

TMS320C5409 를 이용한 G.723.1 음성 코덱의 실시간 구현

이동원, 손창용, 김지생, 조장형, 강상원
한양대학교 전자전기 제어계측공학과

Real-time implementation of the G.723.1 coder using TMS320C5409

Dong-Won Lee, Chang-Yong Son, Ji-Saeng Kim, Jang-Hyung Cho, Sang-Won Kang

Dept. of EECE, Hanyang Univ.

E-mail) swkang@selab.hanyang.ac.kr

요 약

본 논문에서는 국제 통신 표준화기구인 ITU-T에서 인터넷폰과 화상회의를 목적으로 채택된 G.723.1 음성 부호화 시스템을 TMS320C5409를 이용하여 전 과정을 어셈블리어로 실시간 구현하였다.

구현된 G.723.1 음성 부호화기는 6.3kbps 전송률일 때 인코더 25.75MIPS 이고 디코더 1.99 MIPS 의 최대 복잡도를 나타내고, 5.3kbps 전송률일 때 인코더 17.69MIPS 이고 디코더 1.9MIPS 의 최대 복잡도를 나타낸다. 사용된 메모리는 program ROM 11kwords, data ROM(table) 9.45kwords, RAM 2.8kwords 정도이며, 실시간 처리된 출력음성은 C simulation 결과와 같은 음질을 보였다.

구현된 G.723.1 음성 부호화기는 ITU-T에서 제공되는 17개의 테스트 벡터를 모두 bit-exact 하게 통과하였다.

1. 서 론

디지털 이동 통신 시스템에서 기존의 음성 전송 품질을 그대로 유지하면서 시스템의 용량을 증대시키기 위한 핵심 기술로서 음성 부호화기는 매우 중요하다. 기존의 통신 시스템에서는 주로 유/무선 전화를 사용한 통신에 관심을 가졌으나, 이제는 인터넷의 보급이 대부분의 사람들에게

게 확산되면서 인터넷과의 연결이나 화상회의 등 점차 적용분야가 확대되어가고 있다.

G.723.1은 5.3kbps와 6.3kbps의 이중 전송률을 갖는 구조로 현재 인터넷폰과 그 외의 멀티미디어 용 보코더로 사용되어지고 있으며, 낮은 전송률에 비해 우수한 음질을 제공하고 있다. G.723.1은 가변률(VAD) 동작시 평균 3.7kbps로 동작된다. 최적의 환경을 위해 두개의 전송률을 가지고 있기 때문에 다른 보코더 표준안들에 비해 응용성이 높다. G.723.1의 주된 응용분야는 H.324 표준을 포함한 멀티미디어 통신이다. 5.3kbps에서는 algebraic code-excited linear prediction(ACELP)을 6.3kbps에서는 multi-pulse maximum likelihood quantization(MP-MLQ) 방식을 사용하여 고정 코드북 검색을 수행한다. 30ms의 프레임을 사용하여 알고리즘 지연이 37.5ms 이고 잡음/비잡음 환경에서 MOS 3.98로 toll quality 음질을 나타내는 것으로 알려져 있다.

논문에서 사용한 TMS320C5409는 TI사의 고정 소수점 DSP로서 100MIPS의 성능을 가지고 있으며 40bit의 ALU와 dual operand 지원, 개선된 멀티버스 구조 등 기존의 16비트 고정 소수점 DSP보다 높은 성능과 편의성을 제공하고 있다.

본 논문에서 구현한 시스템은 마이크로 입력된 신호를 버퍼에 저장한 후 인코딩을 하여 출력된 비트스트림을 메모리에 저장하고, 이를 다시 매오

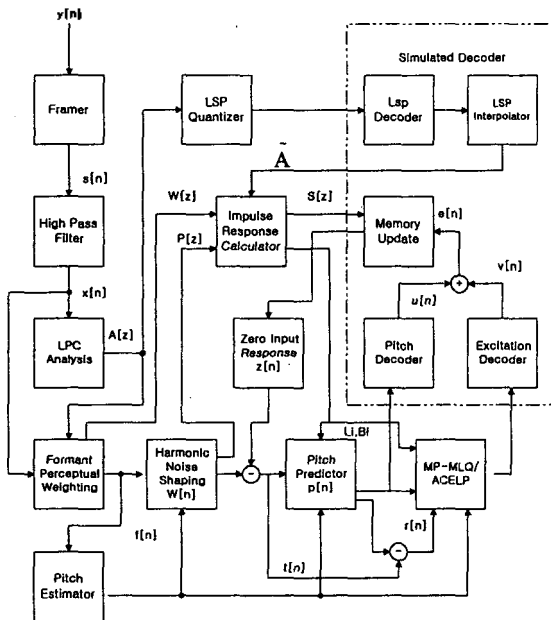
리에서 꺼내어서 디코딩을 하여 버퍼에 저장하고 다시 앞의 과정을 수행하면서 버퍼에 저장된 데이터를 출력하도록 되어 있다.

본 논문의 구성은 2장에서 G.723.1의 기본구조와 특성에 대하여 살펴보고, 3장에서는 실시간 구현에 관해 논하며, 마지막으로 4장에서 결론을 맺도록 하겠다.

2. G.723.1 음성 부호화기

2.1. 부호화기

G.723.1 부호화기는 원 신호와 합성된 신호와의 인지 가중 에러를 최소화하는 선형 예측 합성에 의한 분석 코딩이론에 기초를 두고 있다. 전체 부호화 과정의 기본 개념도는 <그림 1>과 같다.



<그림 1> G.723.1 부호화기 구조도

각 프레임은 먼저 DC 성분을 제거하기 위해서 고대역 통과 필터를 거치고 60샘플 크기의 4개의 부프레임으로 나뉜다. 모든 부프레임에 대해서 10차 LPC 필터가 계산되고 LSP 계수로 변환된다. LP 계수는 predictive split vector quantizer (PSVQ)를 사용하여 양자화되고 1차 AR 예측기가 사용된다. 양자화되지 않은 LSP 계수는 크기가 3,3,4인 3개의 부벡터로 나뉘며 각 부 벡터는 8비트 코드북을 이용해서 벡터 양자화된다. 부 벡터 단위로 에러를 최소화하는 적절한 코드북의 인덱스가 선택되고 채널을 통해 전송된다.

두개의 부프레임(120 샘플)마다 18에서 142 샘플의 범위에서 개구간 피치 주기 L_{OL} 이 가장 음성 신호를 사용해서 계산된다.

미리 계산된 예측 피치 주기를 사용하여 하모닉 노이즈 웨이핑 필터를 구성하며, LPC 합성필

터, 포먼트 인지 가중 필터와 하모닉 노이즈 웨이핑 필터의 결합은 임펄스 응답을 찾기 위해서 사용된다. 피치 주기 L_{OL} 과 임펄스 응답을 사용하여 폐구간 피치 예측값이 계산된다. 여기에는 5차 피치 예측기가 사용되며, 먼저 구해진 개구간 피치 주위에서 계산된다. 초기 대상 벡터에서부터 피치 예측값이 제거되고 피치 주기는 복호화기로 전송된다.

마지막으로 여기 신호의 비주기 성분이 근사화된다. 높은 전송률에 대해서 multi-pulse maximum likelihood quantization(MP-MLQ)가 사용되고 낮은 전송률에 대해서 algebraic-code-excitation(ACELP)가 사용된다.

높은 전송률에 대해서 잔여신호는 새로운 대상 벡터로서 MP-MLQ 블록으로 전송되고 양자화된다. 양자화 과정은 대상 벡터 $r[n]$ 을 $r'[n]$ 으로 근사화하는 것이다. 여기서 $r[n]$ 과 $r'[n]$ 의 에러 신호 $err[n]$ 의 제곱 평균값을 최소화하는 파라미터, G를 측정한다. 파라미터 측정과 양자화 과정은 합성에 의한 분석 방법에 기초한다. G_{max} 파라미터는 다음과 같이 측정되고 양자화된다.

우선 임펄스 응답 $h[n]$ 과 새로운 대상 벡터 $r[n]$ 사이의 교차상관 함수를 계산한다. 여기서 측정된 이득 G_{max} 는 logarithmic 양자화에 의해서 양자화된다. 양자화된 값, G_{max} 주위로 추가적인 이득값이 $[G_{max}-3.2, G_{max}+6.4]$ 의 범위내에서 선택된다. 이러한 이득값들 각각에 대해서 부호와 펄스의 위치는 연속적으로 최적화되고, 이 과정은 홀수와 짝수 grid에 대해서 반복된다. 마지막으로 $err[n]$ 의 최소 제곱 평균을 갖는 양자화된 파라미터의 조합이 선택되고 펄스 위치와 이득의 최적 조합이 전송된다.

낮은 전송률에 대해서 17 비트 대수 코드북이 고정 코드북 여기신호를 만들기 위해 사용된다. 각 고정 코드북은 보통 4개의 non-zero 펄스들로 구성되어 있으며 각각의 펄스들은 <표 1>에서와 같이 고정된 부호와 위치에 놓여있다.

<표 1> 대수 코드북의 부호와 위치

부호	위치
± 1	0,8,16,24,32,40,48,56
± 1	2,10,18,26,34,42,50,58
± 1	4,12,20,28,36,44,52,(60)
± 1	6,14,22,30,38,46,54,(62)

모든 펄스들의 위치는 동시에 한 칸 홀수 위치로 이동될 수 있는데, 이때 추가적인 비트가 필요하다. 각 펄스 위치는 3비트로 코드화 되고, 각 펄스의 부호는 1비트로 코드화 되어, 4개의 펄스에 대해 총 16비트가 필요하게 된다. 그리고, 마

지막 1비트는 이동여부에 대한 코드화에 사용되어 결과적으로 17비트가 사용된다.

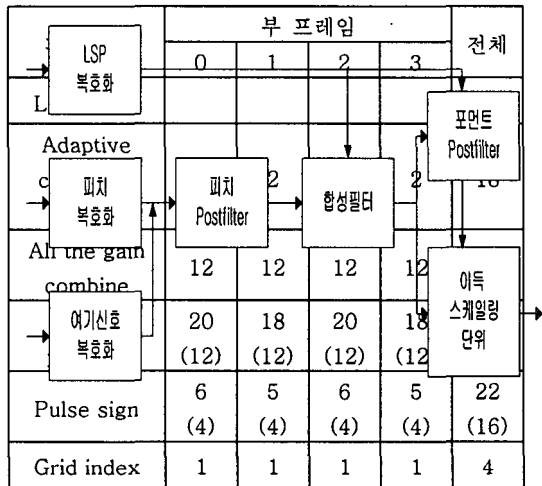
코드북은 가중된 음성 신호 $r[n]$ 과 가중된 합성 음성간의 제곱 평균 에러를 최소화함에 의해 선택된다.

G.723.1의 전송률에 따른 비트 할당을 <표 2>에 나타냈다. 괄호 안의 숫자는 5.3kbps 일 경우를 의미한다. 6.3kbps는 총 189비트, 5.3kbps는 총 158비트가 할당된다.

<표 2> 전송률에 따른 비트 할당

2.2. 복호화기

전체 복호화 과정의 기본 개념도는 <그림 2>와



같다.

<그림 2> G.723.1 복호화기 구조도

복호화기 역시 프레임 단위로 수행된다. 우선 양자화된 LPC 인덱스들이 복호화되고, LPC 합성 필터가 구성된다. 매 부프레임 마다 적응 코드북 및 고정 코드북 여기 신호가 복호화되어 합성된다. 적응 포스트 필터는 포먼트 포스트 필터와 순방향-역방향 피치 포스트 필터로 구성되어 있다. 여기 신호가 입력되는 피치 포스트 필터는 합성된 신호의 질을 높이기 위해 매 서브 프레임마다 수행된다. 이 때 최적의 피치 지연에 대한 신호 에너지와 교차 상관도에 의해 예측된 이득 값이 1.25dB 보다 작으면, 기여도는 무시되어 피치 포스트 필터는 사용되지 않는다. 필터가 사용될 경우에는 전송률에 따라 가중 인자가 최적 이득에 곱해져서 최종 스케일링 이득이 계산된다. 이후 합성 필터에서는 10차의 LPC 계수가 복호화된 피치 포스트 필터링 된 잔여 신호와 함께 음성신호를 합성하는데 사용된다. 이후 일반적으로 ARMA 포스트 필터로 포먼트 포스트 필터가 사용되고, 합성된 음성 벡터와 포스트 필터링된 벡

터로부터 포먼트 포스트 필터의 입력 수준으로 에너지를 유지시켜주는 이득 스케일링 단위가 계산된 후, 프레임 보간이 이루어진다. 본 음성 부호화기는 프레임 손실에 대해 강인하게 설계되어 프레임 손실 에러 은닉 방법이 복호화기에 포함되어 있다. 그러나 연속적인 에러에 대해서 설계되었기 때문에 불규칙적인 비트 에러에 대한 에러 수정은 불가능하다. 프레임 손실이 발생하면, 복호화기는 정규 복호화 모드에서 프레임 손실 은닉 모드로 변환된다. 프레임 보간 과정은 LSP 계수와 잔여 신호에 대해 독립적으로 수행된다.

3. 실시간 구현

DSP프로그램의 개발은 준비단계, 코딩 및 검증 단계, 시스템 연결 단계의 세가지 부분으로 구별되어 수행되었다. 준비 단계에서는 알고리즘을 분석하고 C 소스를 검토하여 인코더 및 디코더에 필요한 모듈을 파악하고 call tree를 작성하였다. 효율적인 구현을 위해서는 알고리즘의 분석이 반드시 선행되어야 한다. 준비 단계에서 가장 중요한 것은 프로그램의 형식, 각 모듈의 메모리의 활용, 지역 변수와 전역 변수의 지정 그리고 함수 호출 시 인자의 전달 방법을 미리 정의하는 것이다. G.723.1의 코딩은 LPC 분석 및 양자화, 피치 탐색, 고정 코드북 탐색, 이득값 양자화의 4가지로 구분 지을 수 있다. 각 모듈에 대하여 최적화가 이루어 졌는데, 특히 많은 복잡도를 요구하는 페퍼로 피치 탐색, 고정 코드북 탐색 등에 큰 비중을 두었다. 본 논문에서는 프로그램의 전 과정이 어셈블리어로 이루어 졌으며 프로그램된 각각의 모듈은 ITU-T에서 제공하는 테스트 벡터를 통하여 검증되었다. 마지막으로 시스템 연결 단계에서는 시뮬레이션상에서 검증된 프로그램을 실시간 동작할 수 있도록 하드웨어와 소프트웨어를 연결하는 작업이 수행된다. 이때 항상 우선 고려되는 부분이 음성의 입출력과 관련된 A/D 및 D/A로부터의 데이터 처리인데, 본 논문에서는 직렬 포트를 통한 더블 버퍼링을 사용하였다. 즉 현재 A 버퍼에 데이터가 차 있는 상태이고 B 버퍼는 비어있다고 가정하자. 이런 상황에서 인터럽트를 이용하는 경우, DSP의 CPU는 A 버퍼의 데이터에 대하여 음성 부호화작업을 수행하는 동시에 직렬 포트로 들어오는 1/8000초 마다의 인터럽트에 대하여 하던 일을 잠시 멈추고 직렬포트에 들어온 데이터를 B 버퍼로 이동시킨 후 다시 하던 일을 계속하게 된다. 따라서 B 버퍼가 가득차는 시간안에 인코딩 및 디코딩이 수행되어야 한다. 이러한 과정을 통해 구현된 G.723.1 코덱의 각 전송률별 복잡도를 <표 3>에 나타내었다. 앞서 언급하였듯이 DSP 어셈블리어 코딩시 이들 모듈에 대한 집

중적인 최적화 작업을 통해, G.723.1 코덱의 성능을 향상시키는데 중점을 두고 구현하였다. 구현된 G.723.1 코덱의 메모리 사용량을 <표 4>에 나타내었다.

<표 3> G.723.1 코덱의 전송률별 복잡도

전송률	Func.	MIPS(MAX)	MIPS(AVG)
5.3K	Encoder	17.69	14.68
	Decoder	1.9	1.73
	Total	19.59	16.41
6.3K	Encoder	25.75	19.42
	Decoder	1.99	1.74
	Total	27.74	21.16

<표 4> G.723.1 코덱의 메모리 사용량 (words)

Program		11k
Data	Static	1728*N
	Stack	1104
Table		9450

실시간 구현 결과는 C 시뮬레이션의 결과와 bit-exact해야 하고, 실시간으로 음성 부호화기가 동작될과 동시에 최소의 메모리를 사용하여야 한다. 실시간으로 음성 부호화기가 동작하기 위해서는 최대 복잡도가 DSP 칩의 처리용량을 초과하지 않아야 하며 메모리는 가능한 on-chip 메모리만으로 동작하는 것이 좋다.

4. 결론

본 논문은 TI사의 TMS320C5409 칩을 이용하여 G.723.1 음성 부호화기의 전 과정을 어셈블리로 실시간 구현하였다. 본 연구를 통해 G.723.1 음성 부호화기 및 복호화기가 실시간 동작하여 스피커로 출력됨을 확인하였고, ITU-T 에서 제공하는 테스트 벡터의 검증을 통해 C 시뮬레이션의 결과와 bit-exact 함을 확인 하였다. 또한 DSP 어셈블리어 코딩 과정에서 G.723.1 알고리즘에 대한 변경 없이 C 시뮬레이션에서의 같은 음질을 유지하면서, TMS320C5409의 구조적 장점과 명령어의 효율적인 사용을 통하여 최소의 복잡도를 가지도록 DSP 어셈블리어 코딩 과정에 중점을 두었다. 구현된 G.723.1 코덱은 최대의 계산량이

6.3kbps 모드일 때 인코더 25.75MIPS 디코더 1.99MIPS 로 최대 27.74MIPS 의 복잡도를, 5.3kbps 모드일 때에는 인코더 17.69MIPS 디코더 1.9MIPS로 최대 19.59MIPS를 나타낸다.

참고문헌

- [1] ITU-T Recommendation G.723.1, "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbits/s," March, 1996.
- [2] ITU-T Recommendation G.723.1, "Digital test vectors for the dual rate LBC G.723.1 speech coder," November 5, 1996.
- [3] Texas Instruments, "TMS320C54x User's Guide," Digital signal processing products, 1995.
- [4] Texas Instruments, "TMS320C54x Assembly Language Tool", Digital signal processing products, 1997.
- [5] Texas Instruments, "TMS320C54x Optimizing C Compiler", Digital signal processing Solutions, 1997.