

TMS320VC5510 DSK를 이용한 2.4kbps MELP 보코더

이 상 원, 김 준, 배 건 성
경북대학교 전자전기공학부

2.4kbps MELP Vocoder with TMS320VC5510 DSK

Sang Won Lee, Jun Kim, Keun Sung Bae

School of Electronic and Electrical Engineering, Kyungpook National University
sleewon@samsung.com

요약

본 논문에서는 저전송율 음성부호화기인 2.4kbps MELP(Mixed Excitation Linear Prediction) 보코더를 TI(Texas Instruments) 사의 고정소수점 DSP인 TMS320VC5510을 이용하여 실시간 구현한 결과를 제시한다. MELP 보코더는 전통적인 LPC 합성방식에 기반한 것으로, 2.4kbps LPC 보코더가 여기신호를 유성음 구간에 대해서는 펄스열로, 무성음 구간에 대해서는 백색잡음 신호로 단순화시켜 합성함으로써 음질이 저하되거나 buzz 현상이 나타나는 단점을 적절히 혼합된 형태의 여기신호를 사용함으로써 보완한 것이다. DDVPC(Defense Digital Voice Processing Consortium)에서 제공하는 ANSI C 소스 코드를 이용하여 TMS320VC5510 DSK에서 실시간 동작이 가능하도록 최적화 작업을 수행하였으며, 구현된 MELP 보코더는 프로그램 메모리 46.5 kbyte와 데이터 메모리 57.9 kbyte를 가지며, 22.5ms의 한 프레임을 처리하는데 1326531 클럭(6.6 ms)이 소요되었다.

1. 서론

이동통신 및 인터넷 전화 등과 같은 디지털통신 기술의 발달과 함께 음성부호화에 대한 연구도 꾸준히 진행되고 있다. 음성부호화 방식은 크게 세 가지로 나눌 수 있는데 그 첫 번째는 파형부호화 방식, 두 번째는 파라미터 부호화 방식이고 세 번째는 혼합부호화방식이다[1]. 기존의 음성부호화 알고리즘을 실제 시스템에서 구현하고자 할 때, 먼저 목적에 부합하는 적절한 전송율을 갖는 부호화 방식을 선정하고, 타겟 DSP에서 최소한의 메모리 용량과 계산량을 가지도록 최적화 작업을 수행하

여 알고리즘을 포팅(porting)하는 과정을 거치게 된다. 본 연구에서는 4 kbps이하의 낮은 전송율에서 우수한 음질을 보이는 파라미터 부호화 방식인 2.4kbps MELP 보코더 알고리즘을 분석하고 TI사의 고정소수점 DSP인 TMS320VC5510에서 실시간으로 동작하도록 구현하였다. 본 논문의 구성은 다음과 같다. 먼저 2장에서는 2.4kbps MELP 보코더의 기본적인 내용에 대해 서술한다. 3장에서는 TMSVC5510 DSK의 특징 및 개발환경에 대해 간략히 설명하고, 이를 이용한 실시간 구현 과정 및 결과를 제시한다. 마지막으로 4장에서 결론을 맺는다.

2. 2.4kbps MELP 보코더

2.4 kbps MELP 보코더는 전통적인 LPC 합성방식에 기반한 것으로, 8 kHz로 샘플링 되어 분석프레임 당 22.5 ms의 길이를 갖는다. 기존의 2.4kbps LPC 보코더에서 유/무성음에 따라 주기적인 임펄스열이나 백색잡음을 합성 필터의 여기신호로 사용함으로써 문제가 되는 합성음의 울림현상(buzz), 톤 성분(tonal noise)등의 부자연스러운 성분을 제거하고 음질을 개선시키기 위해 MELP 보코더에는 다음의 5가지 특징이 추가되었다[2]. 1) Mixed Excitation, 2) Aperiodic Pulses, 3) Adaptive Spectral Enhancement, 4) Pulse Dispersion Filter, 5) Fourier Magnitude Modeling.

Mixed Excitation은 주기적인 펄스열 및 백색잡음 신호를 분석프레임의 유성음 강도에 따라 적절한 비율로 혼합하여 여기신호로 사용함으로써 기존의 LPC 보코더에서 나타나는 합성음의 울림 현상을 제거해준다. Aperiodic Pulses는 주로 음성 신호의 유성음과 무성음 사이의 천이영역(transition region)을 적절히 모델링하기

위해 사용되는데, 톤 성분을 제거하기 위해 분석된 피치의 25% 범위 내에서 피치를 변화시켜 천이구간에서 보다 자연스러운 합성음이 만들어진다. Adaptive Spectral Enhancement는 LPC 합성필터의 pole이 원음성의 포먼트 공명 특성과 비슷한 형태를 가지도록 하여 좀 더 자연스런 합성음을 만들어 낼 수 있게 해준다. Pulse Dispersion Filter는 삼각파의 저주파 통과 특성을 제거한 신호로부터 고정된 필터로 구현되는데, 여기신호의 에너지가 피치구간 내에서 빨리 감소하지 않고 확장되도록 만들어 줌으로써 합성음에서 생길 수 있는 거친 음질을 줄여준다. Fourier Magnitude Modeling은 LPC 잔차신호(residual signal)의 스펙트럼 특성을 모델링하는 것으로써 지각적으로 중요한 의미를 갖는 여기신호의 저주파 대역 정보를 이용하는 것이다[3].

2.1 MELP 인코더

입력 음성신호는 8kHz로 샘플링 되고, 먼저 60Hz 이하의 d.c 성분이 제거된 후 22.5ms 길이의 프레임 단위로 분석된다. 표 1은 2.4kbps MELP 코덱의 한 프레임 당 양자화 되어 할당된 비트수를 비트 스트림 순서로 나타낸 것이며, 그림 1은 MELP 인코더의 블록도이다. 입력신호는 1kHz의 LPF(Lowpass Filter)를 통과한 후, 0-500Hz, 500-1000Hz, 1-2kHz, 2-3kHz, 3-4kHz의 대역을 갖는 5개의 BPF(Bandpass Filter)를 통과시켜 대역별로 음성음 강도와 피치를 구하게 된다. 또한, 입력신호의 분석을 통해 10차의 LPC 계수와 잔차신호, 피치, 이득 값 등을 얻게 되고, 잔차신호에 대한 10개의 Fourier Magnitude를 구한 후 이들 변수를 표 1에 주어진 비트할당에 따라 양자화 하여 프레임 당 54 bit를 전송한다.

표 1. 한 프레임에 대한 비트 할당 (bits)

Parameters	Voiced	Unvoiced
Pitch Index	7	7
Aperiodic Flag	1	0
Bandpass Voicing	4	0
Gain(2 per frame)	8	8
LSF	25	25
Fourier Magnitude	8	0
Error Protection	0	13
Sync Bit	1	1
Total Bits/frame	54	25

2.2 MELP 디코더

각 모드(유/무성음, jittery 유서음)에 따라 피치합성한 후 IDFT 및 Random Noise Generator를 통해 혼합 여기신호를 구성한다. 이렇게 얻어진 여기신호는 1차의 Tilt와 10차의 pole/zero로 구성된 Adaptive Spectral Enhancement 필터를 거친 후 LPC 합성 및 이득조정을 통해 음성을 합성하게 되고, 65차의 Pulse Dispersion 필터를 거쳐 최종 합성음이 얻어진다. 그림 2는 MELP 디코더의 블록도를 나타낸 것이다.

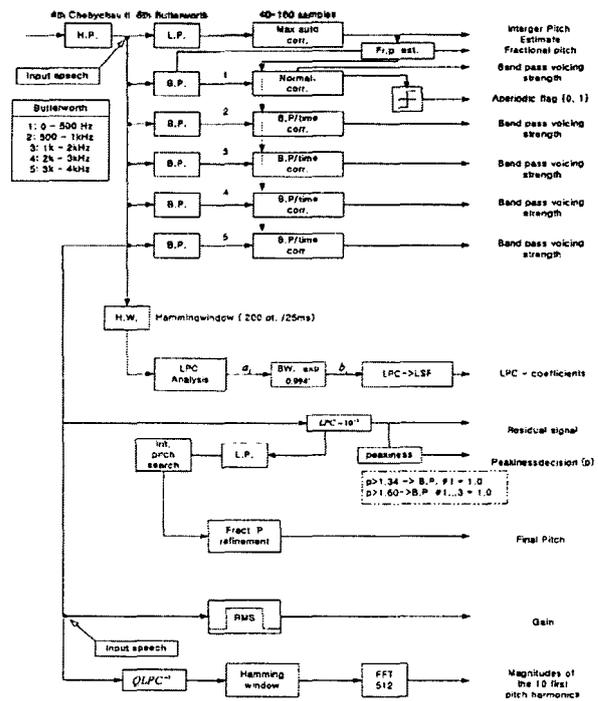


그림 1. MELP 인코더의 블록도

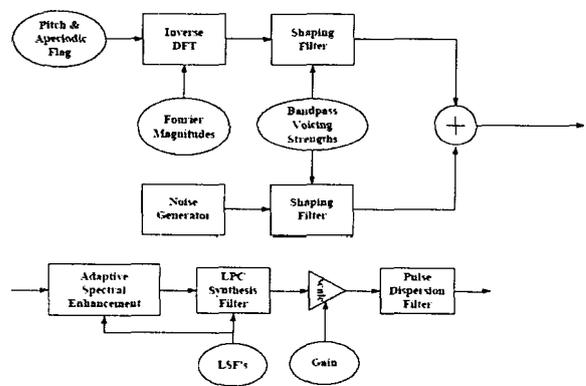


그림 2. MELP 디코더의 블록도

3. TMS320VC5510 DSK를 이용한 실시간 구현

3.1 TMS320VC5510 DSK 및 개발환경

TMS320VC5510은 TI사의 저전력 고성능 모델인 C55x 계열의 고정소수점 DSP로 200MHz의 Clock에서 최대 400MIPS의 성능을 가지는데, 기본적으로 17-bit x 17-bit Multiplier와 40-bit의 Adder, Non-pipelined Single-cycle MAC(Multiply accumulate) 등으로 구성되며, 내부에 64kbyte의 DARAM과 256kbyte의 SARAM을 가지고 있다. 본 연구에서 사용한 DSK는 입/출력 인터페이스를 위해 AIC23 Stereo Audio Codec을 장착하고 DMA (Direct Memory Access)를 지원한다. 또한, 보드상에 8Mbyte의 SDRAM과 512kbyte의 FLASH 메모리를 가지고 있는데, 추가적으로 16Mbyte의 외부 메모리 장착이 가능하다[4]. 시스템 개발을 위한 소프트웨어 툴로는 CCS (Code Composer Studio)[5] ver 2.12.07을 이용하여 프로그램의 Debugging 및 실시간 분석을 수행하였다.

3.2 실시간 처리를 위한 최적화 과정 및 실험결과

MELP 보코더의 실시간 구현은 DDVPC (Defense Digital Voice Processing Consortium)에서 제공한 ANSI C MELP 소스 코드를 기반으로 하여 수행하였다. TMS320VC5510 DSP에 포팅하기 위해 수행한 코드 수정 및 최적화 과정은 다음과 같다.

(1) 소스코드에서 32bit로 처리되는 read, write block을 16bit로 처리되게 하고, 60 Hz 이하의 d.c 성분을 제거하는 고역통과 필터 함수 부분의 정적선언 변수를 초기화하도록 수정하였다.

(2) 소스코드 분석과정을 통하여 실제 음성부호화 알고리즘과 관련 없는 부분을 찾아 프로그램 코드에서 제외하였다.

(3) MELP 코드가 포함하고 있는 ETSI의 기본 산술연산 함수를 TMS320VC5510 DSP에 가장 효율적으로 구성된 intrinsic 함수로 대체하였다.

(4) CCS에서 제공하는 Compile option을 최대한 활용하여 불필요한 코드의 사용을 최소화 하였으며, 간단한 함수를 inline 함수로 대체하였다[6].

(5) MELP 소스코드에 포함된 루프 연산 과정에서의 실행 Clock 수를 줄이기 위해 루프 언롤링(unrolling) 및 루프 인라인(inline)화를 적용하였고, loop trip counter가 작을 경우, 루프 내부의 연산을 직접 반복 하도록 하여 for 문의 사용을 최소화하였다.

위와 같은 최적화 과정을 통하여 구현된 MELP 음성 부호화 시스템의 성능을 평가하기 위해 프로그램의 실행속도는 음성부호화 과정이 수행되는데 필요한 Clock 수를 측정하였으며, 프로그램 및 데이터 메모리 측정은 COFF link 시 생성한 Memory map 파일을 참조하여 측정하였다. 표 2~ 는 MELP ANSI C 소스코드의 최적화 전과 최적화 단계별 결과를 보인 것이다. 복음구간을 비롯하여 유성음, 무성음 및 천이구간을 포함한 음성 데이터를 이용하여 인코딩 및 디코딩 과정에서 프레임 당 최대 소요되는 Clock 수를 측정하여 제시하였고, 수행시간은 식 (1)과 같이 계산하였다. 최적화 과정을 모두 마친 결과 22.5ms 한 프레임을 full-duplex 모드로 처리하는데 6.6ms가 소요되었다. 또한, 사용되는 프로그램 메모리와 데이터 메모리의 크기는 각각 46.5 kbyte, 57.9 kbyte로 나타났다

$$\text{프레임당 수행시간} = \text{최대 Clock 수} \times \frac{1}{200\text{MHz}} \quad (1)$$

표 2. 소스코드에서 불필요한 함수 제거 경우 [clock 수]

함수	원시 소스 코드	불필요한 함수 제거
Encoder	32229487	31325440
Decoder	12137859	11972352
Total	44367346 (221ms)	43297792 (216ms)

표 3. 최적화 단계별 결과 [clock 수]

함수	Intrinsic 함수 적용	컴파일러 최적화(-o3)	루프 최적화
Encoder	4315963	960447	960441
Decoder	2006988	371631	366090
Total	6322951 (31.6ms)	1332078 (6.6ms)	1326531 (6.6ms)

표 4. 최적화 전후의 최종 결과 [clock 수]

함수	원시 소스 코드	최적화
Encoder	32229487	960441
Decoder	12137859	366090
Total	44367346 (221ms)	1326531 (6.6ms)

실시간으로 구현된 MELP 보코더가 올바르게 동작하는지를 확인하기 위해 임의의 음성신호에 대해 PC에서 얻어진 ANSI C 소스코드의 출력과 DSP에 구현된 MELP 음성부호화기를 통해 얻어지는 출력신호를 비교하였다. 그림 3은 각각의 결과를 나타낸 것으로서 출력 값이 오차 없이 서로 동일한 결과를 가짐을 확인하였다. 또한 마이크 입력을 DSK 보드상의 AIC23 Stereo codec[7]을 통해 8kHz로 샘플링하고, 샘플링된 음성신호를 DSK 상에서 구현된 MELP 보코더의 encoding과 decoding 과정을 거쳐 스피커로 출력되도록 하여 음질의 왜곡 없이 실시간으로 구현됨을 확인하였다.

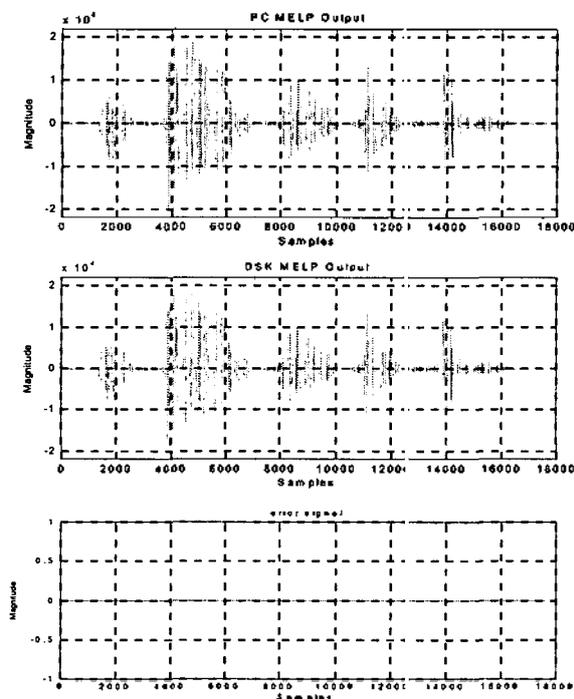


그림 3. PC 및 DSK 상에서 얻어진 출력신호 비교

4. 결론

본 연구에서는 2.4kbps MELP 보코더를 TI사의 고정소수점 DSP인 TMS320VC5510 기반의 DSK에서 실시간으로 구현하고, 그 결과를 제시하였다. 먼저 ANSI C 소스코드에서 불필요한 함수들을 제거하고, ETSI 기본 산술 연산 함수에 대해 Intrinsic 함수를 적용하였다. 그리고 CCS v2.1 컴파일러에서 수행 Clock 수에서 가장 최적화를 시키는 옵션(-o3)을 적용하였으며 루프에 대한 최적화를 수행하여 성능을 향상시켰다.

구현된 시스템은 프로그램 메모리 46.5kbyte와 데이터

메모리 57.9kbyte를 필요로 하며, 22.5ms의 한 프레임을 인코딩/디코딩 하는데 소요되는 Clock 수가 1326531로 200MHz의 DSP Clock rate를 고려할 때 약 6.6ms의 수행시간이 소요되었다. 따라서 full-duplex 모드로 최대 3채널을 수용할 수 있다. DSK상에서 구현된 MELP 음성부호화기의 출력과 DDVPC 에서 공개한 ANSI C 소스코드를 PC 상에서 구현한 결과를 비교하여 서로 일치함을 확인하였으며, DSK상에서 실시간 음성 입출력을 통하여 MELP 음성부호화기가 음질의 왜곡 없이 실시간으로 동작함을 확인하였다.

참고문헌

1. A.S. Spanias, "Speech coding : A Tutorial Review", *Proc. of IEEE*, Vol. 82, No. 10, pp.3071-3075, Oct. 1994
2. A. V. Mccree and T. P. Barnwell III, "A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding", *IEEE Trans. Speech and Audio Processing*, Vol.3, No.4, pp242-250, July 1995.
3. L.M. Supplee, R.P. Chon, J.S. Collura, and A.V. McGree, "MELP The New Federal standard at 2400 bps," *Proceeding of the IEEE ICASSP-97*, Vol.2, pp. 1591-1594, 1997
4. Texas Instruments Inc. *TMS320C55x DSP Functional Overview* 2000년 11월
5. Texas Instruments Inc. *Code Composer Studio Getting Started Guide* 2001년 11월
6. Texas Instruments Inc. *TMS320VC5510 Optimizing C/C++ Compiler User's Guide*
7. Texas Instruments Inc. *TLV320AIC23 Stereo Audio CODEC Data Manual* 2002년 4월