

Disclaimer: This document was part of the DSP Solution Challenge 1995 European Team Papers. It may have been written by someone whose native language is not English. TI assumes no liability for the quality of writing and/or the accuracy of the information contained herein.

Designing an Echo Canceller System Using the TMS320C50 DSP

*Authors: Paulo Marques (pmarques@cc.isel.pt)
Fernando Sousa (fsousa@cc.isel.pt)
Instituto Superior de Engenharia de Lisboa*

*Digital Signal Processing Solutions
September 1996
SPRA322*



IMPORTANT NOTICE

Texas Instruments (TI) reserves the right to make changes to its products or to discontinue any semiconductor product or service without notice, and advises its customers to obtain the latest version of relevant information to verify, before placing orders, that the information being relied on is current.

TI warrants performance of its semiconductor products and related software to the specifications applicable at the time of sale in accordance with TI's standard warranty. Testing and other quality control techniques are utilized to the extent TI deems necessary to support this warranty. Specific testing of all parameters of each device is not necessarily performed, except those mandated by government requirements.

Certain application using semiconductor products may involve potential risks of death, personal injury, or severe property or environmental damage ("Critical Applications").

TI SEMICONDUCTOR PRODUCTS ARE NOT DESIGNED, INTENDED, AUTHORIZED, OR WARRANTED TO BE SUITABLE FOR USE IN LIFE-SUPPORT APPLICATIONS, DEVICES OR SYSTEMS OR OTHER CRITICAL APPLICATIONS.

Inclusion of TI products in such applications is understood to be fully at the risk of the customer. Use of TI products in such applications requires the written approval of an appropriate TI officer. Questions concerning potential risk applications should be directed to TI through a local SC sales office.

In order to minimize risks associated with the customer's applications, adequate design and operating safeguards should be provided by the customer to minimize inherent or procedural hazards.

TI assumes no liability for applications assistance, customer product design, software performance, or infringement of patents or services described herein. Nor does TI warrant or represent that any license, either express or implied, is granted under any patent right, copyright, mask work right, or other intellectual property right of TI covering or relating to any combination, machine, or process in which such semiconductor products or services might be or are used.

TRADEMARKS

TI is a trademark of Texas Instruments Incorporated.

Other brands and names are the property of their respective owners.

CONTACT INFORMATION

US TMS320 HOTLINE	(281) 274-2320
US TMS320 FAX	(281) 274-2324
US TMS320 BBS	(281) 274-2323
US TMS320 email	dsph@ti.com

Contents

Abstract	7
Product Support on the World Wide Web	8
Introduction	9
Conceptual Approach	11
Adaptive Filtering in Echo Cancellation	11
Coarse Time-Delay Estimation	13
Speech Detection	15
DSP Implementation	16
Centered Adaptive Filter	16
Coarse Time-Delay Estimator	18
Speech Detector	21
Hardware and Application	24
Performance Evaluation	26
Summary	29
Acknowledgements	29
References	29

Figures

Figure 1.	Echo Canceller Based on Low Resolution Time-Delay Estimation	10
Figure 2.	Acoustic Echo Cancellation Using an Adaptive Filter	11
Figure 3.	Acoustic Echo Path Impulse Response for an IRISTEL Telephone	12
Figure 4.	Centered Adaptive Filter	12
Figure 5.	Direct Cross-correlation Method (DC)	14
Figure 6.	Centered Adaptive Filter Algorithm.....	17
Figure 7.	Coarse Resolution Time-Delay Estimator.....	18
Figure 8.	Time-Delay Estimation as a Function of Window Length and Correlation Memory Length (D=12.5ms).....	19
Figure 9.	Coarse Time-Delay Estimation Using a Real Speech Signal	20
Figure 10.	Speech Detector Flowchart	22
Figure 11.	Speech Detector in the Presence of Real Speech Signals	23
Figure 12.	Coefficient Drift in the Presence of Double Talk with the Speech Detector Disabled.....	24
Figure 13.	Echo Canceller ERLE in Simulated and Real Echo-Paths	27
Figure 14.	Echo Canceller Working with Real Speech Signals	28

Tables

Table 1.	System Requirements	25
Table 2.	Echo Canceller Performance.....	26

Designing an Echo Canceller System Using the TMS320C50 DSP

Abstract

This application report describes an echo canceller system using the Texas Instruments (TI™) TMS320C50 digital signal processor (DSP). The system is based on low resolution time-delay estimation, using the knowledge that mainly one or two active regions (depending on the number of reflections in the communication channel) characterize the impulse response of a typical echo-path in telephony communications.

These active regions usually have a duration that is only a small fraction of the total supported length. Low-resolution time-delay estimators track those active regions where short-length adaptive filters are centered, assuring an accurate echo-path estimation.

A speech detector is employed to avoid erroneous time-delay estimates and adaptive filter coefficient drift. Using a single TMS320C50 DSP with no external memory, the system detects and cancels an echo with a delay of more than 380 ms. Considering a configuration with 64Kwords of data memory, the maximum supported delay is greater than 2.5s.

The resulting system enables long delay echo cancellation with reduced computational effort, while achieving greater convergence speed and lower residual error, when compared with other systems reported in recent literature.

This document was part of the first European DSP Education and Research Conference that took place September 26 and 27, 1996 in Paris. For information on how TI encourages students from around the world to find innovative ways to use DSPs, see TI's World Wide Web site at www.ti.com.



Product Support on the World Wide Web

Our World Wide Web site at www.ti.com contains the most up to date product information, revisions, and additions. Users registering with TI&ME can build custom information pages and receive new product updates automatically via email.



Introduction

With the development of ISDN, mobile digital telephones, and teleconferencing systems, echo cancellation has become an important consideration. Delays caused by temporal multiplexing, channel coding, and speech coding are significant and compare with delays in long-distance communications. In satellite communications, far-end echoes can have delays greater than half a second.

The typical solution uses an adaptive FIR filter with more than 4000 taps (using the standard 8KHz sampling rate) and the same number of associated coefficients. The computational effort needed makes this kind of echo canceller uncommon in low-cost products.

However, mainly two active regions, corresponding to the near- and far-end signal echoes, characterize the echo-path impulse response. Intermediate delays are possible but not consistent.¹ The length of the active regions is usually much shorter than the total echo-path length. The proposed system shown in Figure 1 uses two echo cancellers, one for each communication direction, and a speech detector. A centered adaptive filter and a time-delay estimator compose each echo-canceller. The delay estimator tracks the position of the most significant signal reflection on each direction.

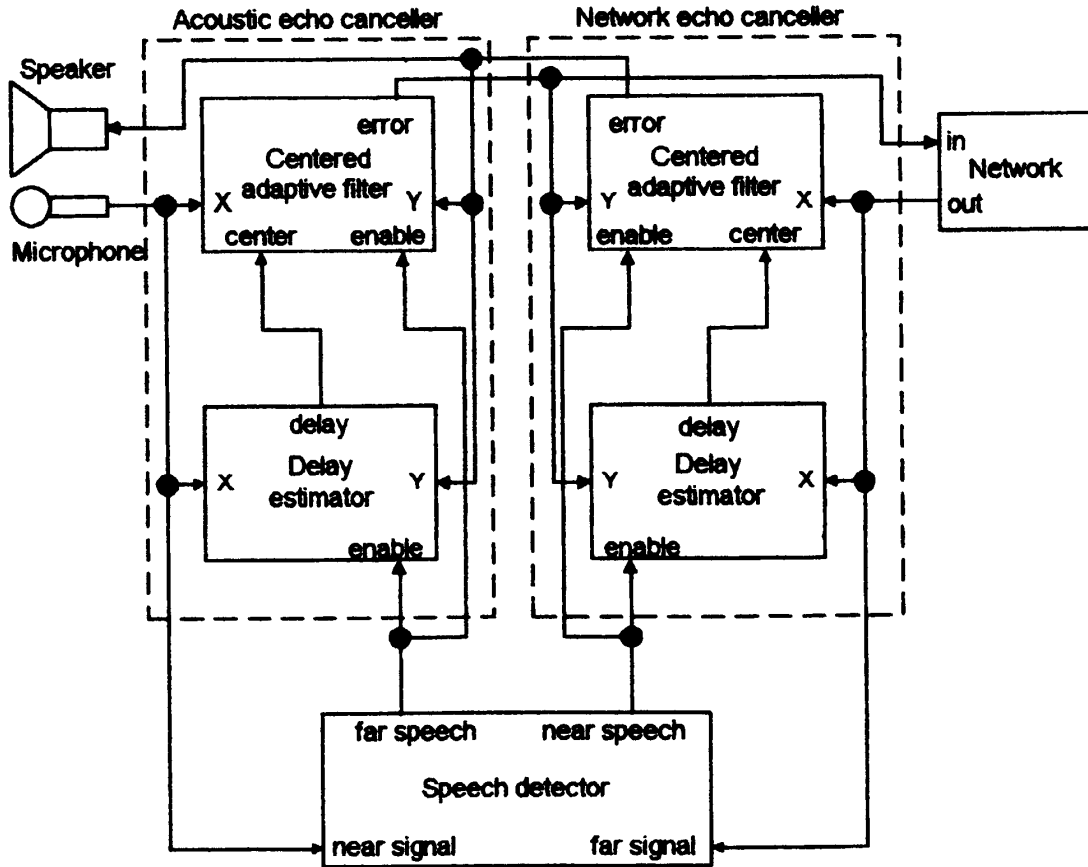
Based on that information the corresponding short-length adaptive filter is centered. The speech detector avoids erroneous delay estimates and adaptive filter coefficients drift. It inhibits the filter coefficients adaptation and delay estimation in unfavorable situations. Compared with the conventional solution, the resulting algorithm has improved computational effort and achieves greater convergence speed and reduced residual error.

The support of long echo path lengths is notorious. Using a single TM5320C50 DSP, without any external memory, the system is able to detect and cancel an echo with a delay larger than 380 ms. Considering a configuration with 64Kwords of data memory the maximum supported delay is greater than 2.5s.

Measurements were made in simulated and real situations. The system exceeds the CCITT G.165 recommendation in simulated echo paths. The echo return loss enhancement (ERLE) is of 41dB in a simulated echo path, 24.5dB in an electrical path and 19.2dB in an acoustic echo path.

A prototype of a secure speech communication system that includes a version of this echo canceller is currently being tested.

Figure 1. Echo Canceller Based on Low Resolution Time-Delay Estimation



Conceptual Approach

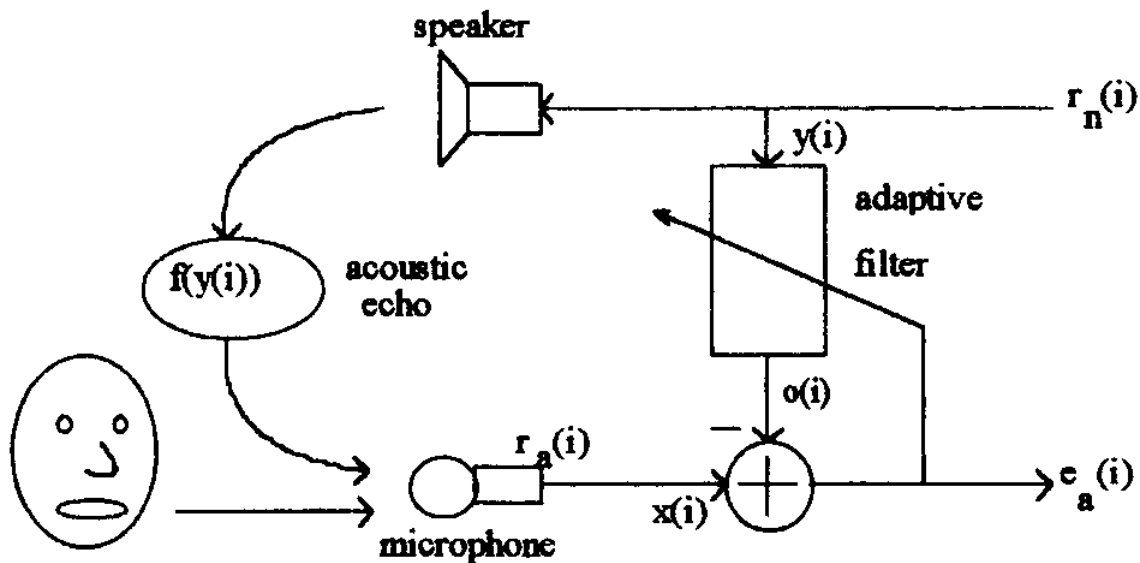
This section discusses the concepts, problems, and solutions involved with echo cancellation.

Adaptive Filtering in Echo Cancellation

Echo cancellation is usually achieved using an adaptive filter. The adaptive filter synthesizes a replica of the echo signal and subtracts it from the returned signal (see Figure 2).

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, $r_a(i)$.

Figure 2. Acoustic Echo Cancellation Using an Adaptive Filter



To achieve the desired result, the maximum supported 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 real time operation of the required computational effort can become compromised.

Additionally, the adaptation step would have to be so small that the convergence speed would be unacceptable.² The finite precision effects would make the achieved unsatisfactory.

However, as stressed by Margo and Etter, most of the echo path impulse response has a null value.¹ Figure 3 shows 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 is active. Knowing its position and length, the adaptive filter only has to adjust the corresponding coefficients.

Figure 3. Acoustic Echo Path Impulse Response for an IRISTEL Telephone

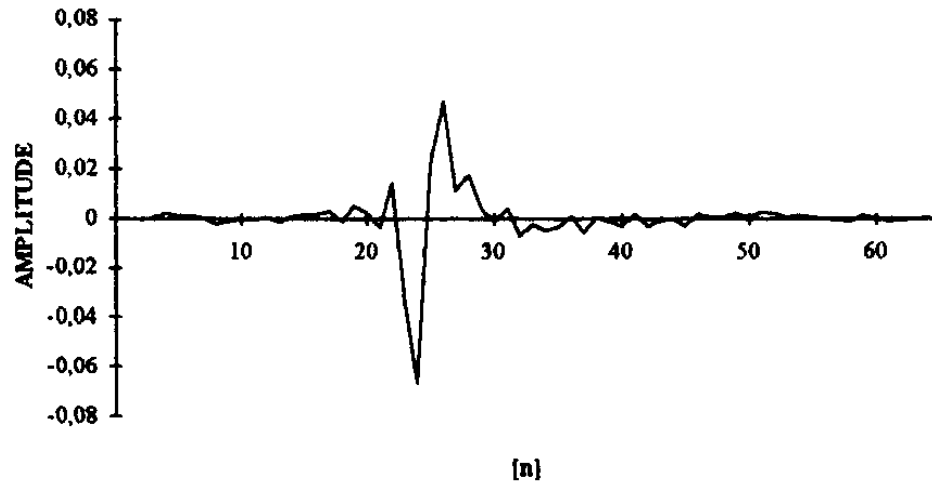
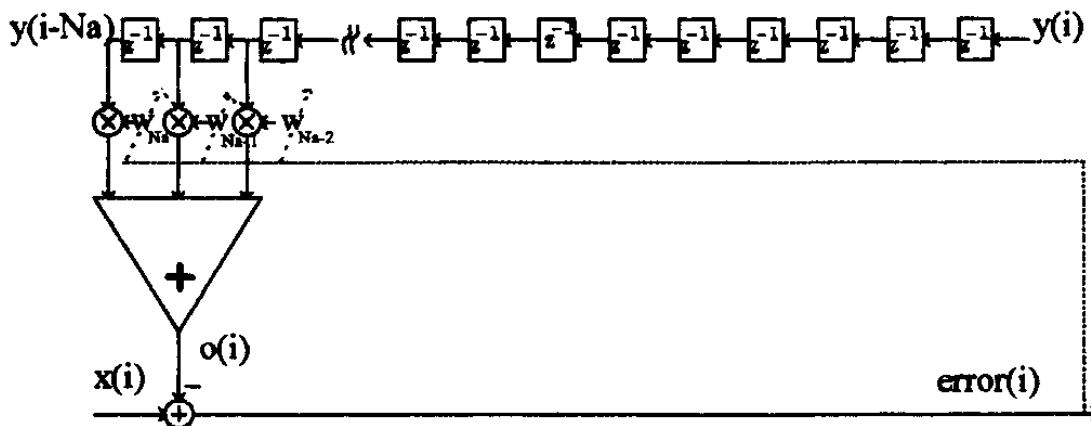


Figure 4 shows a centered adaptive filter example. The supported echo path length is N_a taps, the position of the active region is $(N_a-1)T_s$ and the considered length is 3 taps.

Figure 4. Centered Adaptive Filter





The main advantages are:

- ❑ Reduced computational cost because of the lower number of coefficients needing adjustment when compared to the total supported echo path length
- ❑ Greater convergence speed because the adaptation step can now be larger^{1 2}
- ❑ Reduced residual error because coefficients that 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 called the Normalized Least Mean Squares (NLMS), which is more adequate than the common LMS in the presence of speech signals.² These signals have a power that exhibits a large dynamic range and makes a constant step size ineffective. The variable step size, $\mu(i)$, is computed using equation (1), in which P_x is the signal power and a suitably chosen value to avoid numerical problems.

$$\mu(i) = \frac{\xi}{a + P_x(i)}, \text{ with } a > 0 \quad (1)$$

The low convergence speed is the main weakness of this algorithm when compared with others.² However, this problem is minimized by the lower number of coefficients needing adjustments compared to the total number of memorized elements in the supported delay line, thus enabling a larger step.

Coarse Time-Delay Estimation

The correct estimation of the involved echo delays is an extremely important subject in this work. It is supported on this knowledge that the short length adaptive filters can be centered. Jacovitti and Scarano investigate the basic aspects of time delay estimation (TDE) based on sampled signals.³ The direct cross-correlation method (DC) is analyzed and compared to the average square difference function (ASDF) and the average magnitude difference function (AMDF).

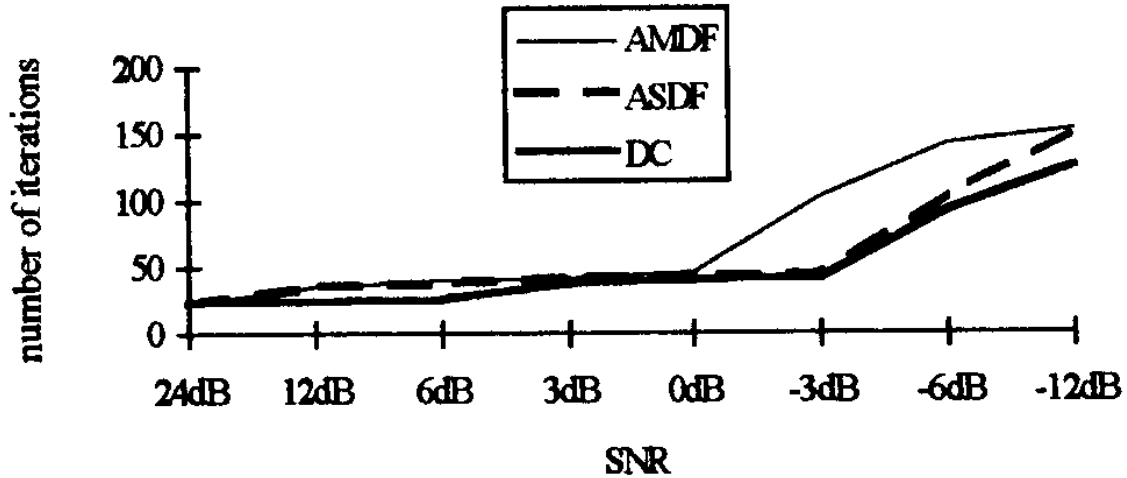
We have chosen the direct cross-correlation method, which outperforms the others for low signal-to-noise ratios (SNR).³ Figure 5 shows the supporting simulation results for a simulated echo path with a transfer function computed by Jacovitti and Scarano (adopted from Margo and Etter).^{1 3}

$$H(z) = K(z^{-15} - z^{-16} - z^{-17} - z^{-18} - z^{-19}) \quad (2)$$

This transfer function simulates a band-pass channel, where K is calculated to achieve the desired channel echo return loss (ERL).

The SNR was decreased from 24dB to -12dB, for a 10dB ERL. Although for high SNRs, the number of iterations needed to estimate the correct delay is similar to all the considered methods, the same is not true for lower SNRs, where the DC estimator gives the correct result in less iterations.

Figure 5. Direct Cross-correlation Method (DC)



TDE problems usually require an accurate value of the considered delay. However, we just need an estimate with a sample period resolution, which avoids the need for interpolation. Thus, the technique consists of searching the absolute extreme of the computed cross-correlation function. For acoustic echo cancellation, in the proposed system, the resulting estimated delay is given by equation (3) which selects the absolute maximum from the cross-correlation function computed by equation (4), where N_a is the number of memorized elements from the received network signal $e_n(t)$.

$$\hat{D}DCa = \arg \max \hat{R}DCa(\tau) \quad (3)$$

$$\hat{R}DCa(\tau) = \frac{1}{N_a} \sum_{k=1}^{N_a} e_n(kT + \tau) r_a(k\tau) \quad (4)$$



Speech Detection

Speech detection is a serious issue in echo cancellation.^{4 5 6} When there is double-talking (that is, both speakers are talking), the received signal is a composition of the received echo and the signal from the other speaker. If no action is taken, adaptive filter coefficient drift can occur. Additionally, in the proposed system, double-talking can originate erroneous time-delay estimation.⁷ The strategy is to inhibit the adaptation and the time-delay estimation when double-talking is detected.

A speech detector is developed to avoid the referred problems. A version of the algorithm presented by Messerschmitt et al. is used.⁴

Let $r_a(i)$ and $r_n(i)$ be the received signal from the acoustic channel and the received signal from the network, respectively. Near-end speech is declared when equation (5) is verified.

$$|r_a(i)| \geq \frac{1}{2} \max\{|r_n(i-1)|, \dots, |r_n(i-N_a)|\} \quad (5)$$

It is necessary to compare the received signal from the acoustic channel, r_a , with the recent past of the far-end signal, r_n , because of the unknown delay in the echo path. The factor 1/2 is based on the hypothesis that the echo path return loss (ERL) is at least 6dB.

Far-end speech is declared based on the same principle:

$$|r_n(i)| \geq \frac{1}{2} \max\{|r_a(i)|, |r_a(i-1)|, \dots, |r_a(i-N_a)|\} \quad (6)$$

If the supported delays are significant, equations (5) and (6) can be computationally demanding. So the comparison is made with recursively computed power estimates, as given by:

where $0 < \alpha < 1$. As the detection is based on power peaks, it is desirable to continue declaring its presence for some time, after detection. According to Messerschmitt et al., 75ms is an adequate value.⁴

DSP Implementation

This section describes optimizations made to the adaptive filter, the coarse time-delay estimator, and the speech detector.

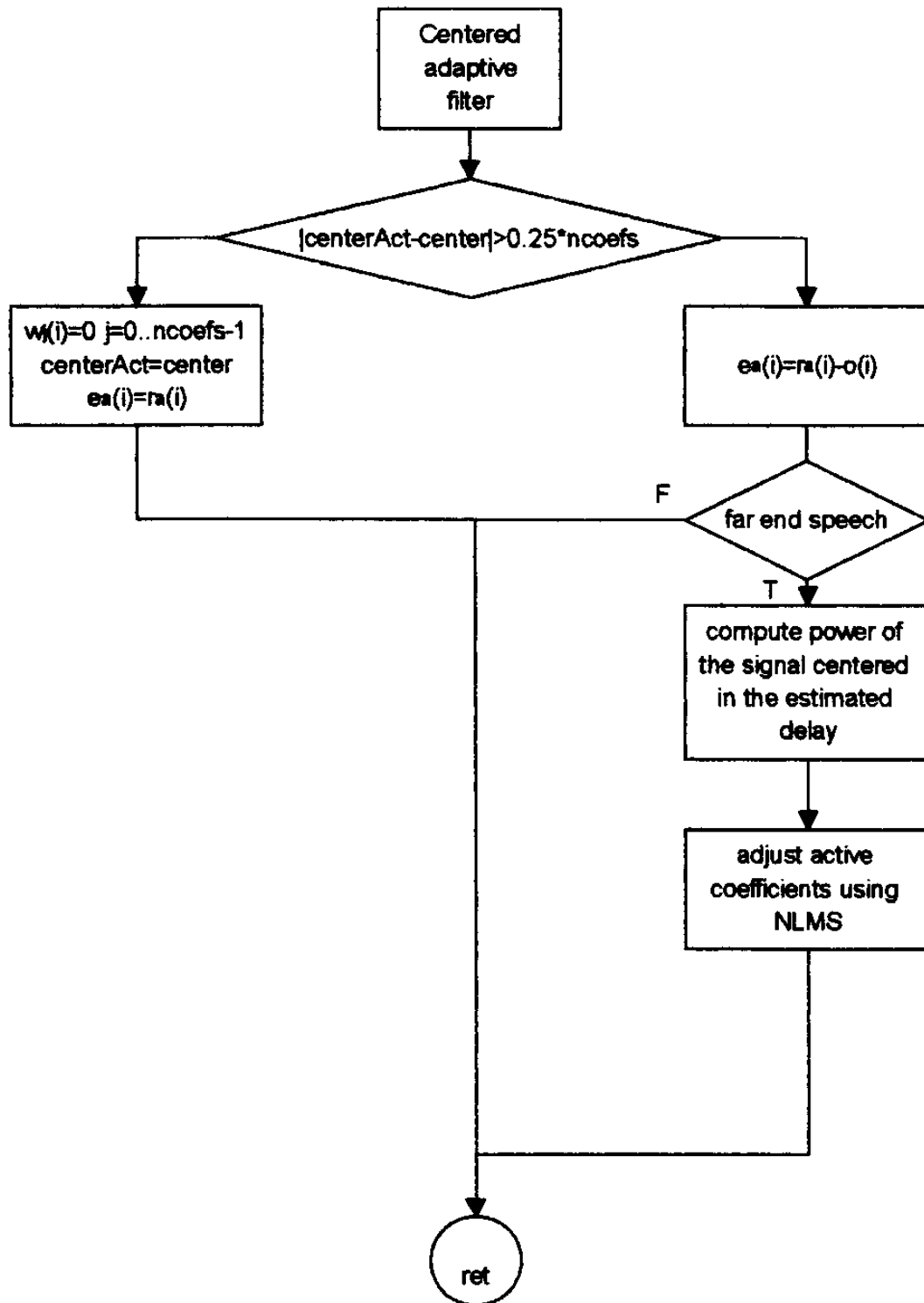
Centered Adaptive Filter

The signal power for the NLMS step calculation is recursively computed, based on the actual delay estimation. Supported by Messerschmitt et al. and real situation tests, the considered active region of the echo path impulse response is 2ms.⁴ We doubled that value to avoid a little change in the delay that could force the adaptive filter position change.

Figure 6 shows the centered adaptive filter algorithm. In each iteration, the adaptive filter checks if the estimated delay (named center) has moved more than 25% of the total active region number of taps. The new position is accepted (centerAct=center) only in that case.

If the position is to be maintained and the speech mode is correct (far end speech in the case of acoustic echo cancellation), the echo estimate, $o(i)$, is computed. The considered power for the step size computation is based on the signal centered in the estimated delay.

Figure 6. Centered Adaptive Filter Algorithm



Coarse Time-Delay Estimator

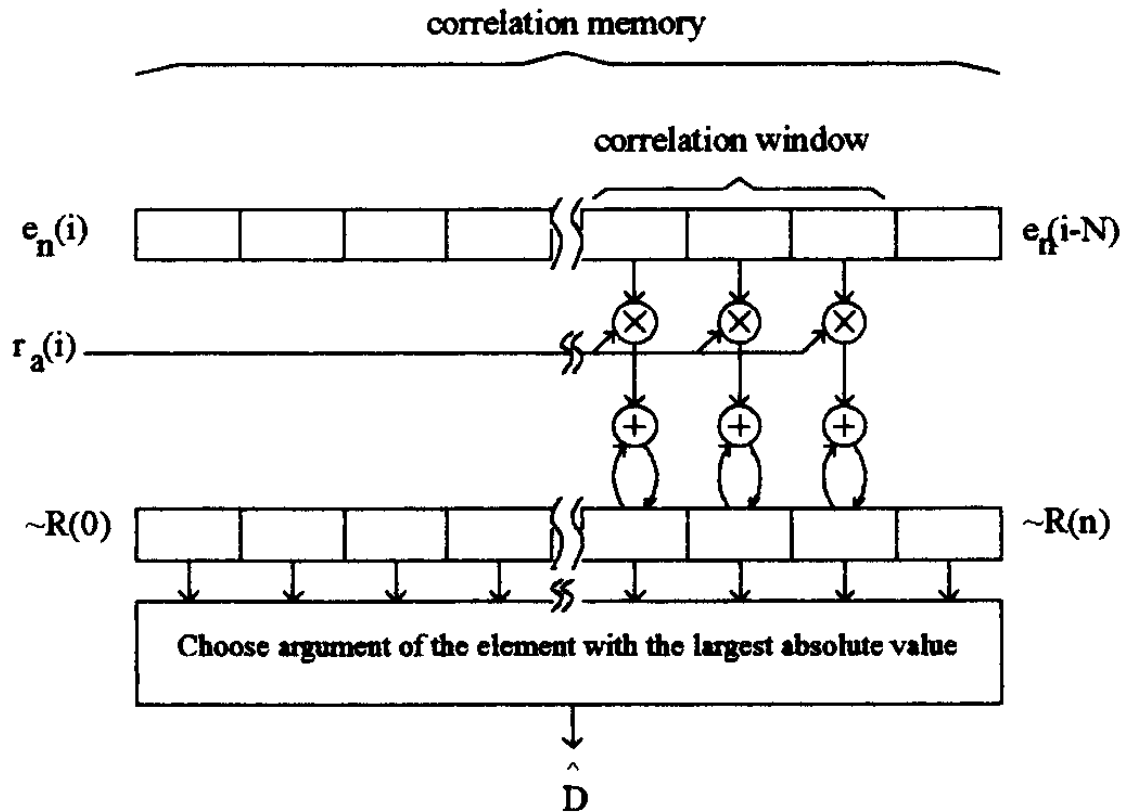
The coarse time-delay estimator is based on the direct cross-correlation method (DC) and uses a 32-bit precision for the cross-correlation computation. The delay line is shared with the centered adaptive filter, thus optimizing the used memory and processor requirements, as only one delay line is used and updated in each direction.

The average computation in equation (4) is not done to allow the contribution of small values that otherwise would be considered zero. Equation (9) is computed instead.

$$\hat{R}DCa(\tau) = \sum_{k=1}^{N_a} e_n(kT + \tau)r_a(kT) \quad (9)$$

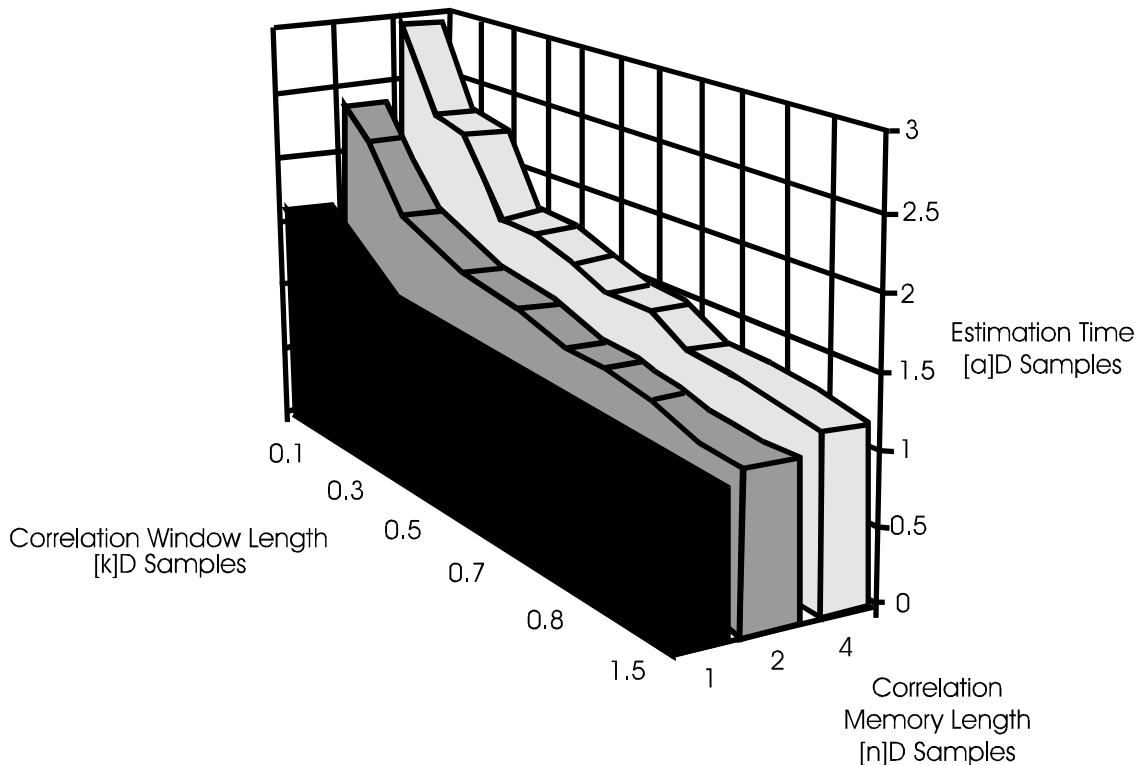
Figure 7 shows the computational model. When the maximum value for the cross-correlation is achieved in any of the function elements, all of them are rescaled, thus holding the function shape. This is important to keep the reflection position information.

Figure 7. Coarse Resolution Time-Delay Estimator



To enable the system real time operation when a long delay is supported, the correlation memory is subdivided in fractions named correlation windows. When a new sample is received, the cross-correlation is computed only with the next correlation window instead of the whole correlation memory. This implies the degradation in the estimator performance, as seen in Figure 8.

Figure 8. Time-Delay Estimation as a Function of Window Length and Correlation Memory Length ($D=12.5\text{ms}$)



Two values must be chosen:

- Correlation memory length
- Correlation window length

The first is limited by the available memory and should not exceed the previously known maximum signal delay, as this degrades the estimator performance. The second value is bounded by the available processor power.

Figure 9. Coarse Time-Delay Estimation Using a Real Speech Signal

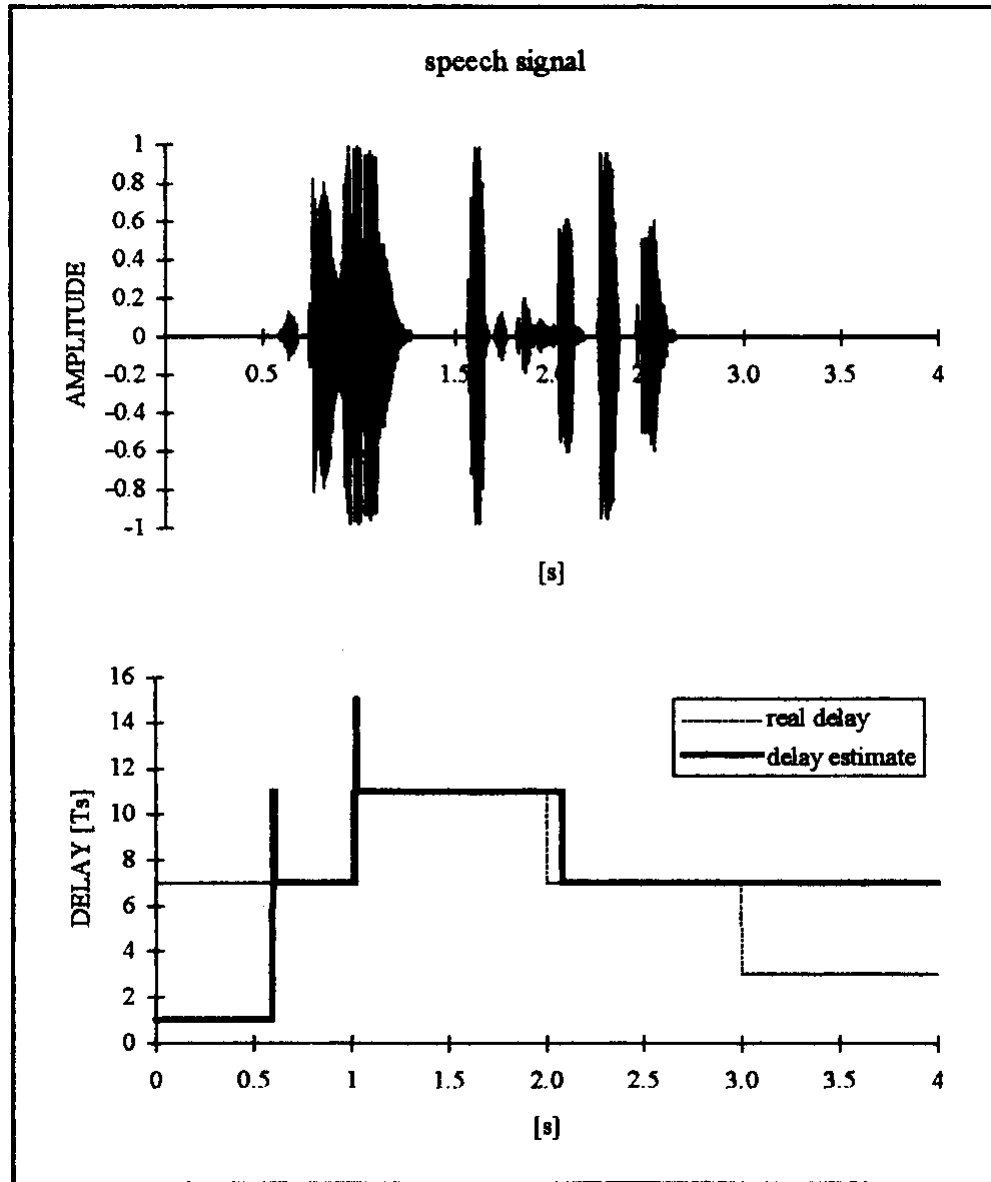


Figure 9 shows the coarse time-delay estimator working with a real speech signal. The correct value is achieved only when there is non-null signal. The correct delay is not detected in the first 0.5 second and in the last second.

The speech power is also important, as seen by the time taken to make the detection in the third second.



Speech Detector

Figure 10 shows the speech detector flowchart. Four modes of operation are discriminated:

- Idle
- Near end speech
- Far end speech
- Double talking

Idle is assumed when none of the signal powers exceeds a noise threshold (**min**), experimentally determined.

Near- or far-end speech is declared when the (estimated) power of the near-end or far-end signal, respectively, exceeds the other by 6dB. The corresponding 75ms counter (**nearCnt** or **farCnt**) is started.

Double talking is assumed when both signal powers are above the noise threshold but one does exceed the other by at least 6dB.

When idle or double-talking is declared in the echo cancellation system, the adaptive filters adaptation and time delay estimators are frozen.

When the situation is of near- or far-end speech, the corresponding filter adaptation and time delay estimator are enabled.

Figure 10. Speech Detector Flowchart

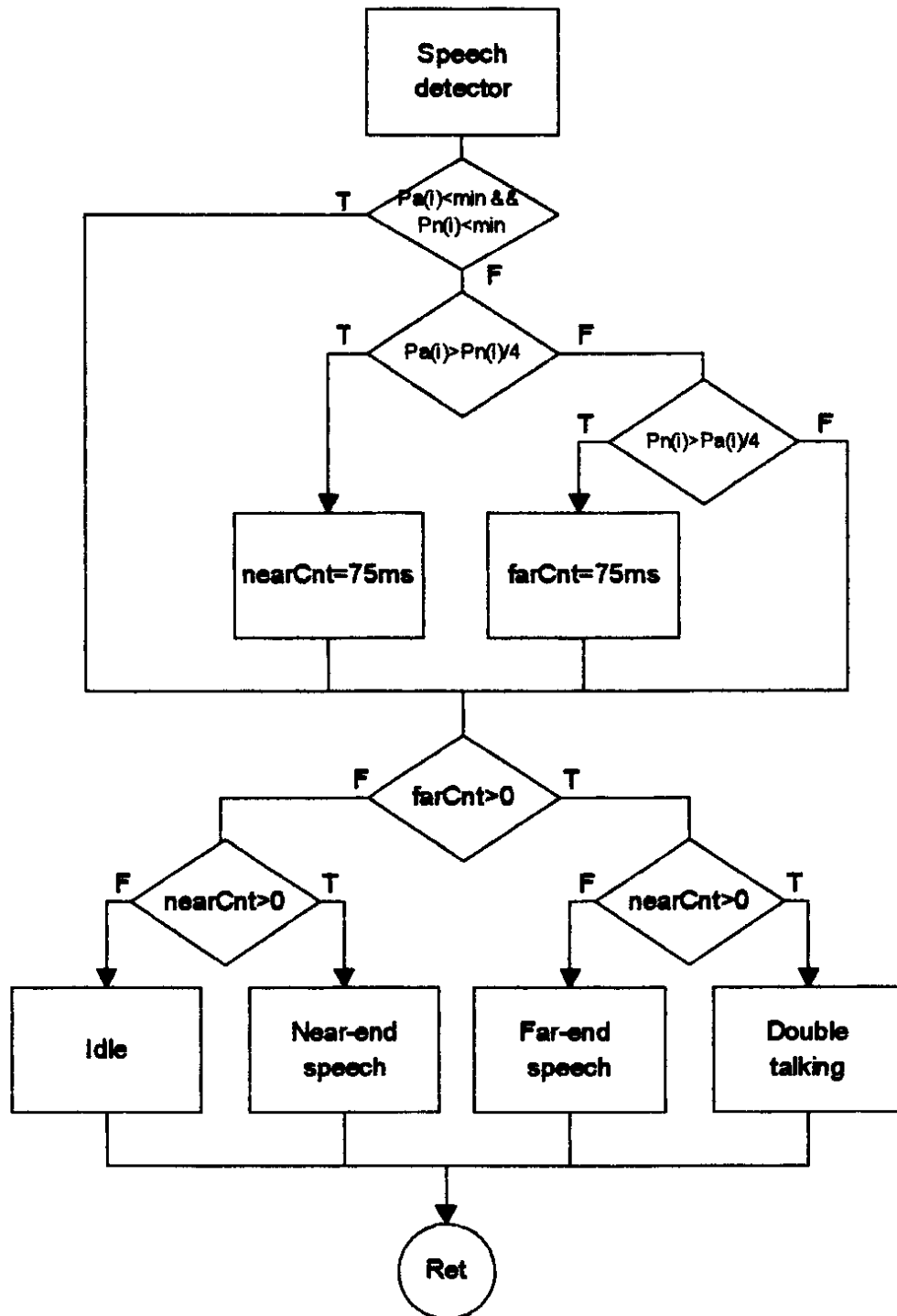


Figure 11 shows the developed speech detector working with real speech signals. Two speech segments with duration of four seconds excite the detector. The four possible situations are correctly determined. Notice that to improve intelligibility, only the near and far end speech detection signals are drawn.

Figure 11. Speech Detector in the Presence of Real Speech Signals

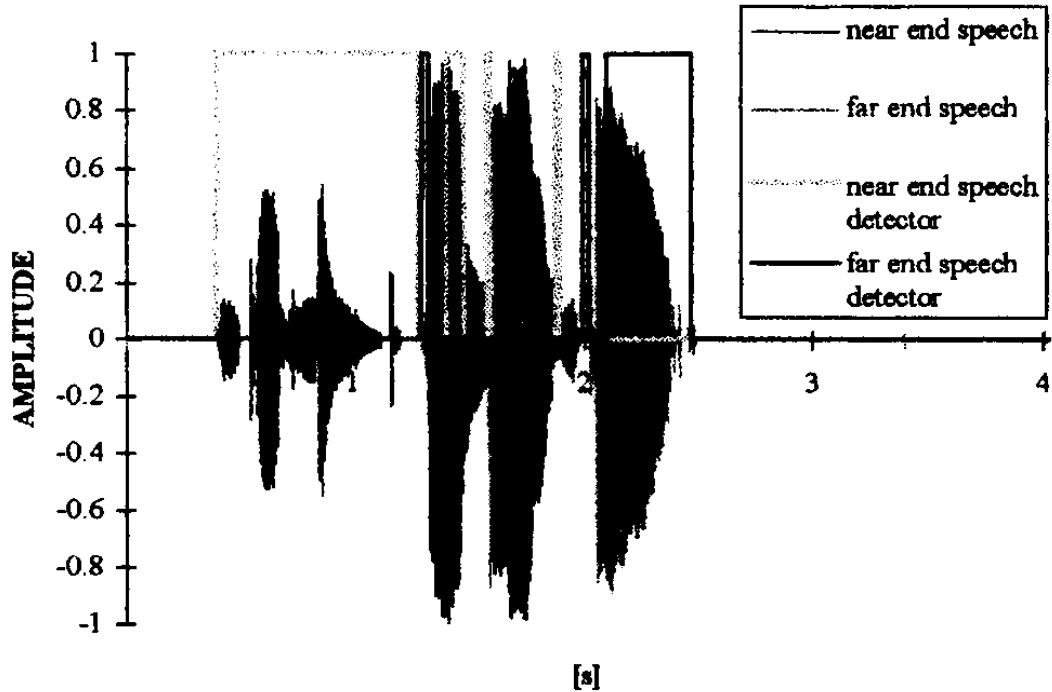
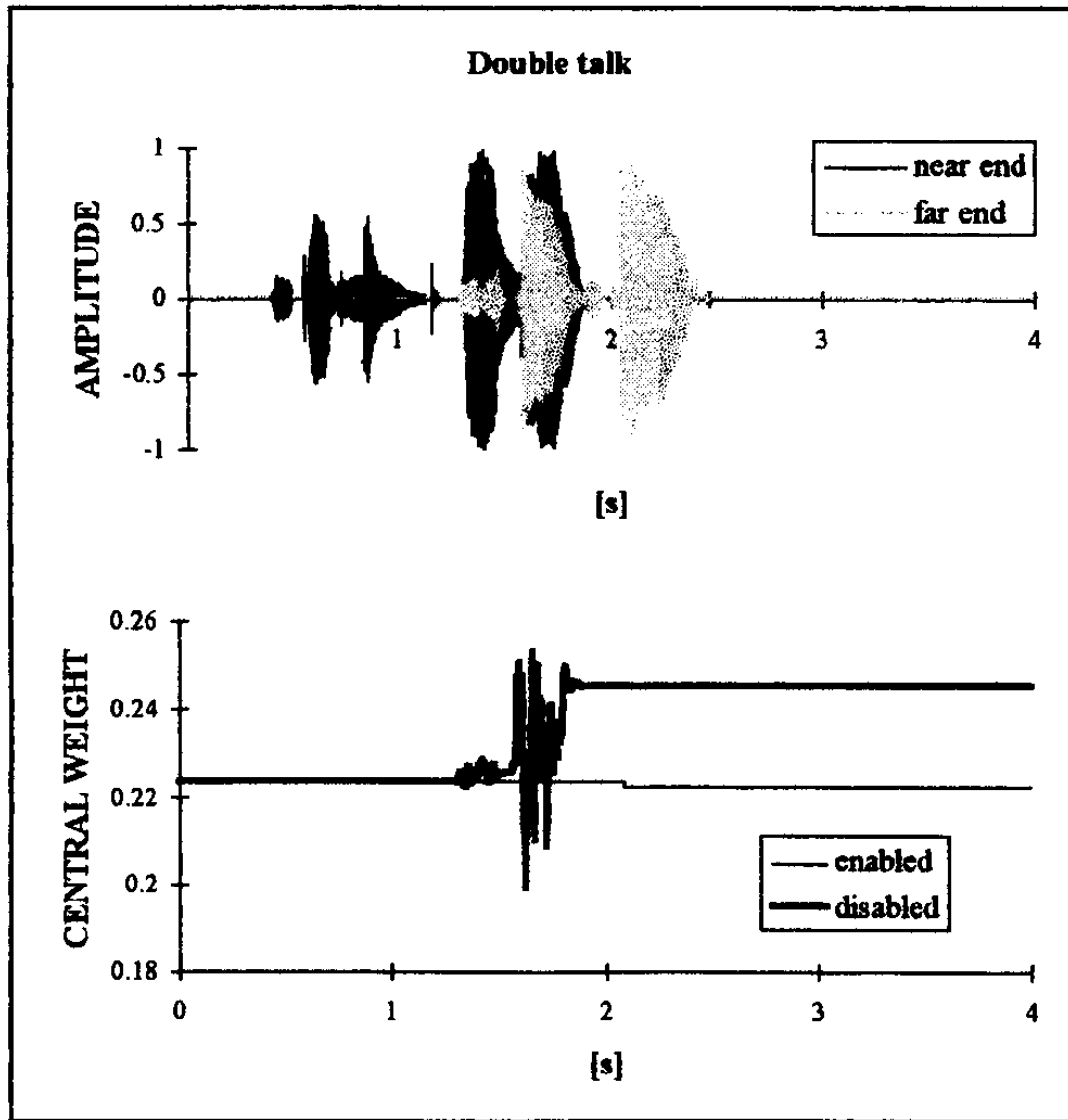


Figure 12 shows the usefulness of the speech detector to prevent the coefficient drift of the adaptive filter. The adaptive filter was initiated with optimal weights for the simulated echo path. In the presence of double-talking and with the speech detector disabled the central weight drift is notorious.

Figure 12. Coefficient Drift in the Presence of Double Talk with the Speech Detector Disabled



Hardware and Application

The echo cancellation system was implemented and tested using the TI TM5320C50 DSP Starter Kit. The application code was developed using the TMS320 fixed point DSP assembly language tools.

The software was written in assembly but modularity was a concern. Three basic modules were built:



- ❑ Speech detector
- ❑ Coarse time-delay estimator
- ❑ Centered adaptive filter

The program modules were optimized to take advantage of the new facilities offered by the TMS320C50 DSP, from which we highlight: zero overhead loops, circular addressing, and the internal Single Access Random Access Memory (SARAM).

Table 1 lists the computational requirements for each module, where **Nactive** is the number of active coefficients in the adaptive centered filter; **CorrL** is the length of the correlation window for the coarse time-delay estimator; **Ntaps** is the delay line length corresponding to memory shared between the filter and the estimator where the reference signal is stored. The considered requirements in CPU clock cycles are accounted for in an environment where the RAM does not need wait states.

Table 1. System Requirements

Module Function	Code Length (Words)	Internal Variables (Words)	Processor Clock Cycles
Speech Detector	85	9	65
Coarse Time-delay Estimator	115	$9+2*Ntaps$	$82+18*CorrL$
Centered Adaptive Filter	124	$14+Nactive$	$114+6*Nactive$
Shared Memory	—	Ntaps	—

As a result of modularity, the system modules can be interconnected to achieve several configurations, such as unidirectional or bi-directional echo cancellation. Additionally, they can be cascaded to support a larger number of active regions in the echo path impulse response.

The supported echo delay can be very significant for a system using a single TMS320C50. Considering the unidirectional configuration, an active region of 4ms, a sample rate of 8KHz, and a 50ns instruction cycle, the maximum supported echo path length is:

- ❑ 381ms for the 9Kword CPU SARAM
- ❑ 2.643 s for 64Kwords of data RAM

Performance Evaluation

An echo cancellation system must exhibit strong attenuation of the returned signal (echo), fast convergence speed after the communication establishment, and little divergence from the solution in the absence of signal or in a double talking situation.

The CCITT G 165 recommendation establishes the tests and goals for echo cancellation products. The CCITT recommendation is achieved and exceeded as shown in Table 2 and Figure 13 for a simulated echo path with ERL of 10dB.

The leak rate in Table 2 refers to ERL degradation two minutes after signals were removed from the fully converged echo canceller. In our system this value is almost null, as the speech detector freezes all the system adjustments in this situation.

Table 2. *Echo Canceller Performance*

Test Description	CCITT G 165 Performance Requirement	System Performance
1. Final echo return loss after convergence, single talk mode	-40dBm0	-51 dBm0
2. Convergence rate; single talk mode	≥ 27 dB	41dB
3. Leak rate	≤ 10 dB	$\cong 0$ dB[MAN29]
4. Infinite return loss convergence	-40dBm0	-48dBm0

The final ERL should be at least 27dB after 500ms to meet the CCITT recommendation. Our system achieves 51dB (41dB ERLE+10dB ERL) only 80ms, as shown in Figure 13.

Figure 13. Echo Celler ERLE in Simulated and Real Echo-Paths

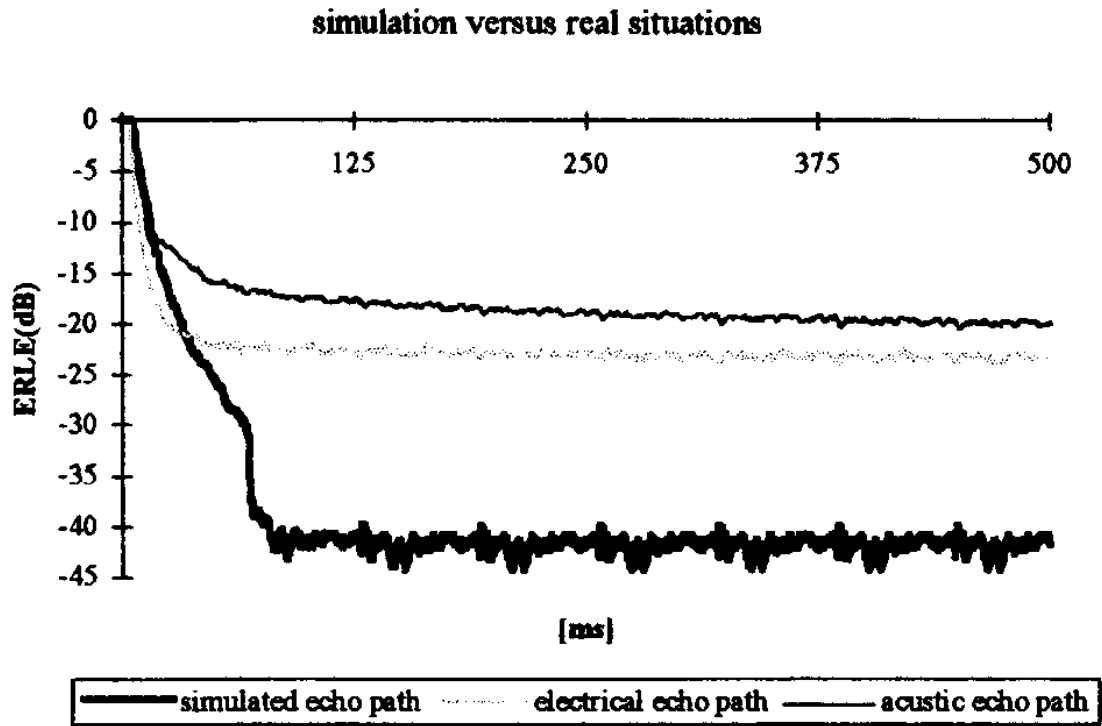
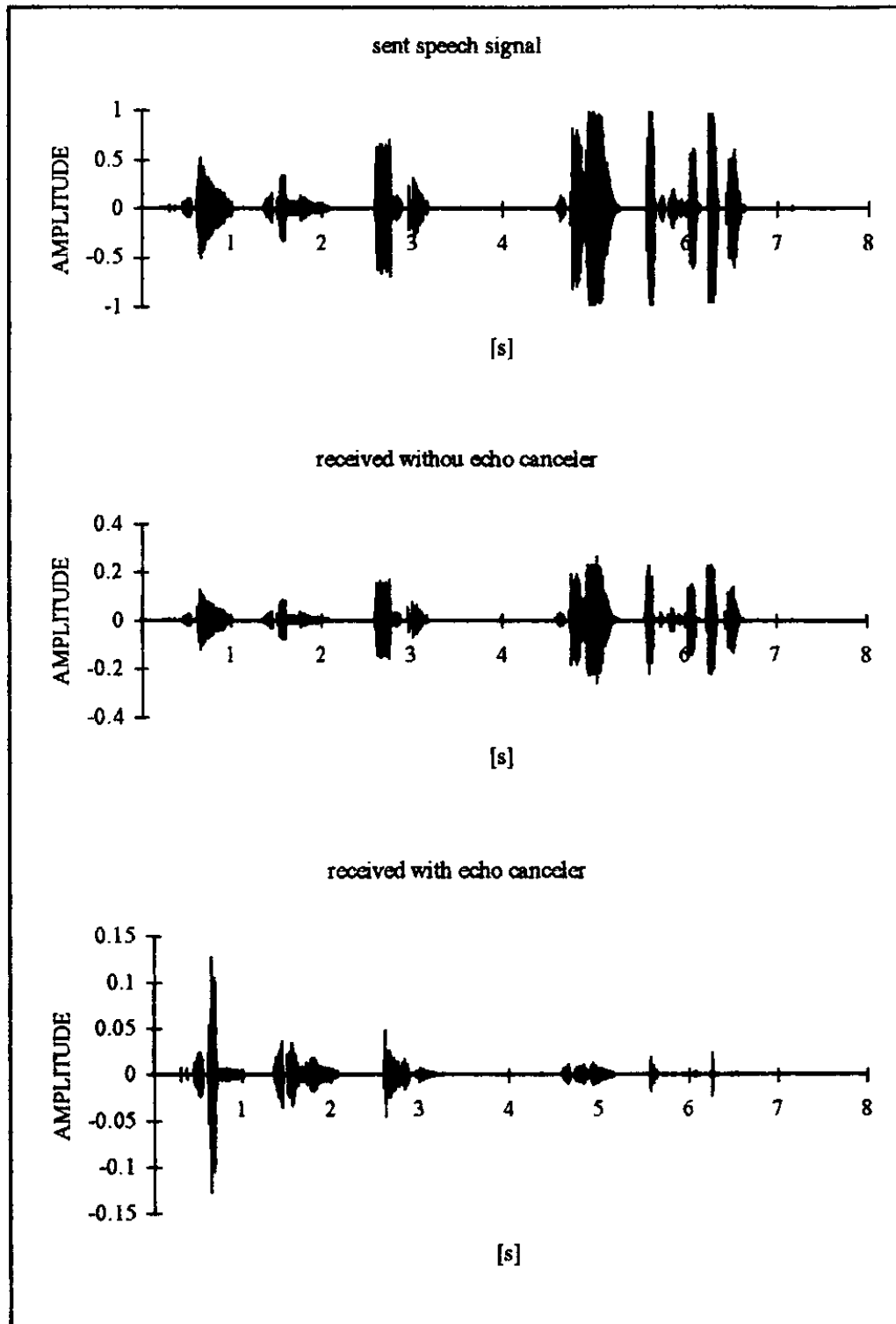


Figure 14 shows the system working with real speech signals. The achieved attenuation is appreciable after the first word return.

Figure 14. Echo Canceller Working with Real Speech Signals





Summary

In this work we present an echo canceller system that exploits the knowledge that the impulse response of an echo path has a set of active regions, which is only a small fraction of the total length.

The system is based on the Texas Instruments TMS320C50 DSP. It uses coarse time delay estimation to track the echo path impulse response active regions positions, where short length adaptive filters are centered.

The echo canceller exceeds the CCITT G 165 recommendation and enables the support of long echo paths, with low computational requirements. Additionally it achieves faster convergence speed and greater ERLE, when compared with other systems reported in recent literature.⁷

Acknowledgements

The authors wish to thank Professor José Leitão for his cooperation, and to Centro de Calculo do ISEL for the research support.

References

- ¹ V. Margo, D. Etter, "Multiple Short-Length Adaptive Filters for Time-Varying Echo Cancellation", *Proc. of ICASSP*, Vol.1, pp.161-164, 1993.
- ² S. Haykin, *Adaptive Filter Theory*, 2nd Edition, Prentice-Hall, 1991.
- ³ G. Jacovitti, G. Scarano, "Discrete Time Techniques for Time Delay Estimation", *IEEE Transactions on Signal Processing*, Vol. 41, No. 2, pp. 525-533, February 1993.
- ⁴ D. Messerschmitt, D. Hedberg, C. Cole, A. Houi, P. Winship, "Digital Voice Echo Canceller with a TMS32020", *Digital Signal Processing Applications*, Vol.1, Texas Instruments, pp.415-454, 1986.
- ⁵ C. Carlemalm, F. Gustafsson, B. Wahlberg, "On the Problem of Detection and Discrimination of Double Talk and Change in the Echo Path", *Proc. of ICASSP*, Vol.5, pp.1295-1298, 1996.
- ⁶ C. Johnson Jr., "Yet Still More On the Interaction of Adaptive Filtering, Identification, and Control", *IEEE Signal Processing Magazine*, pp.22-37, March 1995.
- ⁷ P. Marques, *Long Distance Echo Cancellation Using Centered Short Length Transversal Filters*, Master Thesis (preliminary version in Portuguese), Instituto Superior Tecnico, 1996.