ULTRA LOW-POWER SUB-BAND ACOUSTIC ECHO CANCELLATION FOR WIRELESS HEADSETS

Julie Johnson, Etienne Cornu, Gary Choy and John Wdowiak

Dspfactory Ltd, 611 Kumpf Drive, Unit 200, Waterloo, Ontario, Canada N2V 1K8

ABSTRACT

Integrating voice-processing algorithms into headsets is a challenging task due to size and power constraints. As a result of the small size of these headsets, acoustic leakage between the closely located speaker and microphone is a common problem. To address this problem, an oversampled sub-band adaptive acoustic echo cancellation system is proposed and implemented on an ultra-low resource DSP system. The echo cancellation algorithm is integrated with two voice activity detectors and double-talk detection logic to provide a fully functional system. By using decimation, load balancing and other optimization techniques, the complete system consumes about 6 milliwatts of power. Other signal-processing algorithms can be added with very little impact on power consumption and no impact on size. Tests on a BlueTooth headset show an echo return loss enhancement of about 15 dB for a typical headset signal level.

1. INTRODUCTION

BlueTooth headsets are becoming a popular alternative to handsets and wired headsets due to their small size, portability and convenience. Unfortunately, the small size of these headsets often necessitates that the speaker and the microphone be located close together. This proximity causes considerable acoustic leakage from speaker to microphone, thus creating acoustic echo. Delays inherent in the BlueTooth link exacerbate the problem. Since size and form are important in the headset industry, industrial designers hesitate to solve this problem by altering their designs. Clearly, there is a need for an echo canceller inside these BlueTooth headsets. One possibility is to use a fixed-function device, but since other algorithms such as noise reduction or speech enhancement are usually also required, a flexible, low-power, miniature DSP-based solution seems to be the best approach.

This paper proposes an LMS-based sub-band adaptive echo cancellation system that is implemented on an ultra low-power, miniature DSP that features an oversampled, windowed overlap-add (WOLA) co-processor specifically designed for sub-band processing. This echo cancellation system is suitable for deployment in headsets and leaves sufficient resources for additional sub-band based algorithms.

Algorithms based on sub-band adaptive filtering present a number of challenges due to the limited resources available on the DSP. Whereas Gupta and Hero [1] use quantization techniques to reduce computational load in LMS-based systems, we have chosen a combined approach of spectral whitening by decimation and spectral emphasis filtering [2], which not only reduces the number of filter coefficients, but also has been shown to improve the rate of LMS convergence of an oversampled sub-band adaptive filter (OS-SAF) significantly. Load balancing and other optimization techniques have also been used in order to further reduce the computational requirements. Two voice activity detectors [3] and double-talk detection logic have been integrated with the optimized LMS-based adaptive sub-band filtering algorithm in order to demonstrate a fully functional system.

In this paper, a description of the DSP system architecture is first presented, followed by an introduction to the LMS-based sub-band adaptive filtering algorithm. In Section 4, we describe the effects of precision on the behaviour of the algorithm and we present the system architecture that is introduced to facilitate effective load balancing. Section 5 describes the algorithm evaluation. Finally, conclusions are presented in Section 6.

2. DSP SYSTEM

The DSP system [4] consists of three major components: a weighted overlap-add (WOLA) filterbank coprocessor, a 16-bit fixed point DSP core, and an input-output processor (IOP). These three components run in parallel and communicate through shared memory. The parallel operation of these components allows for the implementation of complex signal processing algorithms with low system clock rates and low resource usage. The system is particularly efficient for sub-band processing. The configurable WOLA coprocessor efficiently splits the fullband input signals into sub-bands, leaving the core free to perform the other algorithm calculations, such as filtering and coefficient adaptation, as described in Section 4.

The echo cancellation algorithm is implemented on the DSP system using a 16-band oddly stacked, 8-times oversampled WOLA filterbank configuration. The selected configuration generates a group delay of 9.5 miliseconds. The system operates on 1.8 Volts with a power consumption of about 6 milliwatts.

3. ALGORITHM DESCRIPTION

The echo cancellation algorithm is deployed on the DSP system described in Section 2. The block diagram for the echo cancellation system including the two voice activity detectors (VAD) for far-end (FE) and near-end (NE) signals, and double-talk detection logic is shown in Figure 1.
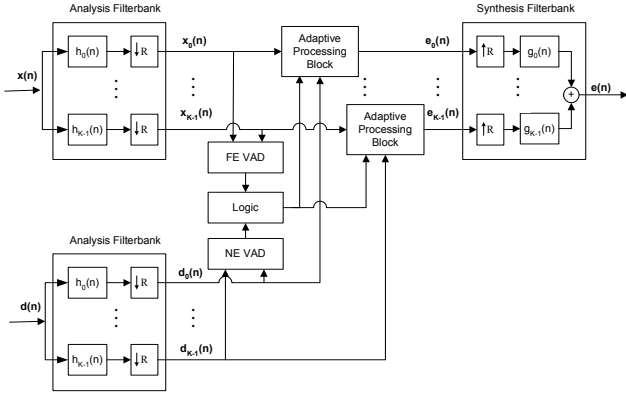

Figure 1: Block diagram of the echo cancellation system.

The DSP is configured with stereo inputs and a mono output. Analysis of the near-end signal $d(n)$ and far-end signal $x(n)$ occurs in parallel on the WOLA filterbank. One input block is analyzed every R samples. The analysis results are used by the two voice activity detectors to determine whether adaptation is to occur. Coefficient adaptation then occurs on each band independently within the adaptive processing blocks shown in Figure 1. Once the filters have been updated, the near-end signal is filtered and synthesized.

The oversampled sub-band adaptive filter echo cancellation algorithm employed in this system uses whitening by both decimation and spectral emphasis (WBDS) as described in [2]. Spectral emphasis whitening improves the convergence rate of the OS-SAF algorithm considerably. A filter, f_{emp}, which amplifies the high three quarters of the spectrum in each sub-band, is employed [2]. Whitening by decimation achieves a data rate closer to critical sampling and provides a whiter signal at the adaptive filter input. The decimation factor D was set to 4 (oversampling of 8) in this implementation. Combined, these steps improve the LMS convergence rate significantly. The use of decimation decreases computation costs by a factor related to D.

A block diagram of the adaptive processing block for a single band used in WBDS is shown in Figure 2. The input signals x_k and d_k are decimated by a factor of D and filtered by f_{emp} to give the whitened signals x'_k and d'_k, respectively.

A side branch is employed in the adaptive system to enable the use of whitened inputs, x'_k and d'_k, for the LMS adaptation (side branch) while conserving the original inputs, x_k and d_k, for the adaptive filtering in the main branch. In Figure 2, the decimated and whitened signals x'_k and d'_k in the side branch are used to calculate the decimated error signal, e'_k. These signals

are, in turn, used to update the side branch adaptive filter, $\mathbf{w}'_k$. When filtering is to occur in the main branch, $\mathbf{w}'_k$ is up-sampled by a factor of D by inserting zeros and copied to the main branch filter, $\mathbf{w}_k$.
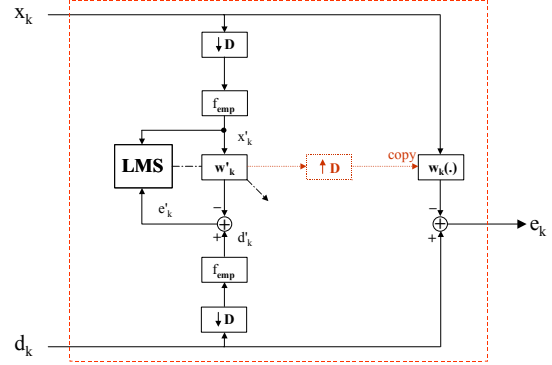

Figure 2: Adaptive processing block used in the WBDS method.

The use of the side branch prevents excessive aliasing [2]. The order of the side branch adaptive filter, $\mathbf{w}'_k$ is set to M/D, where M is the order of the main branch filter. In this implementation M = 32, leading $\mathbf{w}'_k$ to be of order 8.

The relationship between the main branch input signal and the decimated side branch signal, and the corresponding relationship between side branch and main branch filter coefficients may be represented as follows:

$$\begin{aligned}
&(x_k(i)\, x_k(i+1)\ldots) \rightarrow \downarrow D \rightarrow (x'_k(i)\, x'_k(i+1)\ldots) \\
&(w'_k(0)\, w'_k(1)\ldots) \rightarrow \uparrow D \rightarrow (w_k(0)\, 0\, 0\, 0\ldots w_k(D)\, 0\, 0\, 0\ldots w_k(2D)\ldots)
\end{aligned} \tag{1}$$

where

$$\begin{aligned}
&x'_k(i)=x_k(i), \quad x'_k(i+1)=x_k(i+D), \quad x'_k(i+2)=x_k(i+2D),\ldots \\
&w_k(0)=w'_k(0), \quad w_k(D)=w'_k(1), \quad w_k(2D)=w'_k(2),\ldots
\end{aligned} \tag{2}$$

An examination of the system components highlights the problems to be anticipated when porting the algorithm to the DSP. The system uses an analysis window of 128 samples and a synthesis window of 16 samples; an input block is analyzed every R=4 samples. The sampling frequency F_s is 8 kHz. With this configuration, R/F_s=0.5 ms are available during each block to apply the filter, adapt the coefficients, update the VAD flags and perform any other algorithm functions that are required. Thus, effective load balancing is a design issue.

The precision of the filter coefficients is also an area of concern. The data used in the DSP core is represented as 16 or 32-bit numbers. All 16-bit arithmetic operations are executed in a single CPU cycle. Thus, a 16-bit representation for all algorithm data is optimal. We will need to determine the optimal precision for the coefficients that is within the given time and memory constraints.

Another issue is the division operation that is part of the coefficient update calculation. The standard division operation on this DSP core takes 15 cycles, which is relatively expensive, considering that this operation occurs 128 (16 bands x 8 coefficients) times per input block. Alternatives to this division needed to be considered.

4. ALGORITHM IMPLEMENTATION

This section describes an implementation of the echo cancellation algorithm on the DSP that effectively addresses the concerns outlined in Section 3.

4.1. DSP Architecture

As stated previously, the DSP system was specifically designed to support algorithms that operate on sub-bands. In addition to the analog input and output stages, the DSP includes three processing units that operate in parallel: (1) The I/O Processor (IOP) provides an interface to the input and output stages through stereo FIFOs. It groups samples into blocks and interrupts the main processor at every block. (2) The WOLA coprocessor performs analysis and synthesis. For analysis, it reads samples from the input FIFO and writes the analysis results to shared memory; for synthesis, it reads the frequency-domain data from shared memory and write the time-domain samples to the output FIFO. (3) The main processor performs the rest of the operations: filtering in each sub-band, coefficient adaptation and voice activity detection and logic.

4.2. Load Balancing

Adapting and filtering all 16 bands and running the VAD during each block requires a clock speed that would increase power consumption to an undesired level. Therefore, a load-balancing scheme is required in order to fit the algorithms into the available cycles. For the echo cancellation algorithm, the 16 sub-band adaptive filters are divided into groups of four. During each block, four of these filters are updated in the side branch. These updated filters are stored in temporary memory until all of the filters are updated, at which time they are moved to the main branch. All filters are calculated using data from the same block of input data. Thus, a decimation factor of four is introduced by this load-balancing arrangement.

A similar architecture is required for the VAD. The near-end and far-end VADs are each spread over 10 blocks. Therefore, 20 blocks (equivalent to 10 milliseconds) are required for a single adaptation decision. A block diagram showing the echo cancellation system architecture is shown in Figure 3.

4.3. Algorithm Precision

In order to examine the sensitivity of the filter coefficients to precision, a number of tests were performed where the precision of the coefficients was varied. Figure 4 summarizes the effects of adaptive filter coefficient precision on algorithm performance when processing white noise.
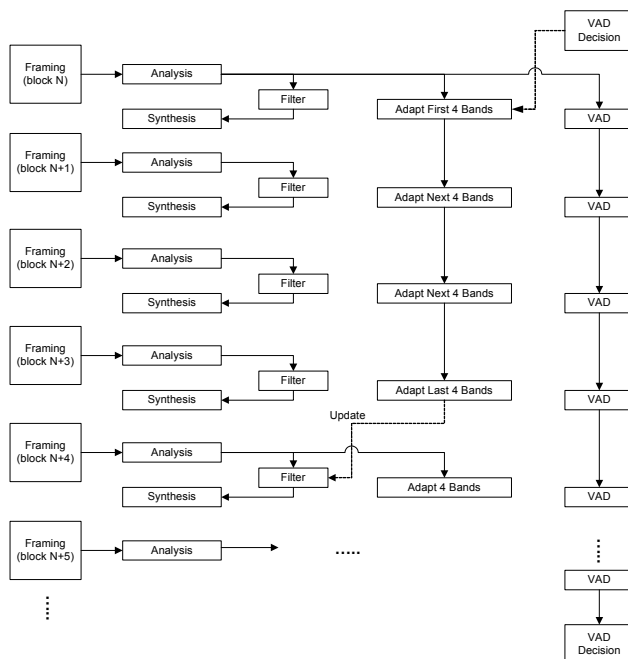

Figure 3: Real-time system architecture

The DSP system used in this implementation has a 40-bit multiply-accumulate instruction. As is evident from the figure, when coefficients are stored in 16 bits, the optimal solution is to calculate the coefficients in 40-bit precision, to perform a round operation on the top 16 bits and then to store the top 16 bits.

Double precision coefficients were contemplated, but not implemented in this version of the algorithm due to the increased time and memory requirements. The trade-off between the performance improvement and the increased resource requirements was not deemed to be worthwhile based on the strong performance provided by the 16-bit representation.
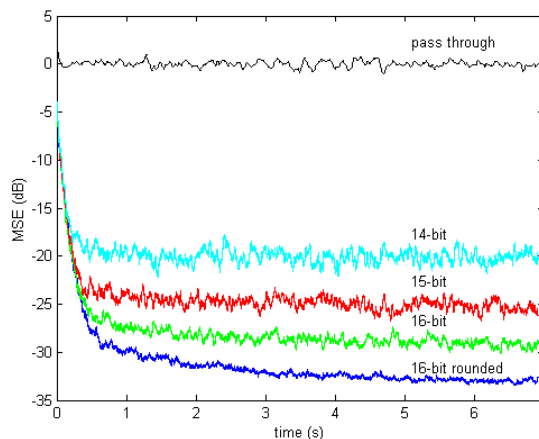

Figure 4: Effects of adaptive filter coefficient precision on performance.

4.4. An Approximation to the Division

The algorithm requires a single division each time a filter coefficient is updated. The problematic division occurs within the adaptive coefficient update calculation:

$$\mathbf{w'_{n+1}} = \mathbf{w'_n} + \mu \frac{e'_n \mathbf{x'_n}}{\mathbf{x'_n}^{H} \mathbf{x'_n}} \tag{3}$$

where μ is the step size, $e'_n = d'_n - \mathbf{w'_n}^{T}\mathbf{x'_n}$, as calculated in the adaptive processing block at time n+1 and H is the conjugate transpose.

In order to reduce computational requirements, a 2-cycle approximation to this division was found. First, we calculate the position of the first non-zero bit in the denominator using the 'calculate leading bit' instruction (CLB), which takes one cycle. Next, we perform the division by shifting the numerator to the left by the value returned by the CLB instruction. This shift is also a single cycle instruction. The maximum error from this approximation is 50%. However, this error was found to be an acceptable trade-off for the savings in cycles. It may be viewed as a reduction in the value of μ. Note that additional steps of the same type can be chained if more precision is needed at this stage.

5. SYSTEM EVALUATION

In order to evaluate the algorithm in a real-world environment, the DSP is integrated into a headset. For this application, the echo cancellation algorithm uses a 32-tap sub-band adaptive filter for each band. This allows the algorithm to handle echo paths up to 14 milliseconds. This filter is long enough to handle the local echo and the delay introduced by the WOLA filterbank. Using of a decimation factor of 4, only 8 filter coefficients in each band are non-zero.

For the tests, a PC is used to apply input signals to the headset and to record output signals from the headset. The headset is affixed to the ear of a head-and-torso (HAT) simulator (Brüel & Kjær Type 4128C). The far-end signal is applied directly to the speaker of the headset. No near-end signal is applied. The far-end signal and the echo captured by the headset's microphone are the input signals to the DSP. The output of the DSP is sent back to the PC for recording and measurement.

Figure 5 shows the results of applying white noise of varying sound pressure levels (SPL) as the far-end input to the system. In this figure, subtracting (in dB) the pass through signal from the input signal gives the echo return loss (ERL) and subtracting (in dB) the echo-cancelled signal from the pass through signal gives the echo return loss enhancement (ERLE). For headset speaker signals that are less than approximately 83 dB SPL, the echo-cancelled signal remains relatively constant as the echo is attenuated to the level of the noise floor, where it is not distinguishable from the noise. For input speaker levels greater than 83 dB SPL, the level of the echo-cancelled signal increases along with the level of the input signal. The maximum ERLE is reached at a speaker level of about 95 dB SPL and remains more

or less constant as the level rises. When the speaker level is at 80-85 dB SPL at the ear, which is a comfortable listening level, the ERLE is approximately 15 dB and the combined loss (ERL+ERLE) is about 35 dB. As the graph shows, the echo cancellation system reduces the echo to the level of the noise floor.
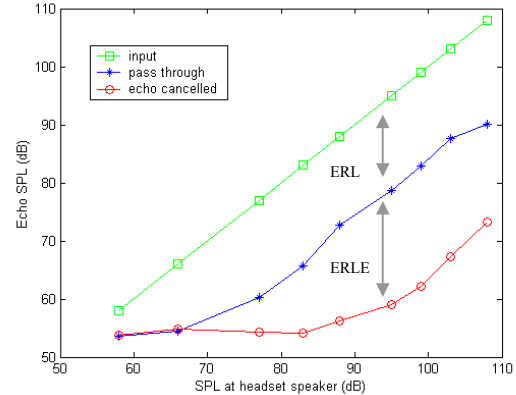

Figure 5: Echo levels before and after echo cancellation

6. CONCLUSIONS

An oversampled sub-band adaptive acoustic echo cancellation algorithm is successfully implemented on an ultra-low power DSP system. This echo cancellation algorithm is integrated with two voice activity detectors and double-talk detection logic to provide a full system solution, with resources still available for the integration of additional algorithms. This system is incorporated into a headset and evaluated. The combined loss for headset speaker levels greater than 83 dB SPL is approximately 35 dB. Thus, this system effectively solves the problem of incorporating an echo canceller into a headset.

Future work will involve integrating other algorithms, such as noise reduction and speech enhancement into the existing architecture. Double precision filter coefficients may also be implemented to further improve algorithm performance. We will also investigate ways of reducing memory and computational requirements in order to accommodate a longer filter.

7. REFERENCES

[1] R. Gupta and A. Hero III, "Theoretical Aspects of Power Reduction for Adaptive Filters", *Proc. ICASSP 1999.*

[2] H. R. Abutalebi, H. Sheikhzadeh, R. L. Brennan and G. H. Freeman, "Convergence Improvement for Oversampled Subband Adaptive Noise and Echo Cancellation," *Proc. Eurospeech 2003.*

[3] E. Cornu, H. Sheikhzadeh, R. L. Brennan, H. R. Abutalebi, E. C. Y. Tam, P. Iles and K. W. Wong, "ETSI AMR-2 VAD: Evaluation and Ultra Low-Resource Implementation, *Proc. ICASSP 2003.*

[4] R. Brennan and T. Schneider, "A Flexible Filterbank Structure for Extensive Signal Manipulations in Digital

Hearing Aids", *Proc. IEEE Int. Symp. Circuits and Systems,* pp. 569-572, 1998.