

TMS320VC5510을 이용한 AMR-WB 음성부호화기의 실시간 구현

조재민, 김준, 김정민, 배건성
경북대학교 전자전기공학부

Real-Time Implementation of the AMR-WB Speech Codec Using TMS320VC5510 DSP

Jae Min Joh, Jun Kim, Jung Min Kim, Keun Sung Bae
School of Electronic and Electrical Engineering, Kyungpook National University
gagjudy@mir.knu.ac.kr

요약

본 연구에서는 ETSI 및 3GPP에 의해 개발된 광대역 음성부호화의 표준안인 AMR-WB 알고리즘을 분석하고, TMS320VC5510 DSP를 이용한 실시간 구현 결과를 제시하였다. AMR-WB 음성부호화기의 실시간 구현을 위해 프로그램 최적화 작업을 수행하였고, 구현된 음성부호화기의 성능을 평가하기 위해서 프로그램 메모리와 데이터 메모리의 크기, 그리고 한 프레임당 수행 시간을 측정하였다. 구현된 시스템의 프로그램 메모리는 약 65.6 kbytes, 데이터 메모리는 약 73.8 kbytes 정도의 크기를 나타내었으며, 한 프레임인 20 msec를 처리하는데 소요되는 cycle 수가 평균 1,247,115 정도로 약 6.24 msec 내에 처리할 수 있음을 보였다. 마지막으로 DSP로 구현한 AMR-WB 음성부호화기의 결과가 PC에서 시뮬레이션 한 결과와 일치함을 검증하였고 실시간으로 동작함을 확인하였다.

1. 서론

차세대 이동통신 및 인터넷 통신 시스템의 가장 큰 특징은 넓은 대역폭을 사용하며, 음성에서 동영상까지 다양한 멀티미디어 데이터를 전송할 수 있다는 것이다. 이에 따라 음성신호의 부호화도 기존의 300-3400Hz 협대역 전화망에서 음성의 명료도와 자연성을 증가시켜 보다 양호한 음질을 얻을 수 있는 50-7000Hz의 광대역 음성부호화에 대한 관심이 증가하고 있다.

광대역 음성신호는 디지털 전송을 위해 16kHz 샘플율로 샘플링된다. 각각의 샘플이 16bit 정수로 표현될 때 실제 비트율은 256 kbit/s가 된다. 이러한 음성 데이터를 낮은 전송률에서 전송하기 위해서는 보다 효율적인 음성부호화 알고리즘이 필요하다. ITU-T는 1988년에 처음으로 광대역 음성부호화 표준안으로서 48, 56, 64 kbit/s의 비트율을 갖는 G.722 방식을 선정하였고, 1999에 24, 32 kbit/s의 비트율을 갖는 G.722.1 방식을 정하였다. 2000년에는 채널상태에 따라 6.6-23.85 kbit/s의 가변전송률을 갖는 AMR-WB(Adaptive Multi-Rate WideBand) 방식이 GSM망 및 3세대 무선통신 시스템에서의 광대역 부호화 방식의 표준안으로 채택되었다[1]. AMR-WB 방식은 ITU-T의 새로운 표준안 선정에 참여하여 2001년에 최종안으로 채택되었다.

본 연구에서는 ETSI에서 공개한 AMR-WB의 ANSI-C 소스 코드[2]를 분석하고 최적화 작업을 수행하여 TI사의 DSP인 TMS320VC5510 DSP(DSP Start Kit)에서 실시간으로 구현하였다. 또한 ANSI-C 소스로 얻어진 결과와 DSP 보드상에서 구현된 결과가 일치함을 확인하였으며, 마이크와 스피커를 이용하여 음질의 왜곡없이 실시간으로 동작함을 확인하였다.

본 논문의 구성은 다음과 같다. 2장에서 AMR-WB 음성부호화 시스템과 알고리즘에 대해 간략하게 서술한다. 3장에서는 TMS320VC5510 DSP의 구조, 개발환경 및 최적화 작업에 관한 내용을 설명한다. 4장에서는 구현결과를 제시하고 마지막으로 5장에서 결론을 맺는다.

2. AMR-WB 음성부호화 시스템

2.1 AMR-WB 시스템 구성

AMR-WB 시스템은 9가지 비트율을 갖는 음성부호화기, SCR(Source Controlled Rate)를 위한 VAD(Voice Activity Detector), 묵음구간에서의 배경잡음 생성을 위한 CNG(Comfort Noise Generation), 손실된 프레임의 보상을 위한 ECM(Error Concealment Mechanism) 등으로 구성된다[1]. AMR-WB 음성부호화기는 ACELP 방식에 기반하고 있으며, 6.6kbit/s 부터 23.85kbit/s 사이의 9가지 전송모드를 갖는다[3]. 또한 DTX (Discontinuous Transmission)[4]에 사용되는 1.75kbit/s의 배경잡음 부호화 모드를 갖는다. AMR-WB 음성부호화기는 20 msec 길이의 프레임마다 비트율을 선택할 수 있으며, 9 가지 전송모드의 비트율은 표 1과 같다.

SCR 동작은 음성프레임이 배경잡음만을 포함하고 있을 경우, 음성부호화기가 아주 낮은 비트율로 프레임을 부호화하는 것을 말한다. 이는 전력소모를 줄여줌으로써 단말기의 배터리 수명을 연장시킬 수 있고, 부호화에 요구되는 평균 비트율을 감소시킴으로써 채널을 효율적으로 활용할 수 있다. SCR 동작을 위해서는 송신단에서 VAD를 통한 음성구간의 검출과 배경잡음 추정이 필요하다. 이때 배경잡음에 대한 정보의 전송은 SID(Silence Descriptor)[4] 프레임의 전송을 통해 이루어지며, 수신단에서 CNG에 이용된다.

VAD는 20ms 길이의 음성프레임마다 음성신호의 포함 여부를 플래그(VAD FLAG)로 표현한다. 최종 플래그는 음성신호를 필터뱅크에 통과시켜서 구한 대역별 에너지 정보와 개루프(open loop) 피치 이득 정보로 구한 문 성분 검출 결과를 모두 고려하여 해결진다.

CNG는 수신단에서 인위적인 잡음성분을 발생시켜 비음성 프레임 구간에서 듣기에 거부하지 않은 소리를 발생시키는 과정이다. 이를 위해 송신단에서는 배경잡음을 추정하여 낮은 비트율의 SID 프레임 형태로 정보를 전송한다.

ECM은 송신단에서 전송한 정보에 오류가 생기거나 손실될 경우 음성신호와 상관성이 없는 단순한 묵음을 재생하지 않고, 이전에 성공적으로 수신된 프레임 파라미터를 바탕으로 합성음을 복원하여, 프레임 오류로 인한 합성음의 음질 저하 방지를 목적으로 수행된다.

표 1. AMR-WB 음성부호화기의 모드에 따른 비트율

Codec mode	Source codec bit-rate
AMR 23.05	23.05 kbit/s
AMR 19.85	19.85 kbit/s
AMR 18.25	18.25 kbit/s
AMR 15.85	15.85 kbit/s
AMR 14.25	14.25 kbit/s
AMR 12.65	12.65 kbit/s
AMR 8.85	8.85 kbit/s
AMR 6.60	6.60 kbit/s
AMR 1.75	1.75 kbit/s

2.2 AMR-WB 인코더

AMR-WB 음성부호화기는 16kHz의 샘플율에서 20 ms 길이의 프레임단위로 동작하는데, 각 프레임은 5 ms 길이로 갖는 4개의 서브프레임으로 나누어진다. 이때, 계산량의 감소와 중요한 정보를 가진 대역폭의 비트 할당에 초점을 맞추기 위해 입력신호를 50~6400 Hz의 저대역 신호와 6400~7000Hz의 고대역 신호로 나누어 각 대역을 독립적으로 부호화한다. AMR-WB 인코더의 구성은 그림 1과 같다.

저대역 신호의 부호화를 위해 입력신호는 12.8kHz의 샘플율로 다운샘플링되고 pre-emphasis 과정이 적용된 후, 협대역 음성부호화기의 표준안에 널리 사용되는 ACELP 방식으로 부호화된다. 선형예측분석은 20ms의 프레임마다 한번씩 수행되고, 선형예측계수는 양자화 및 전송특성을 고려하여 ISP(Immittance Spectral Pairs)[5]로 변환된다. 또한, 고정 여기신호 및 적응 여기신호를 위한 코드북의 탐색은 5ms 길이의 서브프레임 단위로 수행된다. 광대역 음성신호의

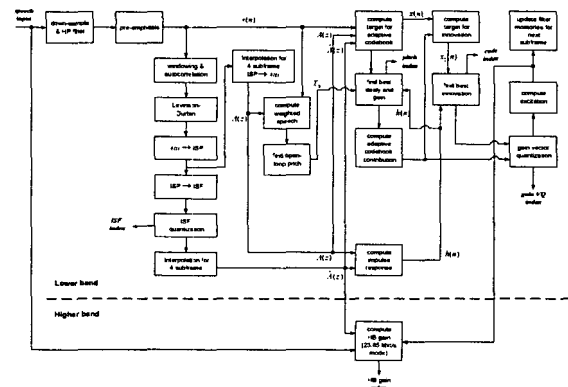


그림 1. AMR-WB 인코더의 구조

스펙트럼은 협대역 신호에 비해 보다 넓은 동적범위를 갖기 때문에 AMR-WB에서는 스펙트럼의 포먼트 특성뿐만 아니라 포락선의 경사도를 고려하는 지각가중 필터가 적용된다. 그리고 유성음 구간에서의 피치 특성을 보다 잘 표현하기 위해 다양한 특성의 필터를 피치 분석에 적용하고 있다[6].

디코더에서는 그림 2와 같이 수신된 정보 비트열로부터 ISP 벡터, 4개의 분수 피치 값, LTP(Long Term Prediction)필터 파라미터, 여기신호 벡터, 그리고 벡터 양자화된 적응 코드북과 고정 코드북의 이득을 추출한다. 이들 파라미터를 이용하여 ACELP 합성 모델을 이용해서 서브프레임 단위로 여기신호를 구성하고 선형예측 합성필터를 통과시켜 12.8kHz 샘플율을 갖는 저대역 음성신호를 복원한다. 또한 복원된 음성신호는 인코더와 같은 값을 사용하여 de-emphasis를 적용하고, 마지막으로 16kHz 샘플율의 신호로 업샘플링된다.

고대역 신호는 저대역 신호에서 분석한 선형예측 필터 특성을 이용하여 재현한 파라미터와 랜덤 여기신호로 합성한다. 23.85kbit/s 모드의 고대역 신호의 이득은 유성음화정보(coicing information)를 이용하여 조정된다. 합성된 고대역 음성신호는 저대역 음성신호와 더해져서 최종 합성음을 구성하게 된다.

3. TMS320VC5510 DSK을 이용한 실시간 구현

3.1 TMS320VC5510 DSK의 구조 및 개발환경

본 연구에 사용된 TI사의 TMS320VC5510은 'C55xx 계열의 고정소수점 DSP 중의 하나로서 200 MHz에서 최대 400 MIPS의 성능을 발휘하며, 1 MIPS에 약 0.05 mW의 전력 소모를 가지는 저 전력 고성능 모델이다. 기본적으로 multiple-bus architecture와 dual 17-bit × 17-bit multiplier, 40-bit adder로 구성된다.

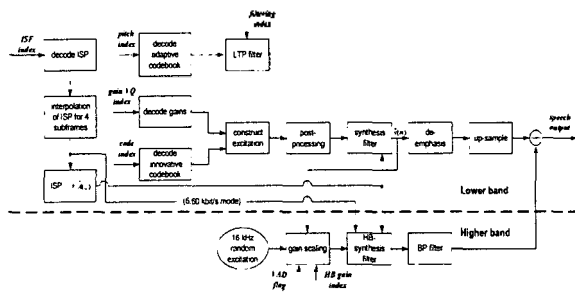


그림 2. AMR-WB 디코더의 구조

본 연구에서 사용한 DSK는 TMS320VC5510 DSP와 더불어 AIC23 stereo audio codec으로 이루어져 있다. 또한 DSK에 장착된 메모리는 내부 메모리와 외부 메모리로 구성되며 내부 메모리는 64 kbytes의 DARAM과 256 kbytes의 SARAM으로 되어 있고, 외부 메모리는 8 Mbytes의 SDRAM과 512 kbytes의 FLASH memory로 이루어져 있다. 개발환경으로 소프트웨어 툴은 2.12.07 version의 CCS(Code Composer Studio)[7]를 사용하고 이를 이용하여 디버깅 작업 및 실시간 분석을 수행하였다.

3.2 TMS320VC5510 DSK를 이용한 실시간 구현

AMR-WB 음성부호화기의 실시간 구현은 ETSI 및 3GPP에서 제공되는 고정 소수점 ANSI-C 코드를 기반으로 하였으며, 소스 코드에는 multi-rate 음성부호화 알고리즘 뿐만 아니라 VAD, CNG, SCR도 포함하고 있다. 소스 코드 파일 중에서 Init_WMOPS_counter(), Init_WMOPS_counter(), WMOPS_output() 등과 같이 알고리즘의 복잡도나 연산 횟수를 계산하는 함수들을 정의하고 있는데, 음성부호화 과정과는 직접적인 연관이 없으므로 실시간 구현에서 제외하였다. 또한 printf(), fwrite() 함수 등과 같은 불필요한 입출력 함수들을 소스 코드에서 제외하였다.

본 연구에서 실시간 동작을 위해 수행한 주요 과정은 다음과 같다.

- 1) 소스 파일에 정의되어 있는 ETSI의 기본산술 연산 함수들을 TI에서 제공하는 intrinsic 함수로 대체하였다.
- 2) CCS의 compile option 기능 중에서 함수의 원형을 inline화 하는 작업을 사용하였다. 이는 함수의 호출이 함수 내부의 코드로 대체된다.
- 3) 내부 메모리가 외부 메모리보다는 좀 더 빠른 성능을 나타내므로 link command file을 사용하여 내부 메모리의 사용을 최대화하였다.
- 4) Loop 문에서 trip counter의 크기가 작을 경우 loop unrolling으로 loop 문의 branch instruction을 줄였다.
- 5) TI에서 제공하는 DSP library[8]를 사용하여 소스 코드 내의 convolution 함수에 대해 DSP library의 convol()의 함수로 대체하였다.
- 6) CCS에서 제공하는 compile option을 최대한으로 적용하였다.

4. 실시간 구현 및 결과

본 연구에서 구현된 AMR-WB 음성부호화 시스템의 성능을 평가하기 위해 프로그램 크기 및 수행속도를 측정하였다. 표 2는 AMR ANSI C 소스 코드의 최적화 작업 없이 DSK에서 구현한 결과와 3절에서 언급한 최적화 작업을 수행한 후의 결과값 나타낸 것이다. 이때 AMR-WB의 8개 모드 중에서 가장 높은 비트율을 가지는 모드, 즉, 23.85kbit/s 에 대한 결과값을 제시하였다. 한 프레임에 대한 수행시간은 프레임 당 최대 소모되는 clock 수와 C5510 CPU의 clock인 200MHz를 사용하여 식(1)과 같이 구해진다.

$$Execution\ Time = \frac{cycles}{200M} \times 1000 [ns] \quad (1)$$

음성부호화기의 전송모드에 따라 연산 시간이 플리지만 가장 높은 비트율을 기준으로 20ms 음성프레임을 인코딩/디코딩 하는데 약 6.24ms가 소요되었으므로, TMS320VC5510 DSK에서 최적화 작업을 통해 구현된 음성부호화기는 full-duplex 모드로 3채널을 수용할 수 있음을 확인할 수 있었다. 또한, 실시간으로 구현된 AMR-WB 음성부호화기가 올바르게 동작하는지를 확인하기 위해 임의의 입력 음성신호에 대해 PC에서 ANSI C 소스로 구현하여 얻어진 인코딩/디코딩된 출력 값과 TMS320VC5510에서 구현하여 얻은 인코딩/디코딩된 출력 값을 비교해 보았다. PC상에서 출력되는 값과 DSK에서 얻어진 값이 일치하는 것을 확인할 수 있었으며, 이를 그림 3에 나타내었다.

5. 결론

본 연구에서는 광대역 음성부호화기의 새로운 표준안인 AMR-WB 알고리즘을 분석하고, 고정소수점 DSP를 사용하는 TMS320VC5510 DSK 보드에서의 실시간 구현 결과를 제시하였다. 구현된 음성부호화기의 성능을 평가하기 위해서 프로그램 메모리와 데이터 메모리의 크기, 그리고 한 프레임 당 수행 시간을 측정하였으며, 프로그램 메모리는 65.58kbytes, 데이터 메모리는 73.8kbytes 정도의 크기를 나타내었다. 또한 20ms의 한 프레임 처리하는데 약 6.24ms가 소요 되었으며, 이는 full-duplex 모드로 3 채널 정도의 AMR-WB 시스템을 실시간으로 처리할 수 있음을 나타낸다.

표 2. 최적화 작업 전, 후의 성능비교 (clock 수)

구분	Total	encoding	decoding	수행 시간	메모리
최적화 작업전	59,119,598	47,948,001	11,031,159	295.6 ms	113.4k/191.4k
최적화 작업후	1,247,115	.	.	6.24 ms	65.58k/73.8k

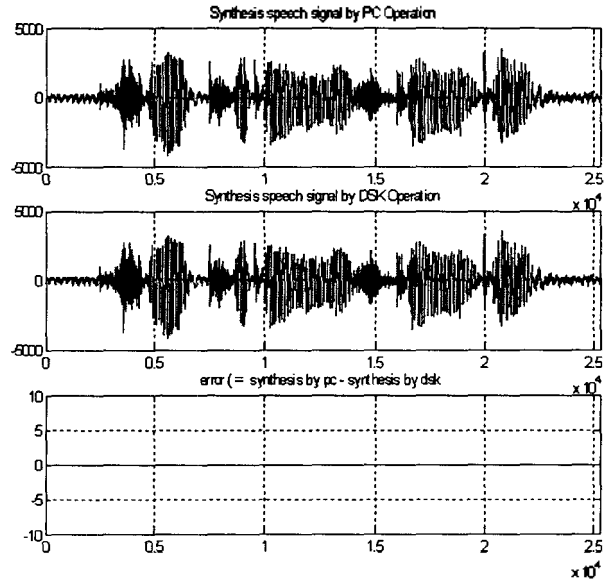


그림 3. PC 및 DSK 상에서 얻어진 출력 파형 비교

참고문헌

- 3GPP TS 26.073 V7.6 : Adaptive Multi Rate(AMR) speech : ANSI-C code for the AMR.
- 3GPP TS 26.171, "AMR Wideband Speech Codec: General Description," 3GPP Technical Specification.
- 3GPP TS 26.190, "AMR Wideband Speech Codec: Transcoding Functions," 3GPP Technical Specification.
- 3GPP TS 26.193 "AMR Wideband Speech Codec: Source Controlled Rate operation," 3GPP Technical Specification.
- Bistritz and S. Peller, "Immittance spectral pairs (ISP) for speech encoding" ICASSP, pp. 9-12, 1993.
- B. Bessette and R. Lefebvre et al., "Techniques for high-quality ACELP coding of wideband speech," Eurospeech 2001, pp. 1997-2000, 2001.
- Texas Instrument, Code Composer Studio Getting Started Guide, 2001.
- Texas Instruments, TMS320C55x DSP Library Programmer's Reference, 2002.