ALGORITHM OF A SINGLE CHIP ACOUSTIC ECHO CANCELLER USING CASCADED CROSS SPECTRAL ESTIMATION

Marco Liem Hyong-Gook Kim O. Manck

Institute of Computer Engineering and Microelectronics / Communication Systems Group
Berlin University of Technology
E-mail: marco@liem.de kim@nue.tu-berlin.de

ABSTRACT

This paper details the algorithm used by a low cost, single chip acoustic echo canceller. The algorithm is based on classical cross spectral estimation. It is employed in a cascaded filter structure where a short, fast filter operates on the output of a longer but slower filter to optimize the tracking performance of changes in the echo path without affecting the steady state performance. This combination allows the use of a fixed configuration for a wide range of acoustic environments.

This contrasts to the predominately employed LMS type of algorithms which are much more sensitive to noise and often require an extensive parameterization specific to the operating environment.

1. INTRODUCTION

Single chip acoustic echo cancellers (AEC) are often employed in mobile communications (e.g. in hands free car kits).

There considerable background noise **n(t)** might be present, which does not only impair the adaptation of the AEC, but also makes the distinction between single and double talk difficult or might sometimes even blur it. Furthermore, even the operating environment itself (and its basic parameters like the expected level of the local speech **s(t)**) is generally not known in advance. Therefore a single chip AEC should be insensitive to noise and should not require an operating environment specific configuration.

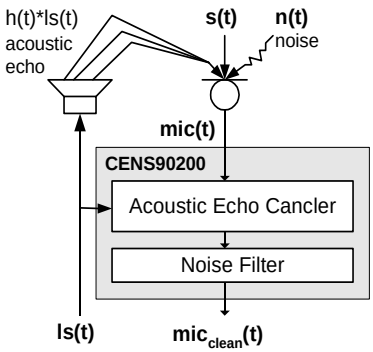

Fig. 1 Operating environment of the acoustic echo canceller chip CENS90200

2. ALGORITHM

Classical cross spectral estimation [1, 2] is employed to provide frequency domain estimates of the unknown echo path **h(t)**. The periodograms $MIC(f)$ and $LS(f)$ of the microphone and loudspeaker signals are used to recursively average the cross and auto spectra $\overline{CS}_{mic,ls}(f)$ and $\overline{CS}_{ls,ls}(f)$, respectively.

$$\begin{aligned}\overline{CS}_{mic,ls}(f) &= \overline{MIC(f)LS^*(f)} \\ &= \overline{H(f)LS(f)LS^*(f)} + \overline{(S(f)+N(f))LS^*(f)} \\ \overline{CS}_{ls,ls}(f) &= \overline{LS(f)LS^*(f)} \end{aligned} \quad (1)$$

$$H(f) = \frac{\overline{CS}_{mic,ls}(f)}{\overline{CS}_{ls,ls}(f)}$$

This cross spectral estimation greatly reduces the error induced by the background noise **n(t)** and the local speech **s(t)**, as both signals are uncorrelated to the loudspeaker signal **ls(t)** and therefore the term $\overline{(S(f)+N(f))LS^*(f)}$ can be averaged out. The forgetting factor used in the recursive averaging determines the tradeoff between adaptation speed and noise immunity.

The bias of the cross spectral estimate $\overline{CS}_{mic,ls}(f)$ is reduced by delaying the loudspeaker with respect to the microphone signal, so that the peak of the time domain cross correlation function of both signals is at zero lag [1 sec.9.3.3.]. This minimizes the effect of non causal short term correlations between **mic(t)** and **ls(t)**.

The classical cross spectral estimation technique described above is employed twice in a cascaded filter structure:

The first stage uses 2048 samples long segments of the incoming **ls(t)** and **mic(t)** signals to compute a first echo path estimate $H0_{1025}(f)$. The long length of this estimate makes it precise by preventing errors caused by under-modeling of **h(t)** and reducing wrap-over effects in the computation of the cross spectra. However the long length also limits the adaptation speed of the first stage as its high computational complexity only allows for one update every 1024 samples. Higher update rates wouldn't improve the tracking performance of echo path changes much, as a new echo path has to dominate an entire segment before it even start influencing the averaged cross spectra.

The echo path estimate of the first stage is subtracted from the microphone signal and the resulting residual echo is fed to the cascaded second stage.

The cascaded second stage uses the residual echo from the first stage and the loudspeaker signal to estimate and subtract the remaining echo. It uses short segments of only 256 samples with 50 percent overlap, and thus has a very high adaptation speed. The length of the resulting echo path estimate $H1_{129}(f)$ cannot sufficiently cover an entire acoustic echo path. Rather the idea is that most acoustic echo paths concentrate much of their energy in a very short section and have this section centered in the cascaded second stage with the delay of the loudspeaker signal described above.

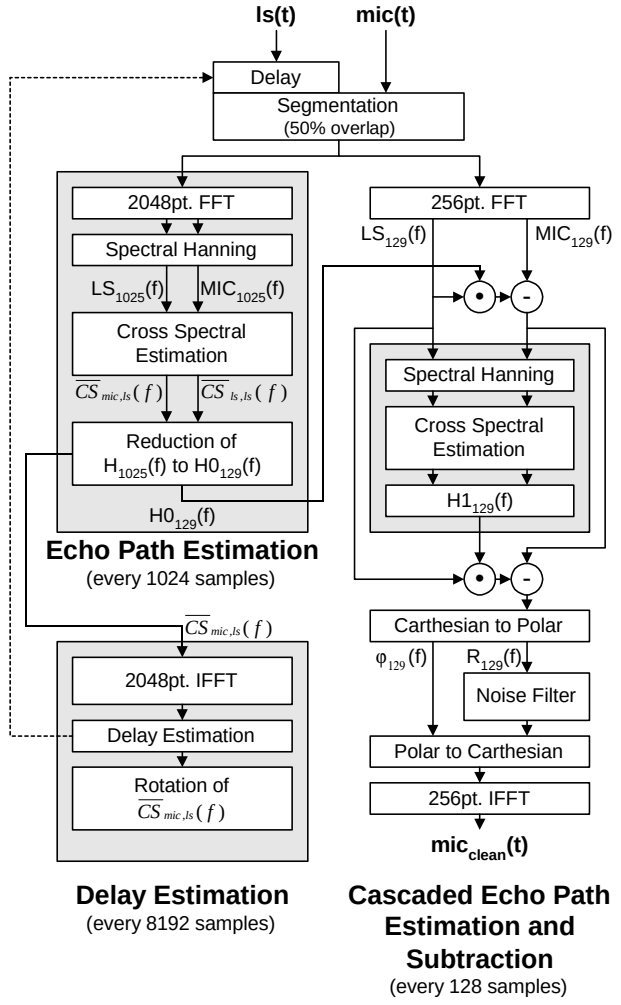

Fig. 2 Overview of the AEC algorithm

This cascaded filter structure has the following characteristics:

In the steady state the long and slow first filter converges and therefore the performance is determined by its length and forgetting factor. In this case the cascaded second filter does not contribute much to the total echo cancellation.

However, during an echo path change the first filter becomes misadapted and reconverges only slowly. Now the cascaded second filter adapts almost immediately to the main portion of the echo let through (or even created) by the misadapted first filter. It steadily tracks the (declining) residual echo from the first filter until it has adapted again.

This cascaded structure allows for a higher tracking quality at a given effective forgetting factor.

3. IMPLEMENTATION

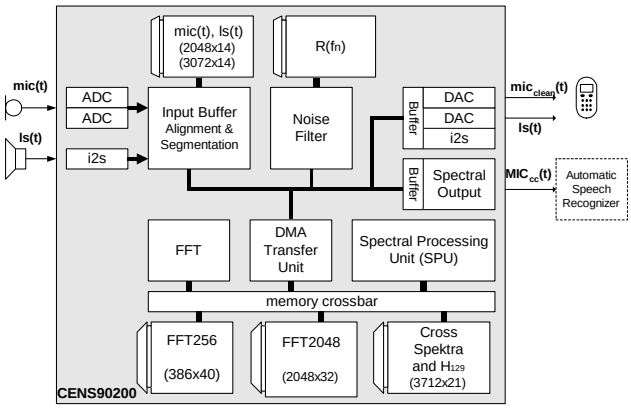

Fig. 3 Architecture of the CENS90200

The chip CENS90200 includes all elements necessary for a stand alone execution of the echo and noise filtering algorithms.

Incoming microphone and loudspeaker signals are either sampled by two integrated 14 Bit G.711 compliant voice codecs or digitally supplied by an I2S interface. The input buffer performs the segmentation of the sample stream into overlapping frames and delays the loudspeaker signal up to 1024 samples with respect to the microphone signal.

The signal processing itself is distributed between dedicated hardware units and a spectral processing unit (SPU). The FFT and noise filter have both been directly mapped to hardware as their processing time directly affects the delay of the microphone signal. The remaining portions of the algorithm are executed in the spectral processing unit. This SPU is a RISC style application specific processor optimized for spectral signal processing. It can execute an arithmetic operation, a division, a memory transfer and a zero cycle conditional jump in parallel.

Finally the computed microphone and loudspeaker signals are transmitted as well in analogue as well as digital form.

This architecture executes the AEC/NF algorithms very efficiently; maximal 1450 clock cycles are required for every incoming sample. This corresponds to a minimal operating frequency of 11.6 MHz at a sample rate of 8 kHz.

Furthermore, the architecture of the CENS90200 has been optimized for maximum portability to allow an easy integration of the filtering algorithms into other designs as an "Intellectual Property" (IP) core. Therefore *every* memory access is given two clock cycles to complete, effectively halving the available memory throughput, but ensuring that any type of integrated memory can be used. The power consumption and operating frequencies compare favorably with other AEC/NF chips [4,5,6]:

Name	**CENS 90200**	**CS 6422**	**MSM 7731-02**	**PSB 2170**
fclk [Mhz]	12.29 / 16.93	20.48	19.2	≈17 / 34,56
i [mA @V]	34 / 43 @3.3	60@5	35 @3.0	≈30/50@3.3
fs [kHz]	8 / 11.025	8	8	8
max AEC length [ms]	128 + 128 / 92.8 + 92.8	63.5	59	50-129 / 96

Table 1 Technical characteristics of integrated AEC/NF chips

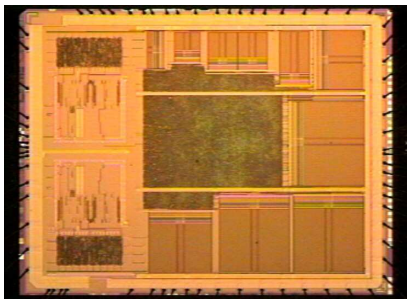

Fig. 4 Die photo of the CENS90200

The dies size of the CENS90200 chip is 36mm^2 in a 0.35µm three layer metal CMOS process.

4. MEASUREMENTS AND RESULTS

The pass through delay of the microphone signal is 245 sample periods, of those only 53 are caused by computations, the remaining 192 are a direct consequence of the overlap and save scheme.

A recording made in an automotive environment achieved an average acoustic echo cancellation performance of around 15db, using only speech as an excitation and without any form of attenuating of **s(t)** or **n(t)**. The impulse response as measured from the same recording with a 16k point FFT is shown in Figure 4. It clearly exceeds the maximum modeling length of the applied echo path estimations $H0_{129}(f)$ and $H1_{129}(f)$.

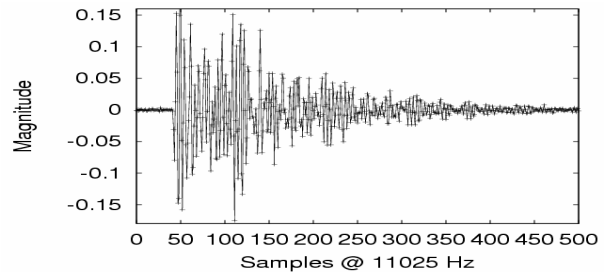

Fig. 5 Measured impulse response

5. REFERENCES

[1] G. Jenkins, D. Watts: "Spectral Analysis and its applications", Holden-Day, San Francisco, 1968

[2] G. Clifford Carter: "Coherence and Time Delay Estimation", Proceedings of the IEEE, Vol.75. No.2, February 1987

[3] Dietmar Ruwisch: "Verfahren und Vorrichtung zur Elimination von Lautsprechersignalen aus Mikrophonsignalen", Patent De 100 43 064 A1, 2000

[4] Cirrus Logic: "CS6422 Enhanced Full-Duplex Speakerphone IC" (Datasheet), July 2001

[5] OKI Semiconductor: "MSM7731-02 Voice Signal Processor" (Datasheet), May 2001

[6] Infinion Technologies: "Acoustic Echo Canceller ACE PSB2170 Version 2.1" (Datasheet), 1999