Real-Time Implementation of Acoustic Echo Canceller Using TMS320C6711 DSK

Wonchul Heo* · Keunsung Bae**

ABSTRACT

The interior of an automobile is a very noisy environment with both stationary cruising noise and the reverberated music or speech coming out from the audio system. For robust speech recognition in a car environment, it is necessary to extract a driver's voice command well by removing those background noises. Since we can handle the music and speech signals from an audio system in a car, the reverberated music and speech sounds can be removed using an acoustic echo canceller. In this paper, we implement an acoustic echo canceller with robust double-talk detection algorithm using TMS-320C6711 DSK. First we developed the echo canceller on the PC for verifying the performance of echo cancellation, then implemented it on the TMS320C6711 DSK. For processing of one speech sample with 8kHz sampling rate and 256 filter taps of the echo canceller, the implemented system used only 0.035ms and achieved the ERLE of 20.73dB.

Keywords: acoustic echo canceller, TMS320C6711

I. Introduction

Automatic speech recognition in car environments is a good interface for a driver to input voice command to the telematics system. But the recognition performance becomes very poor when there exists reverberated music or speech coming out from the audio system. For robust speech recognition in a car environment, it is necessary to extract a driver's voice command well by removing those background noises. Since we can handle the music and speech signals from the in-vehicle audio system, the reverberated music and speech sounds can be removed using an acoustic echo canceller (AEC).

In this paper, we implement an AEC running in real-time with robust double-talk (DT) detection algorithm using TMS320C6711 DSK. It uses an FIR-type adaptive filter with the normalized least mean square (NLMS) algorithm [1]. For robust DT detection, an auxiliary filter is employed [2]. We first developed the echo canceller on the PC for verifying the performance of echo cancellation, then implemented it on the TMS320C6711 DSK. For one speech sample with 8kHz sampling rate and 256 filter taps of the echo canceller, the implemented system needed the process time of only 0.035ms and achieved the ERLE of 20.73dB.

The rest of this paper is organized as follows. In section 2, we give a brief description about the AEC and DT detection algorithms used in the implemented system. In section 3, experimental result for real-time implementation on the TMS320C6711 DSK is explained. Finally, conclusion is given in section 4.

* LG Electronics Inc., Pyungtaek
** School of Electrical Engineering and Computer Science, Kyungpook National University

2. AEC with Robust Detection

<Figure 1> shows the block diagram of the implemented AEC with an auxiliary filter. Far-end signal, $x(n)$, corresponds to a music or speech signal from the in-vehicle audio system. The output signal of the audio system returns to a microphone via room impulse response. Such a signal, what we call an echo, $y(n)$, is added to a driver's voice command, $s(n)$, and constructs a microphone input, $d(n)$. Since the reverberated audio output acts as a noise signal for speech recognition, the role of the AEC is to remove it from the microphone input, and extracts the driver's speech signal. This is accomplished with an adaptive filter which estimates an echo replica, $\hat{y}(n)$, and subtracts it from the microphone input. An FIR-type adaptive filter that works with an NLMS algorithm is used because of its simplicity and good performance [3].

In the AEC, the adaptation should be halted during DT periods in which both near-end and far-end signals are present at the same time. Especially, in the AEC as a front-end for speech recognition, the DT detection is very important since it indicates the region where the driver's voice command exists. In our system, the improved cross-correlation method is used with an auxiliary adaptive filter for robust DT detection [1,2].

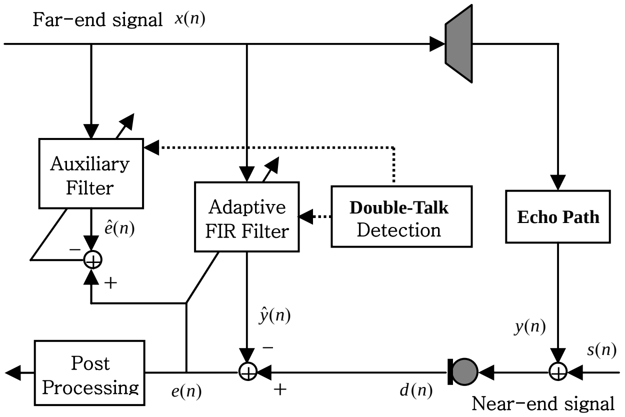


Figure 1. Block Diagram of the Implemented AEC

The DT detection in a very noisy environment like the interior of an automobile is a difficult task. The conventional DT detection algorithm [4] using the cross-correlation between $d(n)$ and $e(n)$ occasionally detects the background noise as a DT in a very noisy environment since the correlation increases when the background noise is dominant as compared with an echo signal. To reduce this type of error, we use the smoothed power, P_{avg}, obtained from the power of the microphone input signal, σ_d^2, as given in eq.(1)-(3).

$$\sigma_d^2(n) = (1-\beta_1)\cdot\sigma_d^2(n-1) + \beta_1 \cdot d^2(n) \tag{1}$$

$$\sigma_{avg}^2(n) = (1-\beta_2)\cdot\sigma_{avg}^2(n-1) + \beta_2 \cdot \sigma_d^2(n) \tag{2}$$

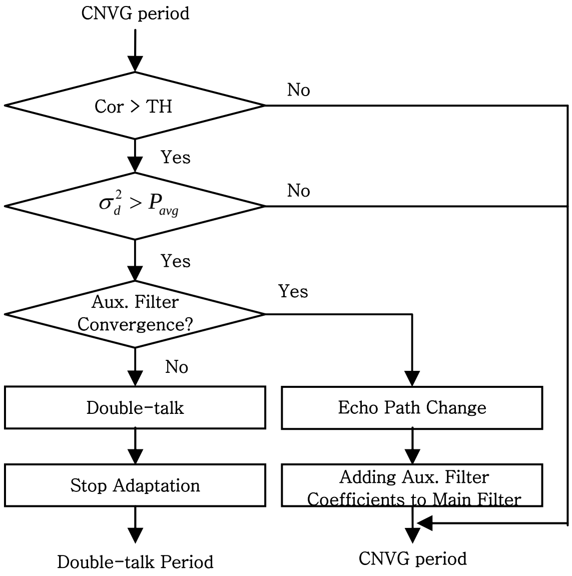
$$P_{avg}(n) = (1-\beta_3)\cdot P_{avg}(n-1) + \beta_3 \cdot \sigma_{avg}^2(n) \tag{3}$$

where β_1, β_2 and β_3 are set to 1/256. It is shown in [1] that σ_d^2 is generally larger than P_{avg} during the DT periods and vice versa.

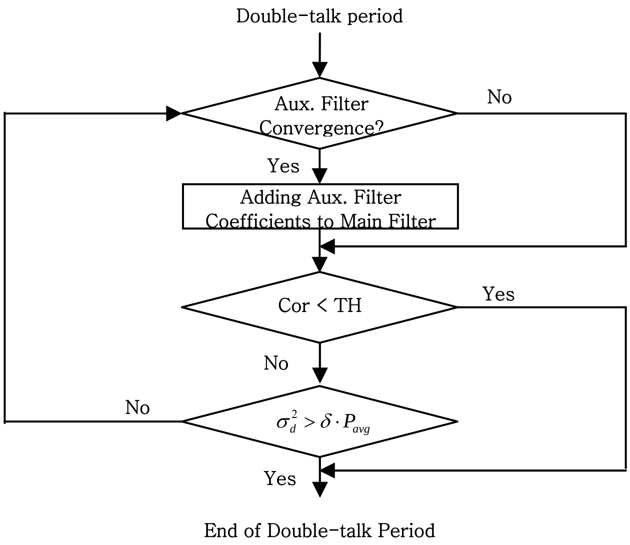
The flowchart of the DT detection algorithm is shown in Figure 2. To detect the starting point of the DT period, it compares the correlation, Cor, with the threshold ($TH = 0.6$) and then compares σ_d^2 with $\delta \cdot P_{avg}$ ($\delta = 1.5$).

$$Cor(n) = \frac{\sigma_{de}^2(n)}{\sqrt{\sigma_d^2(n)\cdot\sigma_e^2(n)}} \tag{4}$$

Here the parameter values are determined empirically, and σ_{de}^2 denotes the covariance between $d(n)$ and $e(n)$, and σ_d^2 and σ_e^2 are the power of $d(n)$ and $e(n)$, respectively. The echo path change is discriminated from the DT using the auxiliary filter.



(a)



(b)

Figure 2. Flowchart of the DT Detection Algorithm (a) Detection of Starting-point, (b) Detection of End-point

During the DT periods, the adaptive filter should stop updating of filter coefficients to prevent the coefficients from diverging. In the very noisy environment like the interior of an automobile, it will be better to stop updating filter coefficients when the background noise is very high even though it is not the DT period. Hence, we set another constraint such that the adaptive filter stops its coefficients adaptation when no echo signal exists or background noise is strong in comparison with the echo signal. The decision whether the filter tap adaptation continues or not is made by comparing the reference signal power, σ_x^2, with the estimated background noise power, σ_n^2, that is estimated from the error signal when the DT is not present, as given in eq.(5) and (6).

$$\sigma_n^2(n) = (1-\alpha)\cdot\sigma_n^2(n-1) + \alpha\cdot e^2(n) \tag{5}$$

$$\begin{cases} adaptation \quad continue\ , & if \quad \sigma_x^2(n) \geq \gamma\cdot\sigma_n^2(n) \\ adaptation \quad stop\ , & if \quad \sigma_x^2(n) < \gamma\cdot\sigma_n^2(n) \end{cases} \tag{6}$$

where α and γ are set to 1/256 and 0.5, respectively.

In general, the center clipping is used as a post-processing for further reduction of a residual echo signal [5]. This method removes compulsory some values that are under pre-defined threshold. It is difficult to remove the residual echo signal effectively when the fixed value is used for the threshold. Thus we use a variable threshold that is calculated by the square root of the average power, P_{avg}. To avoid distortion of a near-end signal, i.e., speech input from a driver, the center clipping should be performed during only single-talk periods.

3. Experiment and Discussion

An adaptive filter needs two input channels for reference and desired signals, respectively. But the TMS320C6711 DSK (Development Start Kit) supports a mono audio codec (AD535) which has a single input and output. For two channel inputs, we used a TMDX326040A that is an audio daughter card having a stereo audio codec (PCM3003). We carried out the echo cancellation experiment with the implemented AEC system in the laboratory. <Figure 3> shows the experimental set up. A reference signal that corresponds to the audio output in a car is generated by the PC, and its radiated and reverberated signal through the room from the speaker will be an echo signal. The reference signal is assigned to the left channel input and the echo signal, which is mixed with the (driver' s) speech and background noise, is assigned to the right channel input of the audio daughter card.

TI's TMS320C6x profiler in code composer studio (CCS) was used to measure the execution time of the implemented system [6]. <Table 1> shows the profiling result for executing one sample of the speech signal with the number of an adaptive filter taps 128 and 256, respectively. Since the signal used in our system is sampled at 8kHz, the execution time for processing one sample must be less than 0.125ms for real-time operation. As we can see in <Table 1>, in the case of 256 filter taps, the profiling result for the implemented system on the TMS320C6711 DSK which runs with 150MHz clock, the maximum number of clocks required for processing one sample is 5,324 which corresponds to 0.035ms. Hence, we can see that the real-time operation is achieved.

Table 1. Execution Time for One Sample of Speech

Filter Tap Number	Number of Clocks			Max. Required Time
	Max.	Min.	Average	
128	2,682	1,970	2,166	0.018ms
256	5,324	2,596	3,815	0.035ms

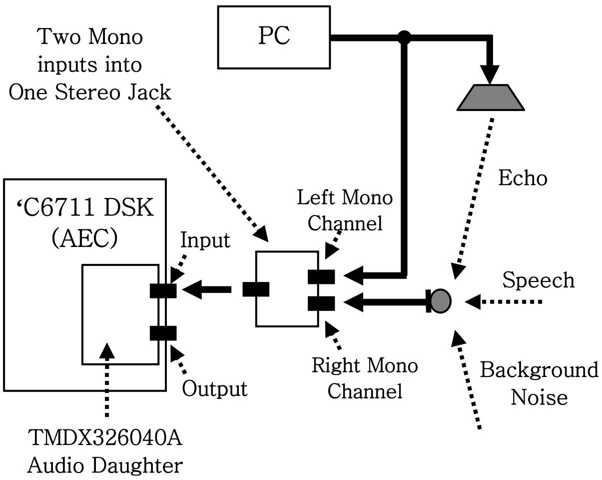


Figure 3. Experimental set up with the Implemented AEC

<Figure 4> shows an example of echo cancellation of the background music using the implemented system in the laboratory environment where the background noise (more than 20dB) such as PC fan exists. The adaptive filter tap is set to 256 and the reference signal is 20 second-long Korean music. The ERLE (Echo Return Loss Enhancement), that is a well-known parameter to evaluate the performance of an acoustic echo canceller as given in eq.(7), is used for performance evaluation. The measurement was made after the first 4000 samples which correspond to 500ms to reach the convergence region for the adaptive filter.

$$ERLE = 10 \cdot \log_{10} \frac{\sum_{n=k+1}^{N} d^2(n)}{\sum_{n=k+1}^{N} e^2(n)} \tag{7}$$

where N is the total length of a signal, and k is set to 4000. As a result of the ERLE measurement, 20.73dB was achieved, and we can say that the implemented system removes the background audio signal quite well.

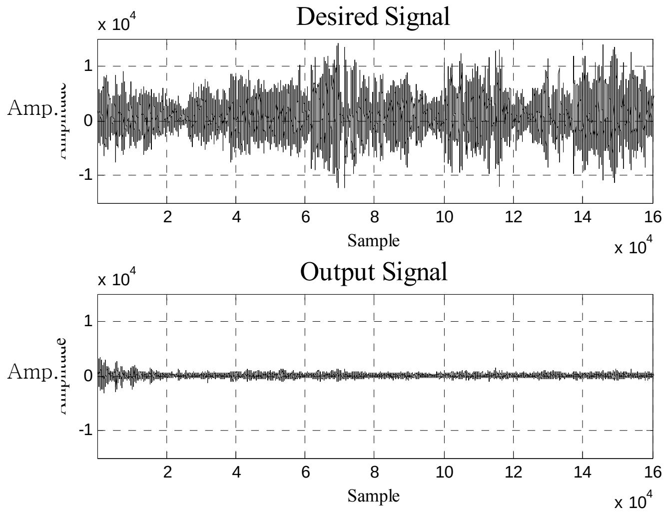


Figure 4. Example of an Echo Cancellation

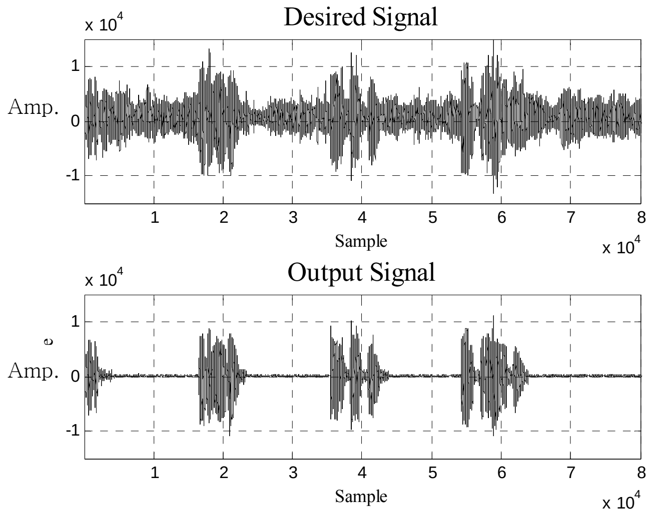


Figure 5. Extracting the Near-end Speech with Echo Cancellation (Center Clipping is Applied)

<Figure 5> shows the result of an experiment when the near-end speech is present, which is the utterance of Korean male. As we can see in <Figure 5>, the near-end speech is well preserved but most of echo signal is removed.

4. Conclusion

We have implemented an AEC with robust DT detection algorithm using the TI's TMS320C6711 DSK. For performance evaluation, the ERLE and execution time were measured. The ERLE of 20.73dB for a 20 second-long background music signal was achieved, and the maximum execution time required for processing one speech sample with sampling rate of 8kHz was 0.035ms. Consequently, the implemented acoustic echo canceller removes the background music echo signal that is treated as a noise signal in the speech recognition system quite well and its real-time operation is validated.

References

[1] Kim, S. H., Kim, J. & Bae, K. S. 2005. "Speech Interface with Echo Cancellation and Barge-in Functionalities for Robust Speech Recognition in the Car Environment." *Proceedings of the third IASTED International conference*, 250-254.

[2] Kim, S. H., Kwon, H. S., Bae, K. S. 2002. "Applying an Auxiliary Filter in the Adaptive Echo Canceller for Performance Improvement of Double-talk Detection." *IEEE 10th DSP & 2nd SPE workshop,* 207-210.

[3] Haykin, S. 1996. *Adaptive Filter Theory,* Prentice Hall, Inc., New York, ch. 9.

[4] Lee, H. W., Eun, M. E., Kim, C. K. 1997. "A Double-Talk Processing Algorithm of Acoustic Echo Canceller." *IEEK*, Vol. 17, No. 3, 10-15.

[5] Kalouptsidid, N., Theodoridis, S. 1993. *Adaptive System Identification and Signal Processing Algorithms,* Prentice Hall International (UL) Limited, ch. 10.

[6] Texas Instrument, 2000. TMS320C6000 Programmer's Guide, ch. 3.

received : January 27, 2008
accepted : March 4, 2008

▲ Wonchul Heo
LG Electronics Inc.
Pyngtaek, Korea
Tel : +82-53-940-8627
E-mail : onlyinkr@mir.knu.ac.kr

▲ Keunsung Bae
School of EECS, Kyungpook National University
Bukgu Sankyukdong 1370, Daegu, Korea
Tel : +82-53-950-5527
E-mail : ksbae@ee.knu.ac.kr