The test was taken by 7 listeners. Four of the listeners were experienced, while three were non-experts. Two of the listeners exhibited inconsistencies in their gradings (e.g. giving high grades to sound signals whose quality apparently were worse than other signals, etc.). Their results were entirely removed (not just the inconsistent gradings).

The results of the subjective listening tests are shown in Fig. 2. Panel (a) shows the results for the male doubletalk items, (b) for the female doubletalk items, and (c) the average of both cases. In each plot, the mean grading at each cut-off frequency is shown, together with a 95% confidence interval. Note that the grading with $f_c = 8$ kHz corresponds to the performance of a full-band AEC and $f_c = 0$ kHz corresponds to the performance of a full-band AES.

For both the female and the male speech the perceived quality increases rapidly at low cut-off frequencies. Already at $f_c = 0.5$ kHz the output quality of HAEC is significantly better than that of AES ($f_c = 0$). In the female case, a cut-off frequency of only 1 kHz results in a perceived quality that is comparable to AEC. In the male case a significantly higher cut-off frequency, $f_c = 4$ kHz, is needed to achieve this level of quality. This may be explained by the fact that female speech in general has more energy at higher frequencies where the auditory system is insensitive to phase corruption. Thus, the impact of applying AES at cut-off frequencies where the auditory system is still sensitive to phase (below about $1 - 2$ kHz) has less negative impact on female speech.

The results of this subjective test indicate that good output quality can be achieved at relatively low cut-off frequencies. Note that the used test scenario is far more critical than a real conferencing scenario. As opposed to a loudspeaker in a possibly noisy office, high quality headphones were used in a sound insulated room. Furthermore, the listeners in our test were presented with critical doubletalk situations, without actively participating in the conversation. In a real scenario, during doubletalk, both conversation participants are active and their own speech will mask a certain amount of distortions.

5.2. Simulations Assuming Non-Ideal Conditions

The misalignment of the echo path response estimate, $\hat{\mathbf{h}}$, varies over time when the echo path changes. When the adaptive filter has converged, the misalignment is low. When echo path changes occur, the misalignment increases and then decreases as the adaptive filter re-converges. This situation is simulated by toggling between two different echo path responses, h_1 and h_2. The misalignment between h_1 and h_2 was -4.6 dB. The echo path responses h_1 and h_2 were measured with a setup shown in Fig. 3(a) [10]. As the estimate of the true echo path response we used h_1. By toggling the estimated echo path response between h_1 and h_2, the misalignment toggles between $-\infty$ and -4.6 dB, simulating a scenario where echo path changes occur at regular time intervals. We toggled the estimate once every second, as indicated in Fig. 3(b).

Another subjective test was carried out in the same manner as the first listening test, now assuming non-ideal conditions. Two additional items with cut-off frequencies $f_c = 440, 750$ Hz were added. The test was taken by 7 listeners, including three experienced listeners.

Figure 4 shows the results of the subjective test, together with 95% confidence intervals. Already at a cut-off frequency of $f_c =$

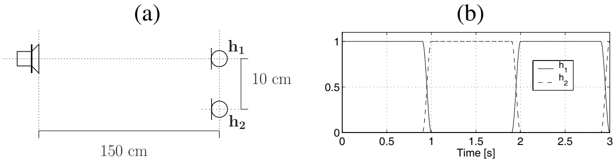

Fig. 3. (a) Echo path changes are simulated by toggling between two echo path responses, h_1 and h_2, measured as indicated. (b) Toggling scheme for the two echo path responses. As the estimate of the echo path response h_1 is used.

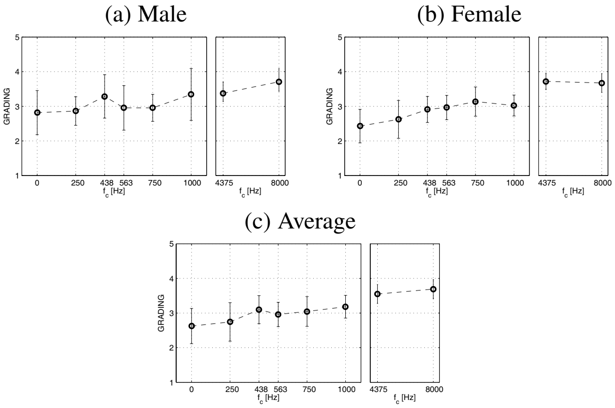

Fig. 4. The subjective test results of HAEC simulated under non-ideal conditions; (a) for male speech, (b) for female speech, and (c) the average test results.

500 Hz HAEC achieves a quality only slightly worse than a full-band AEC, for both male and female speech. Again, the grading with $f_c = 8$ kHz corresponds to the performance of a full-band AEC and $f_c = 0$ kHz to the performance of a full-band AES.

Since the subjective listening test is a "blind" test, the listeners were given no instructions on how to grade the different types of artifacts that appear in the sound signals. At low f_c the distortions arising from AES are the dominant degradation, while at high f_c the residual echoes arising from AEC are the dominant degradation. In this type of test, the listener is only listening to the sound signals, and not taking part in the conversation. The residual echo therefore seems to be preferred over the AES distortions. In a real situation the residual echo is a delayed version of the listeners own voice and would then be perceived as more annoying than in the test scenario used here.

6. CONCLUSIONS

Usually, acoustic echo cancelers (AEC) and acoustic echo suppressors (AES) are combined in series for preventing that residual echoes are let through when minor echo path changes occur. We are proposing a different way of combining AEC and AES. The microphone signal is decomposed into two subbands and AEC is applied at the low frequency subband and AES at the high frequency subband. The motivation for this is that AEC is more robust when applied to only low bandwidth audio signals and AES performs better for high frequency signals. Thus, the proposed combination improves the robustness of AEC (by only applying it at low frequencies), while improving the quality of AES (by only applying it at high frequencies).

We carried out a number of subjective tests for evaluating the quality of the proposed hybrid system, compared to full-band AEC and AES. Doubletalk situations were assessed for male and female speech in two different scenarios. In one scenario it was assumed that ideal estimates of the echo signal were given. In the other scenario, echo path changes were simulated at regular time intervals. The results indicate that the proposed hybrid system provides a good compromise between quality and robustness.

Acknowlegements
We would like to thank Jingdong Chen for his valuable suggestions regarding this work and manuscript.

7. REFERENCES

[1] M. M. Sondhi, "An adaptive echo canceler," *Bell Syst. Tech. J.*, vol. 46, pp. 497–510, March 1967.

[2] C. Avendano, "Acoustic echo suppression in the STFT domain," in *Proc. IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust.*, October 2001.

[3] C. Faller, "Perceptually motivated low complexity acoustic echo control," in *Proceedings of the 114th AES convention*, Amsterdam, The Netherlands, March 2003.

[4] J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, *Advances in Network and Acoustic Echo Cancellation*, Springer, 2001.

[5] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. ASSP-27, no. 2, pp. 113–120, 1979.

[6] W. Etter and G. S. Moschytz, "Noise reduction by noise-adaptive spectral magnitude expansion," *J. Audio Eng. Soc.*, vol. 42, no. 5, 1994.

[7] C. Faller and J. Chen, "Suppressing acoustic echo in a sampled auditory spectral envelope space," *IEEE Trans. on Speech and Audio Processing*, Submitted Aug. 2003.

[8] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, Academic Press, 4th edition, 1997.

[9] W. C. Ward, G. W. Elko, R. A. Kubli, and W. C. McDougald, "The new varechoic chamber at AT&T Bell Labs," in *Proc. Wallance, Clement, Sabine Centennial Symposium*, Woodbury, NY, USA, 1994, pp. 343–346.

[10] A. Härmä, T. Lokki, and V. Pulkki, "Drawing quality maps of the sweet spot and its surroundings in multichannel reproduction and coding," in *Proc. AES 21st Conf. on Architectural Acoustics and Sound Reinforcement*, June 2002, pp. 317–325.

[11] ITU-R, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Recommendation BS.1116-1, 1997, Available online: `http://www.itu.org`.