ROBUST ACOUSTIC ECHO CONTROL USING A SIMPLE ECHO PATH MODEL

Christof Faller and Christophe Tournery

Audiovisual Communications Laboratory, EPFL, Lausanne, Switzerland

ABSTRACT

In handsfree tele or video conferencing acoustic echoes arise due to the coupling between the loudspeaker and microphone. Usually an acoustic echo canceler (AEC) is used for eliminating the undesired echoes. The weaknesses of AEC are that it is relatively complex and that it is not robust against non-linearities occurring when using low-end components such as small loudspeakers. We are describing an algorithm which models the acoustic echo path by means of an overall delay and a coloration effect filter. The echo is removed by means of short-time spectral modification. The algorithm design aims at low computational complexity and high robustness. The effect of echo path changes and non-linearities on the performance of this algorithm are investigated.

1. INTRODUCTION

Handsfree tele or video conferencing systems need an algorithm for eliminating the undesired acoustic echoes which result from the acoustic coupling between the loudspeaker and microphone. Usually, an *acoustic echo canceler* (AEC) [1] is used to remove the undesired echo signal component from the microphone signal. An AEC achieves the echo removal by modeling the echo path impulse response with an adaptive filter and subtracting an echo estimate from the microphone signal. It is not uncommon that an adaptive filter with a length of 50 – 300 milliseconds needs to be considered, which makes an AEC highly computationally expensive.

In practice, the acoustic echo path is constantly changing due to body movements, temperature fluctuations [2], user modifications of sound volume, etc. Thus an AEC is only effective when its adaptive filter constantly adapts to the current acoustic echo path. Only a few seconds of pause in the adaptation process often already leads to residual echoes. During near-end activity (which includes time instants of doubletalk), the adaptive filter is not able to effectively adapt to the current acoustic echo path. Thus, there will always be pauses in the adaptation leading to residual echoes. To address this problem, an AEC is usually combined with a post processor aiming at removing the residual echoes.

Another problem of an AEC is that its adaptive filter only converges if the linearity assumption of the acoustic echo path is reasonably accurately fulfilled. Thus, the performance of an AEC for devices with low end components is often not acceptable. A particular problem is the use of small loudspeakers at high volume causing a high degree of non-linearity (clipping).

It has recently been shown that not only an AEC can provide echo control for duplex communication, but that also an *acoustic echo suppressor* (AES) can provide a high degree of duplex capability. The above mentioned problems, indicating the weaknesses of AEC in practical situations, may suggest that an AES may be a viable alternative to an AEC despite of its potential for artifacts during doubletalk. In [3] AEC and AES were compared in terms of speech quality during doubletalk, indicating a relatively small quality difference under non-ideal conditions.

In [4] echo removal by means of spectral modification was proposed for improving the robustness of a frequency domain acoustic echo canceler. An AES without a need for estimating the acoustic echo path impulse response was proposed in [5]. Recently, we proposed an improved AES using a simple echo path model [6], aiming at low computational complexity and high robustness. The echo path is modeled by an overall delay parameter and a coloration effect filter which captures the effect of the echo path in terms of short-time spectral modification. Sections 2 and 3 describe this AES. Simulations, with an emphasis in assessing robustness and non-linearities, are presented in Section 4.

2. ACOUSTIC ECHO SUPPRESSOR (AES)

Unlike AEC, an AES achieves echo attenuation through manipulating the magnitude spectrum of the microphone signal in the frequency domain, while leaving the phase spectrum untouched. For noise suppression, a widely adopted spectral manipulation algorithm is the parametric Wiener filter (or sometimes called spectral subtraction [7]). If $|\hat{Y}(i,k)|$ denotes an estimate of the magnitude spectrum of the echo signal with frequency index i and time index k, a parametric Wiener filter based echo suppression algorithm can be

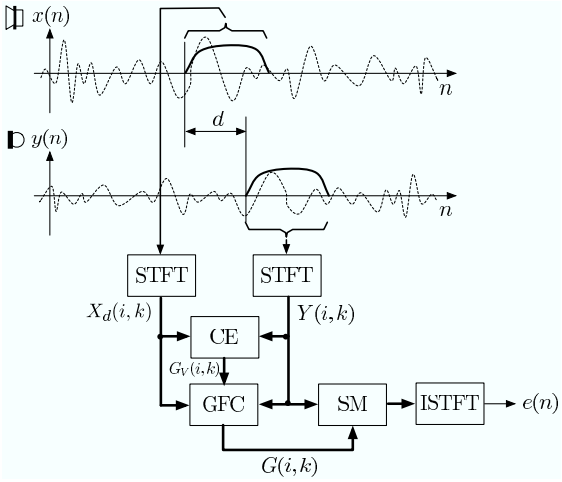

Fig. 1. Block diagram of the proposed acoustic echo suppression algorithm. STFT, ISTFT, CE, GFC, and SM stand for short time Fourier transform, its inverse, coloration effect estimation, gain filter computation, and spectral modification, respectively.

expressed as

$$e(n) = \mathrm{F}^{-1}[G(i,k)Y(i,k)]\,, \tag{1}$$

where e(n) is the echo-suppressed outgoing signal, $Y(i,k)$ is the short time spectrum of the microphone signal, $\mathrm{F}^{-1}[\cdot]$ denotes the inverse Fourier transform, and

$$G(i,k) = \left[\frac{\max(|Y(i,k)|^\alpha - \beta|\hat{Y}(i,k)|^\alpha, 0)}{|Y(i,k)|^\alpha}\right]^{\frac{1}{\alpha}} \tag{2}$$

is a gain filter, where α and β are design parameters to control the echo suppression performance [8]. If the echo is under-estimated, $\beta > 1$ is used, and $\beta < 1$ if it is over-estimated.

3. THE PROPOSED AES

The proposed scheme is illustrated in Figure 1. Short time Fourier transform (STFT) spectra are computed from the loudspeaker and microphone signals. A delay d between the STFTs is chosen such that most of the effect of the acoustic echo path is captured. Then, a real-valued "coloration effect filter" $G_V(i,k)$ [6], mimicking the short-time spectral modification effect of the echo path on the loudspeaker signal, is estimated. For obtaining an approximate echo magnitude spectrum, the estimated delay and coloration effect filter are applied to the loudspeaker signal spectra,

$$|\hat{Y}(i,k)| = G_V(i,k)|X_d(i,k)|\,, \tag{3}$$

where d indicates that the spectrum is computed with a waveform that is delayed by d samples. Underestimation of the echo signal magnitude spectrum due to ignoring the late reflections [6] can be compensated for by using a $\beta > 1$ (2). Note that this is not a precise echo spectrum or magnitude spectrum estimate. But it is precise enough for applying echo suppression, i.e. (1) with (2), with appropriate time and frequency smoothing.

The coloration effect filter is computed as the magnitude of the least squares estimator

$$G_V(i,k) = \left|\frac{\mathrm{E}\{X_d^*(i,k)Y(i,k)\}}{\mathrm{E}\{X_d^*(i,k)X_d(i,k)\}}\right|\,, \tag{4}$$

where * denotes complex conjugate. Since the acoustic echo path is likely to vary in time, $G_V(i,k)$ is estimated iteratively in time using a single pole low pass filter for estimating the expectations in (4). For details refer to [6].

To prevent that during periods of doubletalk the coloration effect filter $G_V(i,k)$ diverges, we use two coloration effect filters, similarly as two echo path models have been used for conventional AEC [9]. This type of doubletalk control has been used for AES in [5].

The main difference between the proposed scheme and other approaches is that not the acoustic echo path is estimated, but merely a global delay parameter and a filter characterizing the coloration effect of (the early part of) the acoustic echo path. This representation (delay and coloration effect filter) is largely insensitive to acoustic echo path changes and more resilient against non-linearities than the echo path itself.

4. SIMULATIONS AND EVALUATIONS

The audio signals are processed in blocks of length 10 ms. The simulations are carried out with 16 kHz sampling frequency, for which blocks of 160 samples are processed at a time. A FFT of size 512 is used with a sine window (analysis and synthesis) of length 320 with 50% window hop size. The computational complexity of the proposed scheme is much lower than a conventional AEC, since only the real-valued coloration effect filter values $G_V(i,k)$ need to be estimated as opposed to the echo path with many more parameters (filter taps).

A dialogue sequence is used, starting with far-end only speech (loudspeaker signal), followed by near-end only speech, and concluding with far-end and near-end speech simultaneously (doubletalk). The signal-to-noise ratio of the microphone signal is 20 dB. Measured impulse responses with 4096 taps are used for generating the loudspeaker and microphone signals. The impulse responses were measured using a computer-based desktop audio system.

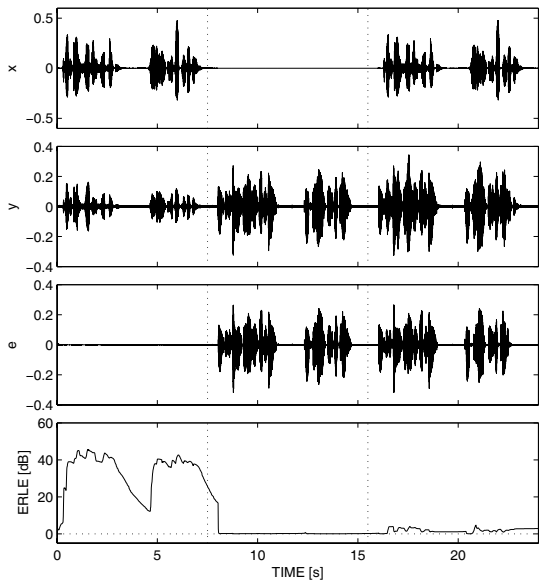

Fig. 2. The loudspeaker signal x, microphone signal y, AES output signal e, and ERLE are shown.

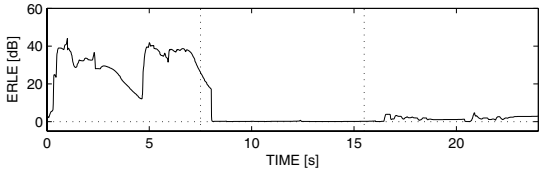

Fig. 3. The ERLE for the same simulation as shown in Figure 2, but simulating echo path changes every second. The ticks on the x-axis indicate the time instants when the echo path is changed.

4.1. Echo suppression and doubletalk performance

The top two panels of Figure 2 show the loudspeaker and microphone signals, $x(n)$ and $y(n)$, resulting from the described dialogue sequence. The vertical dotted lines separate the three parts of the simulation: Far-end only speech, near-end only speech, and doubletalk. The bottom two panels show the echo suppressed AES output signal $e(n)$ and the echo return loss enhancement (ERLE) in dB, respectively.

The ERLE and $e(n)$ imply that during far-end only speech the echo is instantly suppressed. The instant suppression is due to the initial values for $G_V(i,k)$ which are such that the echo is overestimated initially and thus suppressed. The near-end only speech gets through unimpaired. The doubletalk is let through as indicated by $e(n)$ and the ERLE in the figure.

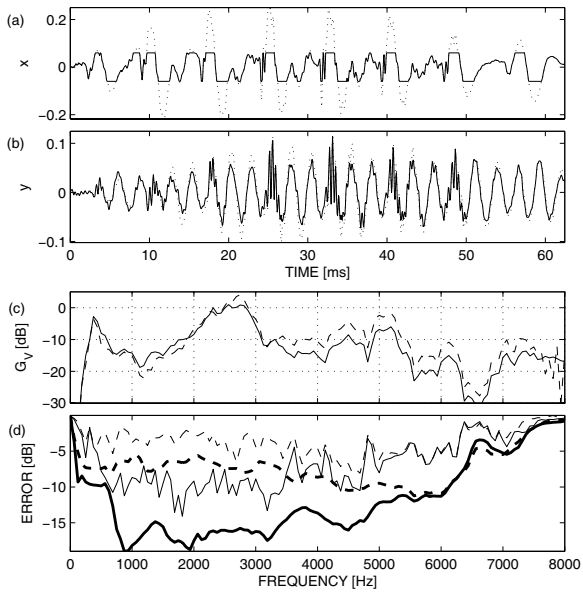

Fig. 4. (a) Original (dotted) and clipped (solid) loudspeaker signal. (b) Corresponding microphone signals. (c) Coloration effect filter computed from the true echo path (solid) and estimated for the case with clipping (dashed). (d) Echo magnitude prediction error (details see text).

4.2. Robustness

In a first experiment, we toggled between two echo path impulse responses every second. The two echo path responses used were measured for the two loudspeakers of a desktop stereo system. The resulting ERLE is shown in Figure 3. Comparison of this result with the bottom panel of Figure 2 indicates that the performance of the proposed algorithm is very similar for the case when echo path changes occur, indicating high robustness.

In a second experiment, we simulated non-linear distortion of a loudspeaker that is over-driven. That is, we simulated clipping. The clipping threshold was chosen such that the error caused by the clipping was 6 dB below signal level. Panel (a) in Figure 4 shows the original (dotted) and clipped (solid) loudspeaker signals. Panel (b) shows the corresponding microphone signals. Panel (c) shows the magnitude response of the coloration effect filter $G_V(i,k)$ directly computed from the true echo path (solid) and its least squares estimate for the case with clipping (dashed)[1]. Clipping results in that the loudspeaker emits less low frequencies and more high frequencies. This effect can clearly be seen: The estimated $G_V(i,k)$ is smaller for low frequencies and larger for high frequencies than the $G_V(i,k)$ computed from the

[1]Note that in this case the non-clipped loudspeaker signal and the microphone signal computed with the clipped loudspeaker signal are supplied to the AES and used for estimating $G_V(i,k)$.

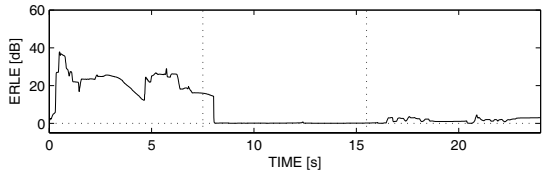

Fig. 5. The ERLE for the same simulation as shown in Figure 2, but simulating non-linear loudspeaker distortions (clipping).

true echo path.

The thin solid line in Panel (d) shows the prediction error of $G_V(i,k)$ computed directly from the echo path (thin solid). The prediction error of the least squares estimate of $G_V(i,k)$ for the case of clipping is indicated by the thin dashed line. Note that prediction performance is significantly affected by the clipping. The thick solid and dashed lines in Panel (d) show the same as the corresponding thin lines, but before computing $G_V(i,k)$ and the error, the STFT spectra $X(i,k)$ and $Y(i,k)$ were smoothed over frequency. Note that for both, with and without clipping, the error is significantly smaller. While the error is still significantly affected by the clipping, it is getting small enough that also for clipping a gain filter can be computed for effectively removing echo. Another advantage of the smoothing is that the resulting gain filter (2) changes smoothly as a function of frequency, resulting in less artifacts during doubletalk.

Figure 5 shows the ERLE of the proposed AES when it is supplied with the clipped signal. All parameters were the same as for the non-clipped simulations. Comparing this result with the bottom panel of Figure 2 indicates that the ERLE is lower than for the case without clipping. But the echo is still suppressed by about 20 dB. Considering the severeness of the clipping that we used, this result confirms the high robustness of the proposed algorithm. For increasing the ERLE, the parameters for the gain filter computations can be modified (2). But if β is chosen too large doubletalk quality will be more impaired.

4.3. Real-time implementation

We implemented a library of the proposed algorithm compatible with all major operating systems (Linux, MacOS X, Windows). Sampling rates up to 48 kHz are supported. For 16 kHz sampling rate the proposed AES consumes only about one percent of the processing power on a 1.6 GHz Pentium M processor. Informal testing with the real-time conferencing system indicates effectivity of the proposed algorithm in terms of echo suppression, doubletalk performance, and robustness.

5. CONCLUSIONS

A low complexity acoustic echo control algorithm, using a simple echo path model, was reviewed and analyzed. As opposed to identifying the acoustic echo path, only an overall delay parameter and a gain filter mimicking the coloration effect of the echo path are estimated. Given this, a spectral modification algorithm is applied for removing the acoustic echo signal from the microphone signal.

Numerical simulations and testing with a real-time implementation indicate that the proposed algorithm effectively suppresses echo and that it is robust against non-linearities and minor echo path changes. The computational complexity of the algorithm is very low.

6. REFERENCES

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

[2] G. W. Elko, E. Diethorn, and T. Gänsler, "Room impulse response variation due to thermal fluctuation and its impact on acoustic echo cancellation," in *Proc. Intl. Workshop on Acoust. Echo and Noise Control (IWAENC), Kyoto, Japan*, Sept. 2003.

[3] F. Wallin and C. Faller, "Perceptual quality of hybrid echo canceler/suppressor," in *Proc. ICASSP*, May 2004.

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

[5] C. Faller and J. Chen, "Suppressing acoustic echo in a sampled auditory envelope space," *IEEE Trans. on Speech and Audio Proc.*, vol. 13, no. 5, pp. 1048–1062, Sept. 2005.

[6] C. Faller and C. Tournery, "Estimating the delay and coloration effect of the acoustic echo path for low complexity echo suppression," in *Proc. Intl. Works. on Acoust. Echo and Noise Control (IWAENC)*, Sept. 2005.

[7] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE trans. Acoust. Speech Sig. Processing*, vol. 27, no. 2, pp. 113–120, Nov. 1979.

[8] W. Etter and G. S. Moschytz, "Noise reduction by noise-adaptive spectral magnitude expansion," *J. Audio Eng. Soc.*, vol. 42, pp. 341–349, May 1994.

[9] K. Ochiai, T. Araseki, and T. Ogihara, "Echo canceler with two echo path models," *IEEE trans. on Communications*, vol. 25, no. 6, pp. 589–595, June 1977.