An Efficient Long Distance Echo Canceller

Artur Ferreira Paulo Marques
Instituto de Telecomunicações - Pólo de Lisboa
Av. Rovisco Pais, 1049-001, Lisboa, PORTUGAL
Instituto Superior de Engenharia de Lisboa
Rua Conselheiro Emídio Navarro n.1, 1959-007 Lisboa, PORTUGAL
e-mail: arturj@cc.isel.ipl.pt pmarques@isel.ipl.pt

Abstract— **This paper describes an implementation of a long distance echo canceller, operating on full-duplex with hands-free and in real-time with a single Digital Signal Processor (DSP). The proposed solution is based on short length adaptive filters centered on the positions of the most significant echoes, which are tracked by time delay estimators, for which we use a new approach. To deal with double talking situations a speech detector is employed. The floating-point DSP TMS320C6713 from Texas Instruments is used with software written in C++, with compiler optimizations for fast execution. The resulting algorithm enables long distance echo cancellation with low computational requirements, suited for embbeded systems. It reaches greater echo return loss enhancement and shows faster convergence speed when compared to the conventional approach. The experimental results approach the CCITT G.165 recommendation levels.**

Keywords: echo cancellation, adaptive filtering, digital signal processing, digital signal processor.

I. INTRODUCTION

The existence of undesired echo is a well-known problem that frequently arises in telecommunications, being most problematic in voice conversations. Therefore, echo cancellation is needed for long-distance conversations such as those based on VoIP (Voice over Internet Protocol)[1], [2] teleconferencing and satellite communications, which are in growing use these days. The network load changes the transmission time and the time delay of the echo(s). On the other hand, the echo path is not static because the channel characteristics change over time. Well-known sources of echos are the two/four-wire conversion [3] in telephony and the VoIP setup.

In order to model the referred time-changing echo characteristics and to cancel its undesired effects on the conversation, adaptive filtering [4] has been used extensively in the past [3], [5], [6], [7]. The problem of echo cancellation has been addressed using the LMS (Least Mean Squares) algorithm and its variants [4]. Recently, an approach named frequency-response-shaped LMS (FRS-LMS) [8] was introduced and shown to have good convergence properties.

This work consists on an implementation of a long-distance echo canceller, using an up-to-date DSP architecture, the Texas Instruments TMS320C6713 [9], with code optimized for real-time processing. The echo canceller, built on adaptive filtering, has three building blocks: a centered adaptive filter to replicate the echo behavior, a time delay estimator and a speech detector. The filter adaptation is carried out only when speech is not present, according to the speech detector. Three different techniques are used for the time delay task: cross-correlation, euclidian distance and a maximum delay filter.

The paper is organized as follows. Section 2 presents the concepts associated with the echo cancellation problem. Section 3 presents the proposed system architecture, identifying its blocks. Section 4 describes the DSP hardware and software. Some experimental results of simulation and real-time operation are presented in section 5. Finally, Section 6 presents some concluding remarks and future work.

II. ECHO CANCELLATION

Echo cancellation is usually achieved using an adaptive filter which attempts to synthesize a replica of the echo signal and subtract it from the returned signal. This is the principle illustrated in figure 1 in the acoustic echo cancellation context. The received signal from the far end talker, $r_n[i]$, is transmitted through the speaker to the near end. A version of this signal is received from the microphone (due, for example, to direct coupling between the speaker and the microphone), together with the near end speech, constituting the received signal from the acoustic channel, $r_a[i]$.

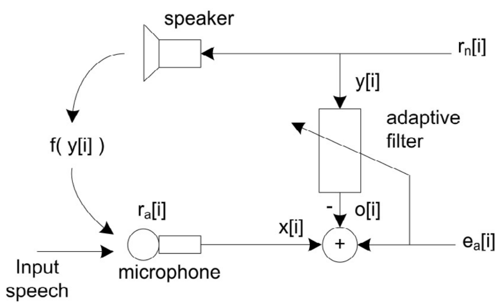

Fig. 1. The acoustic echo cancellation scenario.

To achieve the desired result, the maximum supported

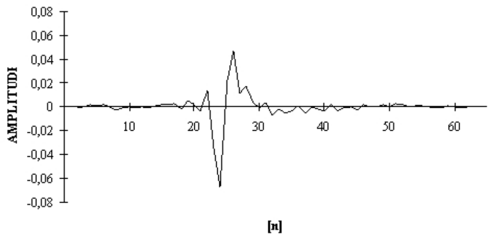

Fig. 2. Acoustic echo path impulse response for an IRISTEL telephone. Most of the coefficients are near zero.

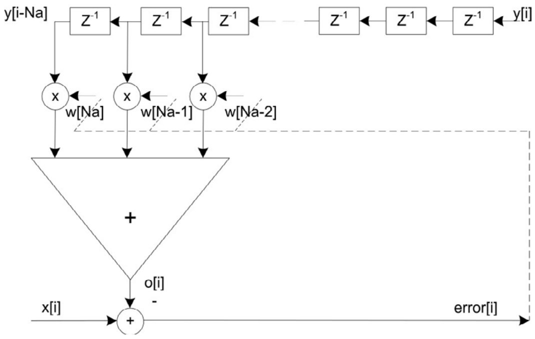

Fig. 3. Centered adaptive filter. The supported echo path length is N_a taps, but considering an active region of 3 taps, only the corresponding three coefficients need adjustment.

adaptive filter impulse response should have a length greater than the longest echo path that needs to be accommodated. For a delay long enough, the required computational effort may be such that its real time operation becomes compromised. Additionally, the adaptation step would have to be so small that the convergence speed would be unacceptable, and due to finite precision effects the achieved solution would be unsatisfactory. However, as stressed in [7], most of the echo path impulse response has a null value. See for example figure 2 which is the impulse response of an acoustic echo path, resulting from the direct coupling between the speaker and the microphone of an IRISTEL telephone. Although the supported echo path length is 64 delay elements, only a small region is active. Knowing its position and length, the adaptive filter has to adjust only the corresponding coefficients.

In figure 3 a centered adaptive filter example is shown. The supported echo path length is N_a taps, the position of the active region is $(N_a - 1)T_s$ and for illustration purposes only, the considered length is 3 taps. The main advantages are: reduced computational cost, due to the lower number of coefficients that need adjustment, when compared with the total supported echo path length; greater convergence speed, since the adaptation step can now be larger; reduced residual error because the coefficients which would otherwise converge to zero, now take precisely that value.

To adjust the short length adaptive filter coefficients we use a modified version of the Least Mean Squares (LMS) algorithm, the Normalized Least Mean Squares (NLMS), which

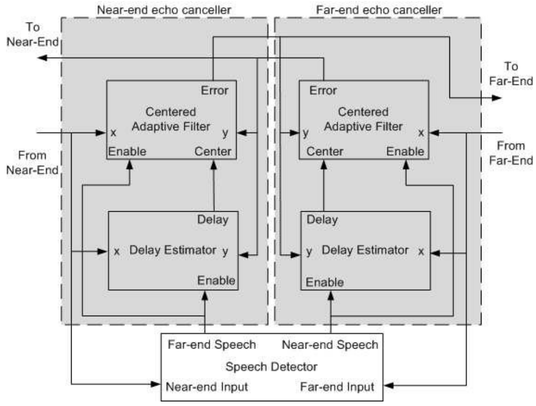

Fig. 4. The block diagram of the proposed system.

is more adequate than the common LMS in the presence of speech signals [4]. In the LMS algorithm, the filter coefficients $\mathbf{w}$ are updated according to

$$\mathbf{w}_{i+1} \leftarrow \mathbf{w}_i + 2\mu e_i \mathbf{x}_i, \tag{1}$$

where μ is the step size, e_i is the error signal at time i, $\mathbf{x}_i$ represents the previous samples and N is the length of the filter. Speech signals have a power which exhibit a large dynamic range, making ineffective the use of a constant step size. On the NLMS algorithm we have a variable step size, μ_i, which is computed by

$$\mu_i = \frac{\eta}{a + P_i}, \tag{2}$$

in which $\eta > 0$, P_i is the instantaneous power of signal $\mathbf{x}$ at time index i and $a > 0$ is a suitably chosen value to avoid numerical problems. This way, we have an adaptation step given by

$$\mathbf{w}_{i+1} \leftarrow \mathbf{w}_i + 2\mu(i) e_i \mathbf{x}_i, \tag{3}$$

for NLMS, which has a relatively low convergence speed but is quite stable and has low complexity. The number of coefficients that need adjustment is low when compared with the total number of elements in the supported delay line, thus enabling a larger step.

III. PROPOSED SYSTEM

Figure 4 shows the proposed system which is a combined echo cancellation structure which includes two echo cancellers, one for each communication direction, and a speech detector. Each echo canceller is composed by a centered adaptive filter and a time delay estimator. The delay estimator tracks the corresponding main signal reflection position where the short length adaptive filter is to be centered. The near-end and far-end speech detectors inhibit the adaptation of the filter whenever speech is present.

A. The echo canceller

The centered adaptive FIR filter has a small number of coefficients, corresponding to the length of the active area. We have considered the use of 32 and 64 coefficients, with a sampling rate of 8 kHz. This way, the filter presents fast convergence. Our tests also showed that is preferable to set to (absolute) zero all the filter coefficients that are near zero and to keep with a non-zero value, only those coefficients on the active area.

The delay estimator is based on a similarity measure (inner product) between the original signal and the echoed signal; we have used cross-correlation and euclidian distance. We also considere a new approach with a *maximum delay FIR filter*, that has a (large) number of coefficients $\mathbf{c} = [c_0, c_1, \ldots, c_L]$ corresponding to the maximum expected delay. These coefficients are updated by

$$\mathbf{c}_{i+1} \leftarrow \mathbf{c}_i + 2\mu e_i \mathbf{x}_i, \tag{4}$$

but only a small fraction ($p \ll L$) of these coefficients is updated between two consecutive sampling times (125 μs), in order to meet real-time requirements. For the typical scenario, there is no need to update the entire set of coefficients in order to get an accurate estimation of the time delay. The main component of the echo is given by the coefficient with the highest (absolute) amplitude value. Using the DSP parallel instructions [10], this update is carried out simultaneously with the update of the centered filter coefficients. In our tests, this delay estimator with low complexity, has obtained good results even in situations of low signal-to-noise ratio.

B. Speech detector

The speech detector copes with double talking situations. When both speakers are talking, the received signal is a composition of the received echo and the other speaker's signal. If no action is taken, the adaptive filter coefficient drift may happen as well as erroneous time delay estimation, so we have to inhibit the coefficients adjustment as well as delay estimation. The speech detector criterion, based on [11], is as follows:

- the far-end speaker is considered present when the power of the original signal is above a given threshold, established by the noise average power; we consider that the speaker is still present for more 75 ms, after the power level gets below this threshold;
- the near-end speaker is considered present when the power of the echoed signal is 6 dB below the power of the original signal.

Each filter is adapted only when the corresponding speaker is detected. This way, when speech is detected on both directions, echo cancellation is performed using the coefficients that were updated just before the presence of speech.

IV. IMPLEMENTATION DETAILS

The DSP hardware and software used in this work is briefly described in this section. The real-time implementation of the system was based on the floating-point Texas DSP TMS320C6713[1] [10], with a master clock of 225 MHZ, delivering 1800 MIPS and 1350 MFLOPS. The analog stereo interface is carried out by the AIC 23 codec, with sampling rates from 8 to 96 kHz, with 16, 20 and 24 bits per sample. The code, written in C++ programming language, is located on the 192 kB internal RAM, along with the data. We compiled the code with level-3 optimization [10], for faster execution.

[1]`www.c6000.spectrumdigital.com`

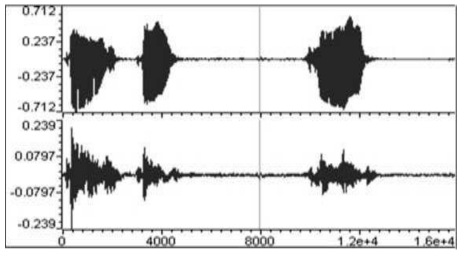

Fig. 5. Echo (top) and error (bottom) signal (simulation on CCS).

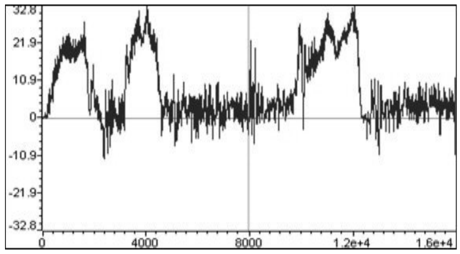

Fig. 6. Attenuation obtained for the speech signal of fig. 5 (simulation on CCS).

The filters are managed as circular buffers and inline functions are used whenever possible. The sampling rate is 8 kHz, and the number of bits per sample is 16 (the minimum allowed by the AIC23 codec), suited for speech signals. This way, we have 125 μs between two consecutive samples, and this is the maximum processing time to meet real-time requirements (28125 instructions, under a 225 MHz clock). The time delay estimator has the largest ammount of total processing time, being not possible to completely update the time delay estimation, within 125 μs. Between two consecutive samples, we update only a small portion of the filter coefficients.

V. EXPERIMENTAL RESULTS

This section presents results obtained with our system on simulated and real scenarios. The simulation tests were carried out in DSP Code Composer Studio (CCS) environment, with code written in C++, using real signals. Figure 5 displays the echo and error signals for a speech signal, while figure 6 displays the achieved attenuation, of about 20 dB, for the speech signal on its voiced parts. Table 1 compares our system with the CCITT G.165 recommendation, for a real situation, according to: CR - Convergence Rate; FERLAC - Final Echo Return Loss After Convergence; IRLC - Infinite Return Loss Convergence; LR - Leak Rate. We conclude that our system approaches the recommendation levels for FERLAC and IRLC measures, matches for CR and exceeds it for the LR measure. The CR and FERLAC measures are taken on the single-talk scenario. Table 2 summarizes the developed system features. The total ammount of memory needed for the

TABLE I
SYSTEM PERFORMANCE - COMPARISON WITH CCITT G.165

Measure	CCITT G.165	System
CR	$\geq$ 27 dB (500 ms)	27 dB (125 ms)
FERLAC	-40 dBm0	-37.39 dBm0
IRLC	-40 dBm0	-37.39 dBm0
LR	$\leq$ 10 dB	$\approx$ 0 dB

TABLE II
ECHO CANCELLER FEATURES

Feature	Value
Absolute delay	0.375 ms
Convergence speed	27 dB (125 ms)
Digitalization	$F_s = 8000$ 16 bit
Hold time after speech	75 ms
Max. length	256 ms
Max. length of dispersion area	4 ms
Max. memory (data + code)	$<$ 192 kB
Residual echo	-42.26 dBm0
Returned echo minimum loss	6 dB
Speech detector	6 dB below threshold

echo canceller data and code is low (and proportional to the maximum expected delay) making it suited for an embedded system. The total of memory required can be reduced, using a fixed-point DSP. The adaptive centered filters have 32 or 64 coefficients, while FIR-based time delay estimator uses up to L=4000 coefficients (delays up to 0.5 seconds), giving a reasonable range of delays, suited for several applications.

Figure 7 displays the attenuation obtained for several electric and acoustic echo paths, using the average power of the received echo as the reference value, because the attenuation on the acoustic channel is not constant along these tests. The attenuation for the simulated echo path is much larger than the other two, as expected. On the other hand, the attenuation for the electric echo path is around 30 dB, which is a considerable value. Finally, for the acoustic path we get about 10 dB of attenuation, which is also an acceptable practical value. This result was expected due to the strong non-linearities in the acoustic echo path, caused by the speakers and microphone.

VI. CONCLUSIONS

We presented and evaluated a real-time DSP based long distance full-duplex echo canceller system. The proposed system exploits the fact that the typical impulse response of the echo is sparse, therefore the echo path has an active region with a small fraction of non-zero coefficients. This way, short length adaptive FIR filters, centered on the estimated echo positions, assure an accurate echo path estimation. For time delay estimation we have evaluated three different algorithms: cross-correlation, euclidian distance and maximum delay FIR filter; the last one, makes the code quite simple, and achieves good results even in the scenario of low signal-to-noise ratio. A power based speech detector is employed to disable adjustments of the adaptive filters in the presence of speech.

The code, written in C++, for TMS320C6713 floating point DSP of Texas Instruments, was optimized to operate in full-duplex in real-time, using a sampling rate of 8 kHz. The echo canceller code is modular, being possible to use it in different DSP architectures, with minor modifications. The data and code modules rely on the 192 kB internal memory of the DSP, being suitable for applications with low computational requirements, supporting delays up to 0.5 seconds.

The developed system approaches the CCITT G.165 recommendation, with long echo paths and shows fast convergence. We have better echo cancellation results for the electric echo path than for the acoustic path, given the presence of non-linearities on the later. As future work we consider to model the non-linearities of the echo path and to modify the speech detector in order to accommodate higher delays.

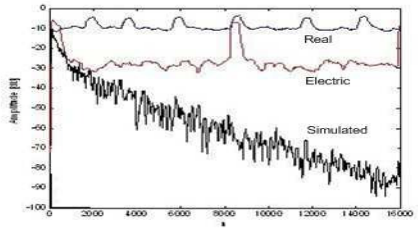

Fig. 7. Attenuation for the echo paths real (acoustic), electric and simulated (real-time on CCS).

REFERENCES

[1] R. Jeanns, P. Scalart, G. Faucon, and C. Beaugeant, "Combined noise and echo reduction in hands-free systems: A survey," *IEEE Transactions on Speech And Audio Processing*, vol. 9, no. 8, pp. 808–820, November 2001.

[2] W. Witowsky, "Understanding echo cancellation in voice over IP networks," in *International Conference on Signal Processing Applications and Technology (ICSPAT)*, Orlando, November 1999.

[3] M. Sondhi and D. Berkeley, "Silencing echoes on the telephone network," *Proceedings of the IEEE*, vol. 68, no. 8, pp. 948–963, August 1980.

[4] S. Haykin, *Adaptive Filter Theory*, 4th ed. Prentice-Hall, 2002.

[5] J. Benesty, T. Gaensler, D. Morgan, M. Sondhi, and S. Gay, *Advances in Network and Acoustic Echo Cancellation*. Springer-Verlag, 2001.

[6] J. Greenberg, "Modified LMS algorithms for speech processing with an adaptive noise canceller," *IEEE Transactions on Signal Processing*, vol. 6, no. 4, pp. 338–351, July 1998.

[7] P. Marques, F. Sousa, and J. Leitao, "A DSP based long distance echo canceller using short length centered adaptive filters," in *IEEE International Conference on Acoustics, Speech, and Signal Processing - ICASSP'97*, Munich, April 1997.

[8] O. Kukrera and A. Hocanin, "Frequency-response-shaped LMS adaptive filter," *Digital Signal Processing*, vol. 16, no. 6, pp. 855–869, November 2006.

[9] Spectrum Digital Inc., DSP Development Systems, "TMS320C6713 DSK technical reference," May 2003.

[10] N. Kehtarnavaz, *Real-Time Digital Signal Processing*. Newnes, 2004.

[11] D. Messerschmidt, D. Hedberg, C. Cole, A. Haoui, and P. Winship, "Digital voice echo canceller with a TMS320C30," *Digital Signal Processing Applications*, vol. 1, pp. 415–454, 1986.