VLSI Implementation of a LMS Based Adaptive Noise Canceller

F. Gittel 1,2, T. D. Smith 1, A. Th. Schwarzbacher 1 and E. Hilt 2

1Dublin Institute of Technology, School of Electronic and Communications Engineering, Dublin 6, Ireland
2Deutsche Telekom University of Applied Sciences, Leipzig, Germany

Abstract: To better integrate disabled persons is a continuous aim in a modern society. For handicapped people, robots are used to support the personal freedom and provide more convenience. These robots need to be controlled by voice which requires a reliable working speech recognition system. Therefore, algorithms that can improve the quality of speech and thus support the detection of the speech information are highly desirable.

This paper introduces a hardware implemented and optimised Adaptive Noise Canceller (ANC), which can be utilised in speech detection devices to reduce the noise intensity of the speech to be recognised. In addition, it can also be used to improve the speech quality in information transfer systems. The evaluation results show how the circuit is able to reduce the unwanted components within a speech signal and therefore, the system is able to increase the speech quality. Furthermore, any prior knowledge of the surrounding environmental properties is not needed.

Keywords: Noise Cancelling, LMS, VLSI

I. INTRODUCTION

This paper describes the VLSI hardware implementation of an Adaptive Noise Canceller (ANC) which is able to filter an input speech signal to provide noise reduced output speech. To achieve this goal, the digital filter, which is the main section of the ANC, needs to adjust its frequency response continually to the changing conditions of the surrounding environment. Therefore, an update functionality must be introduced. This functionality is based on the Least Mean Square (LMS) algorithm [1]. The method of least mean square adaptive filtering takes advantage of the quasi – periodic nature of the speech signal to form an estimate of the clean speech signal at time t. This estimation is derived from the value of the signal at time t-T which represents the actual time shifted by one estimated pitch period. The principle of this method is shown in Figure 1. To describe this approach, some considerations have to be taken into account. In practice, an a priori knowledge to adjust the filter response is not available. The output of the FIR filter used for this implementation is given by

\[ y(n) = \sum_{i=0}^{L} b_i x(n-i-T) \]  \hspace{1cm} (1)

where \( x \) is the noisy speech signal, \( L \) is the filter order and \( T \) is the analysed pitch period for the speech signal. The \( b_i \) represent the filter coefficients updated sequentially according to the LMS algorithm. The filter provides an estimate of the clean input signal \( y(n) \). One possibility to extract the necessary reference from the input signal is to estimate the additive noise during the silent speech segments when only the noise occurs. The problem is, that noise is rarely stationary and the detection of silence speech parts is not error free. In addition, this method can not be applied for quantisation noise. The difficulty of forming a reference noise signal is solved by extracting a reference signal from the original speech \( x(n) \). Due to the quasi periodic nature of speech, a section of speech delayed by its pitch period \( x(n-T) \) is highly correlated to the original speech \( x(n) \) but uncorrelated to the additive noise \( x_{ad}(n) \). The derivations of [1] describe that by minimising the energy of the estimation error \( e(n) \), the output of the filter, and consequently the system output will be a signal \( y(n) \) that is the best least square fit of the input speech signal \( x(n) \). As can be seen in (2), the error signal is defined as the difference between input signal and estimated filter output.

\[ e(n) = x(n) - y(n) \]  \hspace{1cm} (2)

This error signal is used to update the filter coefficients and thus to adjust the filter response.

\[ b_{n+1} = b_n + 2 \mu \cdot e(n) \cdot X_{n-T} \]  \hspace{1cm} (3)

Each coefficient \( b_{n+1} \) is updated using the corresponding present coefficient and a correction term which is formed by the filter tap values shifted by the pitch period \( X_{n-T} \), the estimation error and the step size \( \mu \). This factor controls stability and rate of convergence. The ANC starts with an arbitrary coefficient vector, the algorithm converges in the mean and will remain stable as long as the parameter \( \mu \) is greater than zero but less than the reciprocal largest eigenvalue \( \lambda_{max} \) of the matrix R [1].
The correct pitch period of the input speech is extracted using the Average Magnitude Difference Function (AMDF) [6]. This function was chosen because no multiplications are performed and thus, lower area and power consumption can be achieved. In addition, a voiced / unvoiced classification was implemented [2] which is based on a short term energy determination, zero crossings count and a min / max ratio calculation of the values within an AMDF frame. This classification functionality is used to bypass the filter until the first estimated pitch period and to keep the filter coefficients constant during unvoiced sections of the speech.

II. IMPLEMENTATION

For the hardware implementation the ANC was split into different modules with separate functionalities. They were implemented using the ES2 ECPD 0.7µm CMOS technology. All designs were written in abstract VHDL and synthesised using the Synopsys Design Compiler without any design constraints. Figure 2 shows a block diagram which describes the structure. The ANC consists of two main sections, the Pitch Detector [2] and the Adaptive Filter [3]. It uses the incoming speech samples to provide the noise reduced output signal. Two different clock frequencies are used, an 8kHz clock to read in the input samples and a clock frequency of 22MHz to perform all the necessary calculations within one slow clock period. The filter order was chosen to be 10 and its coefficients are represented by a 23 bit vector to ensure sufficient accuracy of the filtering process. A converter was implemented to change the default two’s complement number system to a signed magnitude system. This procedure lowers the amount of switching in the Adaptive Filter section and thus it reduces the power consumption.

III. SYSTEM PERFORMANCE & RESULTS

The performance of the Adaptive Noise Canceller is demonstrated and benchmarked using noisy synthetic speech signals which were filtered by the ANC. The synthetically produced input signals that were used are a vowel ‘A’ with a fixed pitch frequency of 125Hz and a real speech phrase “Her wardrobe consists of only skirts and blouses” with variable pitch frequencies as well as voiced / unvoiced sections. The formant frequencies which form the vowel ‘A’ are f1=730Hz, f2=1090Hz, and f3=2440Hz. All signals are distorted by White Gaussian Noise (WGN) and have a signal to noise ratio (SNR) of 5dB to 10dB. The results of those tests are presented in Figures 3 to 12. The magnitude spectra of a noisy and filtered vowel (after detecting the pitch and convergence of the adaptive filter) with a SNR of 10dB are shown in Figures 3 and 4. For reasons of clarity, they have been normalised and only the range from 0 to 0.5 is presented. It can be seen that the filtered version retains the spectral shape of the input signal with the formant frequencies remaining prominent. Hence, the perceptual characteristics of the signal are, in the main, unchanged. Furthermore, the noise component is reduced in the filtered signal, being particularly noticeable in the higher frequencies from 2000Hz to 4000Hz where it is nearly completely reduced. However, in the region 0Hz to 1500Hz, although the adaptive filter manages to reduce the noise, remnants of the noise component and some attenuation of the lower harmonics can be observed.
Figure 3: Enlarged Spectrum of a noisy Vowel ‘A’ with Pitch Freq. 125Hz and SNR=5dB

Figure 4: Enlarged Spectrum of a filtered Vowel ‘A’ with Pitch Freq. 125Hz and SNR=5dB

Figure 5: Spectrogram of a noisy Vowel ‘A’ with Pitch Freq. 125Hz and SNR=5dB

Figure 6: Spectrogram of a filtered Vowel ‘A’ with Pitch Freq. 125Hz and SNR=5dB

Figure 7: Spectrogram of a noisy Vowel ‘A’ with Pitch Freq. 125Hz and SNR=10dB

Figure 8: Spectrogram of a filtered Vowel ‘A’ with Pitch Freq. 125Hz and SNR=10dB

Figure 9: LAR Distance Measure Result, Pitch=125Hz, Input Signal SNR=5dB

Figure 10: LAR Distance Measure Result, Pitch=125Hz, Input Signal SNR=10dB

Figure 11: Spectrogram of noisy Synthetic Real Speech, SNR=10dB

Figure 12: Spectrogram of filtered Synthetic Real Speech, SNR=10dB
Figures 5 to 8 show the performance from the spectrogram perspective of the distorted signal and of the filtered signal respectively. Again, the figures show clearly that the noise component of the higher frequencies 2000Hz to 4000Hz is well reduced but that in the region of 0Hz to 1500Hz some noise remains and the harmonic attenuation persists over time. Thus, it must be concluded that the filter reduces the noise component while not significantly changing the perceptual characteristic of the speech signal. Figure 9 and 10 present the corresponding results of applying the Log Area Ratio (LAR) distance speech quality measure [4] to the distorted and filtered signals. This measure is based on finding a set of Linear Predictive Coefficients (LPC) for each frame of the distorted/filtered speech signals and the original clean speech, transforming them into Log Area Ratio (LAR) coefficients [4] and then calculating the difference between them. This measure was shown to have a correlation coefficient of 0.62 with subjective speech quality assessment data [5]. Figures 9 and 10 demonstrate the speech quality improvement using vowels with SNRs of 5dB and 10dB. It is shown that the LAR distance is shortened by 17% (5dB) and in the case of 10dB SNR the distance is even reduced by 28%. Finally, the performance of the ANC is visualised using a distorted real speech phrase. The spectra of the original and distorted signal are shown in Figures 11 and 12. The spectral shape of the voice information remains after the filtering process and the broad band noise energy is noticeably reduced. Furthermore, it can be seen that the energy of the speech information containing spectral sections of the speech signal are kept after filtering.

IV. CONCLUSIONS

The objective of this paper was to describe an Adaptive Noise Canceller which was successfully developed and implemented in hardware using VLSI design techniques in conjunction with a VHDL development environment. Two components, a Pitch Detector and an Adaptive Filter were incorporated into the ANC with additional hardware optimisation of the structure. It has been shown that the developed structure is able to reduce noise in a distorted speech signal using objective speech measures. Furthermore, the frequency components of the signal, which are the bearers of information, are almost unaffected. Additionally, subjective listening tests have shown that the audibility of a noisy speech signal is significantly improved after processing. An insertion into hearing aids, speech recognition systems that aid the handicapped or mobile telephony devices is unproblematic as the silicon area of the whole system is only 14.5mm² (based on the 0.7µm library) and is therefore suitable for such purposes.

In summary, this paper has presented that an effective filtering performance under real conditions is given by the ANC device. It is able to adapt its behaviour to suit different input signals and environments without the need to provide an additional reference source.

V. REFERENCES