적응 후처리 과정을 갖는 마이크로폰 배열을 이용한 잡음제거기의 DSP 구현

권 홍 석, 김 시 호, 배 건 성
경북대학교 전자전기공학부
전화 : 053-940-8627 / 핸드폰 : 017-519-7585

DSP Implementation of Speech Enhancement System Using Microphone Array with Adaptive Post-processing

Hong-Seok Kwon, Si-Ho Kim, Keun-Sung Bae
School of Electronics and Electrical Engineering, Kyungpook National University
E-mail : hskwon@mir.knu.ac.kr

Abstract

In this paper, a speech enhancement system using microphone array with adaptive post-processing is implemented in real-time with TMS320C6201 DSP. It consists of delay-and-sum beamformer and adaptive post-processing filters with NLMS (Normalized Least Mean Square) algorithm. THS1206 ADC is used for collection of 4-channel microphone signals. Sizes of program memory, data ROM and data RAM of the implemented system are 15,744, 748 and 47,540 bytes, respectively. Finally 21.839×10^6 clocks per second is required for real-time operation.

I. 서론

음성인식과 음성통신 등 다양한 응용분야에서 간섭신호와 주변잡음에 강인한 특성을 갖는 입력시스템이 요구된다. 이러한 요구를 충족시키기 위하여 음성개선(speech enhancement)에 대한 연구가 많이 수행되어 왔으며 마이크로폰의 개수에 따라 크게 단일채널 기법과 다채널 기법으로 나눌 수 있다.

최근에는 단일채널 기법과 더불어 다채널 기법의 우수한 잡음제거 성능으로 인하여 다채널 기법에 대한 연구가 많이 진행되고 있다. 다채널을 이용한 잡음제거 기법은 적응빔형성(adaptive beamforming), 지연합빔형성(delay-and-sum beamforming), 적응후처리를 갖는 방법으로 구분할 수 있다[1,2]. 특히 적응후처리를 갖는 방법 중에서 Zelinski는 지연합빔형성된 신호에 LMS(Least Mean Square) 알고리즘을 갖는 적응필터를 적용하여 잡음을 제거하는 방법을 제안하였다[2]. 본 연구에서는 Zelinski가 제안한 방법을 4-채널 선형 마이크로폰 배열로 구성된 TMS320C6201 EVM에서 실시간으로 구현하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 Zelinski에 의해 제안된 적응후처리를 갖는 마이크로폰 배열에 대하여 설명하고 3장에서는 TMS320C6201 DSP와 4채널 ADC(Analog-to-Digital Converter)인 THS1206에 대하여 서술한다. 4장에서는 잡음제거 성능 평가를 위한 실험조건과 결과를 제시, 분석하고 마지막으로 5장에서 결론을 맺는다.

II. 적응후처리를 갖는 지연합빔형성기

후처리과정으로서 Wiener 필터링을 수행하는 마이크로폰 배열은 주변잡음이 포함된 음성신호에서 잡음을 제거하는데 우수한 성능을 보이고 있다[2]. 마이크로폰 배열에 입사된 음성신호는 지연합빔형성기를 통

하여 일차적으로 잡음이 제거되고 Wiener 필터를 이용한 후처리로 좀 더 향상된 잡음제거 성능을 얻을 수 있다. 그림 1은 [2]에서 제안한 지연합빔형성과 Wiener 필터를 사용한 잡음제거 과정을 보이고 있다.

그림 1에서 잡음을 제거하는 과정은 크게 두 단계로 나누어진다. 첫 번째 단계에서는 4개의 마이크로폰에 입력된 신호의 잡음성분을 지연합빔형성기로 줄인다. 지연합빔형성은 마이크로폰 배열에서 잡음을 제거하는 가장 기본적인 방법으로서 입사되는 방향에 대한 도달 지연시간을 보상하여 동위상으로 신호를 맞추어서 잡음을 개선시키는 방법이다. 이때 각 마이크로폰에 입사되는 잡음간에 서로 상관성이 없다고 가정하면 4개의 마이크로폰 배열에 대해서는 약 6dB정도의 SNR이 개선된다. 두 번째 단계에서는 지연합빔형성기에 의해서 개선된 음성인 X_s에서 원음성에 가까운 신호를 WFS로 추정하여 잡음을 더 제거한다. 여기서 WFS의 계수는 WF의 계수를 복사하여 사용한다.

3개의 마이크로폰 M_1, M_2, M_3에 대한 지연합빔형성된 d는 단일채널 입력음성인 v에 비해서 원음성에 가깝기 때문에 WF의 주입력신호로 d를 사용하고, v를 WF의 기준입력신호로 사용한다. 이때 각 마이크로폰의 입력신호에 존재하는 잡음이 서로 상관성이 없다고 가정하면 v가 d를 추정하도록 WF의 필터 계수는 수렴한다. WFS의 입력음성인 X_s는 단일채널 입력음성인 v보다 원음성에 더 가깝기 때문에 수렴한 WF의 계수를 WFS에 복사하여 X_s를 WFS에 통과시킴으로써 더 많은 잡음을 제거할 수 있다[2].

그림 1에서 T_w는 시스템의 실현성을 보장하기 위한 시간지연으로서 WF의 임펄스응답 길이의 반으로 두고 전처리(pre-emphasis)는 신호의 고주파성분을 증가시키기 위해 사용된다. WF의 필터 계수는 식 (1)과 식 (2)로 주어지는 NLMS 알고리즘을 사용하였다[3].

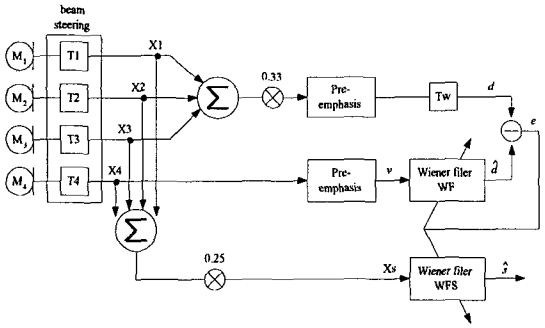

그림 1. 적응후처리를 갖는 잡음제거 시스템

$$e(i) = d(i) - \hat{d}(i) = d(i) - \boldsymbol{w}^T(i) \cdot \boldsymbol{v}(i) \quad (1)$$

$$\boldsymbol{w}(i+1) = \boldsymbol{w}(i) + \beta \cdot \frac{e(i) \cdot \boldsymbol{v}(i)}{\boldsymbol{v}^T(i) \cdot \boldsymbol{v}(i)} \quad (2)$$

여기서 i는 i번째 샘플을 의미하며, d, e, $\boldsymbol{w}$, $\boldsymbol{v}$는 각각 주입력신호, 오차신호, WF의 필터계수, WF의 기준입력신호를 의미한다. 그리고 β는 적응상수로서 $0<\beta<2$의 범위를 가져야 하며 WF의 필터 계수는 오차신호가 줄어드는 방향으로 수렴한다.

III. DSP 구현

TMS320C6201 DSP는 Texas Instrument사의 TMS320C62xx 고정소수점 DSP 제품군의 하나로서 그림 2와 같이 CPU, 메모리, 주변장치 등으로 구성된다. CPU는 하나의 클럭 사이클당 최대 8개의 32-비트 명령어를 동시에 수행할 수 있기 때문에 최대 200MHz의 클럭에서 1600 MIPS의 성능을 가질 수 있다[4]. TMS320C62xx EVM은 64k-바이트의 프로그램 메모리와 64k-바이트의 데이터 메모리로 구성되는 내부 메모리와 SBSRAM, SDRAM 및 비동기식 메모리인 SRAM과 EPROM으로 구성되는 외부 메모리가 있다[5].

THS1206 ADC는 12 비트 양자화간격과 6-MSPS (MegaSamples Per Second)를 갖는 저전력용 ADC로서 최대 4-채널까지 입력구성이 가능하다. 그리고 순환버퍼 특성을 갖는 16-워드 FIFO를 이용하여 고속으로 데이터를 수집할 수 있으며 프로세서의 부하를 줄이기 위하여 DMA를 통한 동기전송에 사용될 수 있다. THS1206 EVM은 THS1206 ADC를 이용하여 제작된 보드로서 DSP의 EMIF에 연결된 daughter card connector와의 인터페이스를 제공한다[7].

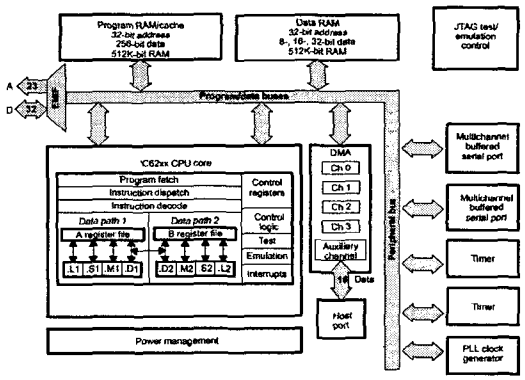

그림 2. TMS320C6201 DSP의 내부구조

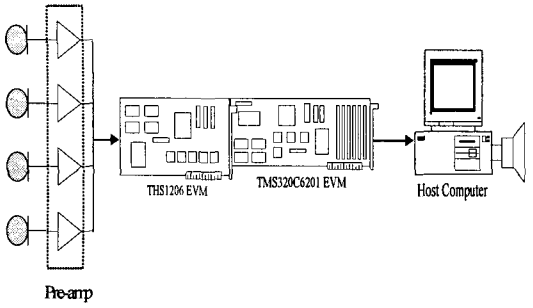

그림 3. 음성개선 시스템 구성도

그림 3은 구현된 음성개선 시스템의 구성도를 보인 것이다. 4-채널 마이크로폰은 무지향성 마이크로폰을 사용하여 THS1206 EVM의 입력에 연결한다. 이때 ADC 과정에서 발생하는 신호의 aliasing을 피하고 입력신호의 증폭을 위하여 프리앰프(pre-amplifier)를 구성하였다. 또한 각 채널의 입력을 동시에 A/D한 다음 그 결과를 DSP의 DMA를 이용하여 TMS320C6201 EVM의 daughter card connector를 통해 수집된다. TMS320C6201 DSP에서는 수집된 데이터를 우선 지연합빔형성기로 입력신호의 입사각도를 구하고 그 각도에 맞도록 각 마이크로폰의 시간지연을 보상하여 지연합빔을 형성시킨다. 지연합빔형성된 결과는 적응후처리를 통하여 더 많은 잡음이 제거되고 그 결과는 호스트 컴퓨터에서 재생된다.

Ⅳ. 실험 및 결과

구현된 잡음제거 시스템의 성능을 평가하기 위하여 잡음제거 정도를 SNR로 측정하였으며 실시간 동작을 확인하기 위하여 DSP의 연산량를 조사해 보았다. 실험에 사용된 신호의 표본화주파수는 각 채널에 대하여 100kHz로 하였으며 12-비트 양자화를 수행하였다. 마이크로폰 사이의 거리는 6.8cm로 선형으로 위치시켰으며 조향위치는 식 (3)을 이용하여 추정하였다. 추정된 조향위치에 대한 지연시간을 각 채널 입력신호에 보상하면서 64차를 갖는 antialiasing 디지털 필터를 이용하여 10kHz로 다운샘플링하였으며 후처리를 위한 적응필터의 차수는 128로 두었다.

$$\hat{d} = \arg\max_{n} \sum_{i=0}^{I-1} \left(\frac{1}{4} \sum_{m=1}^{4} x_m(i + \overline{D_m}(n)) \right)^2 \qquad (3)$$

여기서, x_m은 m번째 마이크로폰에 입력된 신호를 말하며 n은 위치 추정을 위한 지연시간으로서 100kHz 신호에서 10kHz로 다운샘플링하므로 $-10 \le n \le 10$의 범위를 가진다. 그리고 $\overline{D_m}$은 m번째 마이크로폰의 n에 따른 상대적인 지연시간을 나타내며 I는 위치 추정에 사용된 각 채널의 샘플수에 해당된다.

구현된 시스템의 잡음제거 정도를 측정하기 위하여 각 채널별로 100kHz의 원음성에 백색 가우시안 잡음을 첨가하여 각 채널별 잡음음성을 만들어서 모의실험을 하였다. 표 1은 100kHz에서 입력 잡음음성이 각각 -15dB, -10dB, -5dB, 0dB일 때 10kHz의 잡음음성과 지연합빔형성기의 출력음성, 적응후처리를 갖는 잡음제거기의 출력음성에 대한 SNR을 보이고 있다. 이때 SNR은 식 (4)와 같이 정의하였으며, 식에서 s는 원음성을, $\hat{s}$은 잡음이 제거된 음성을 말하며 N은 실험에 사용된 음성의 전체 샘플수를 말한다.

$$SNR[dB] = 10\log_{10}\left(\frac{\sum_{i=0}^{N-1} s^2(i)}{\sum_{i=0}^{N-1} (\hat{s}(i) - s(i))^2} \right) \qquad (4)$$

표 1에서 지연합빔형성기의 출력음성은 약 6dB 정도의 SNR 개선을 보였으며 후처리 과정을 거침으로써 더 많은 잡음이 제거된다는 것을 볼 수 있다. 특히 입력음성의 SNR이 낮을수록 음성개선 정도가 우수하게 나타났으며 입력음성의 SNR이 높은 경우에는 후처리 과정에서의 SNR이 지연합빔형성의 SNR보다 낮았다. 이는 SNR이 낮으면 잡음을 제거하는 방향으로 적응필터가 잘 수렴하지만 SNR이 높으면 적응필터가 수렴하면서 신호의 크기를 줄이기 때문이라고 생각된다.

실시간으로 동작할 때 잡음제거 성능을 평가하기 위하여 그림 4와 같이 7.0m×4.8m×2.6m를 갖는 실험실 환경에서 잡음제거 실험을 수행하였다. 입력음성은 스피커를 통하여 S에서 남성 목소리를 12초 정도 재생하였으며 잡음은 N의 위치에서 백색잡음을 재생시켰다.

표 1. 입력 SNR에 따른 잡음제거 정도 [dB]

100kHz 입력음성	-15	-10	-5	0
10kHz 입력음성	-3.96	0.34	6.01	11.23
지연합빔형성기 출력	0.67	6.70	11.22	16.11
적응후처리의 출력	6.03	8.90	13.60	15.20

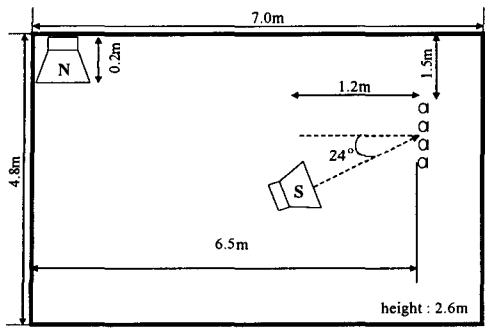

그림 4. 실험실 배치도

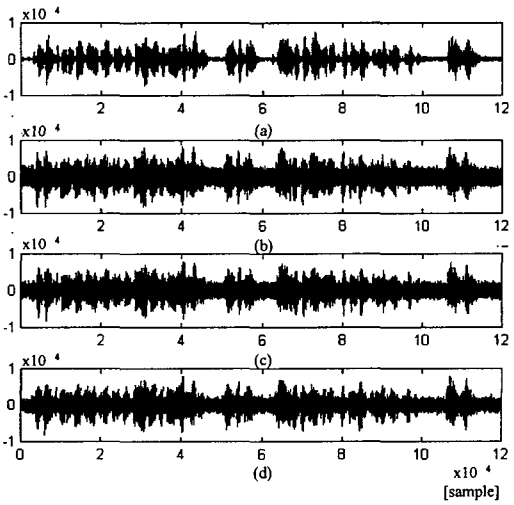

그림 5. 실험실 환경에서의 잡음제거 정도
(a) 원음성, (b) 잡음음성,
(c) 지연합빔형성 결과, (d) 적응후처리 결과

그림 5는 입력음성, 잡음음성, 지연합빔형성기의 출력음성, 적응후처리로 제거된 출력음성을 각각 그린 것이다. 그림을 보면 알 수 있듯이 모의실험에 비해서 잡음제거 성능이 저하되는데 이는 각 마이크로폰에 입력된 신호의 저주파 성분간의 상관성과 실험실에서 발생하는 반향으로 인한 저하라고 생각된다. 그러나 지연합빔형성기의 출력음성보다 적응후처리 출력음성에서 잡음이 더 잘 제거되고 있음을 확인할 수 있다.

실시간 구현을 위하여 본 연구에서 사용한 메모리의 크기를 측정해 보았을 때 프로그램 메모리는 15,774 bytes, 데이터 ROM은 748 bytes, 데이터 RAM은 47,540 bytes를 사용하였으며 대부분이 100kHz로 샘플링된 4-채널 입력신호를 저장하기 위한 버퍼로 사용되었다. 그리고 24ms에 해당되는 9600 샘플로 구성된 한 프레임을 처리하는데 약 524,126 클럭을 사용하였다. 이는 초당 21.84×10^6 클럭에 해당되며 200MHz까지 지원하는 TMS320C6201 DSP의 사용 가능한 연산량의 약 10.92%에 해당됨을 의미한다.

V. 결론

본 연구에서는 4-채널 지연합빔형성기와 적응필터를 이용한 후처리를 갖는 잡음제거기를 TMS320C6201 DSP를 사용하여 실시간으로 구현하였다. 화자의 위치는 조향빔형성 기법으로 추정하였으며 후처리에 사용된 적응알고리즘은 NLMS알고리즘을 사용하였다. 모의실험에서는 낮은 SNR에서 우수한 잡음제거 성능을 보였다. 그리고 실시간으로 동작시킨 경우에는 마이크로폰에 입력되는 잡음 사이의 상관성과 실험실의 반향으로 인하여 높은 성능을 얻을 수 없었지만 지연합빔형성보다 우수한 성능을 나타내었다. 실시간 동작에 사용된 프로그램 메모리, 데이터 ROM, 데이터 RAM은 각각 15,744 bytes, 748 bytes, 47,540 bytes이며 초당 필요한 연산량은 약 21.839×10^6 클럭이었다.

본 연구는 한국과학재단 목적기초연구(R01-1999-00233) 지원으로 수행되었음.

참고문헌

[1] Sven Fischer, Klaus Uwe Simmer, "Beamforming microphone arrays for speech acquisition in noisy environment," Speech Communication, Vol.20, pp.215-227, 1996

[2] Rainer Zelinski, "Noise Reduction Based on Microphone Array with LMS Adaptive Post-Filtering," IEE letters, Vol.26, No.24, pp.2036-2037, 1990

[3] Simon Haykin, *Adaptive Filter Theory*, Prentice-Hall, 1996

[4] Texas Instruments, TMS320C6201: Fixed Point Digital Signal Processor, 2000

[5] Texas Instruments, TMS320C6201/6701 Evaluation Module Technical Reference, 1998

[6] Texas Instruments, THS1206 12-Bit 6 MSPS, Simultaneous Sampling Analog-to-Digital Converters, literature number SLAS217E, 2000

[7] Texas Instruments, THS1206, THS12082, THS10064, THS10082 Evaluation Module User's Guide, 2001