

다채널 음성 감시 시스템 개발

A Multi-Channel Speech Surveillance System

김성수, 김용석, 조성호

Sung-Soo Kim, Yongseok Kim, Sung Ho Cho

한양대학교 전자공학과
(Phone) +82-345-400-5178
(FAX) +82-345-419-3042
shcho@casp.hanyang.ac.kr

요약

본 연구에서는 TMS320C31 DSP를 사용하여 다채널 음성신호를 실시간으로 압축/저장 및 재생 시스템을 구현하였다. 각각의 전화선으로부터 입력되는 음성 신호는 실시간으로 VSELP (Vector-Sum Excited Linear Predictive) 음성 부호화 알고리즘을 이용하여 8 Kbps의 전송률로 압축된다. 압축 음성 데이터는 host computer로 전송되어 각각의 채널별로 HDD나 MOD같은 저장 장치에 저장되어진다. PC Microsoft Windows95 환경에서 동작하는 호스트 프로그램은 압축 음성 데이터를 하드 디스크나 광자기 디스크에 저장하고, 저장된 압축 음성 데이터는 쉽게 검색하여 재생이 가능하다.

I. 서론

현재 기업, 은행, 증권회사, 통신판매회사 등에서 고객과의 전화상담 내용을 자동으로 녹음하거나 군이나 경찰, 정부기관 연구소 등의 비밀취급기관의 통화 감시 용 등의 다양한 용도로 다채널 음성 저장 및 검색 시스템에 대한 수요가 증가하고 있다. 이러한 다채널 음성 저장 및 검색 시스템은 다채널의 음성을 낮은 전송률로 압축을 할수 있어야 하고, toll-quality의 음성 품질을 제공하여야 하며, 저장/검색 및 재생을 편리하게 할 수 있어야 한다.

본 논문에서는 TMS320C31 DSP를 기반으로 하여 64 채널의 전화 음성 신호를 실시간으로 압축하고 저장 또는 재생할 수 있는 시스템을 구현하였다 [1]. 각각의 전화선을 통하여 입력되는 음성 신호들은 VSELP 음성 부호화 알고리즘을 이용하여 8 Kbps의 전송률로 실시간

압축된다 [2][3]. 압축된 음성 데이터는 채널 정보, 상태 정보 등을 가지고 있는 헤더가 붙여지게 되고, 다른 채널의 압축 음성 데이터와 다중화 된다. 이 다중화 된 데이터들은 PC로 전송되어 HDD나 MOD같은 저장 장치에 저장되어진다.

구현된 시스템은 최저 8 채널에서 최대 64 채널의 음성 신호를 실시간으로 압축 및 저장이 가능하다. 수시로 저장된 음성 데이터를 재생하기 위해 6GB의 HDD를 탑재하였고, 2개의 2.6GB MOD를 가지고 있어서 많은 양의 압축 음성 데이터를 편리하게 저장할 수 있다. 구현된 시스템은 음질 저하 없이 많은 양의 데이터를 저장할 수 있도록 디자인 되었다. HDD와 MOD는 각각 1,440시간과 720시간의 음성 데이터를 저장할 수 있어 총 2,880시간의 음성 신호를 연속해서 저장할 수 있다.

구현된 시스템의 전체 블록도를 그림 1에 나타내었다. 8 채널의 음성 신호를 동시에 실시간으로 VSELP 음성 부호화를 위해 8 채널 DSP 보드를 제작하였다. 이 8 채널 DSP 보드는 8개의 1 채널 DSP 모듈로 구성되어 있고, 각각의 1 채널 DSP 모듈은 TMS320C31, SRAM과 EPROM 같은 메모리 소자, AGC (Automatic Gain Control) 회로, hook-off 검출 회로, A/D 변환기들로 구성되었다. 앞에서도 언급했지만 구현된 시스템은 총 8장의 8 채널 DSP 보드를 지원할 수 있다.

8 채널 DSP 보드에서 출력되는 압축 음성 데이터들은 MUX 보드에 의해서 다중화 되어진다. 이 MUX 보드는 FPGA (Field Programmable Gate Array) 와 버퍼로 구성되어 있다. FPGA (Altera FLEX EPF10K50GC403-3) 는 VHDL (Very high speed Hardware Description Language) 를 이용하여 어드레스 디코딩, 채널 ON/OFF 제어, 버퍼 제어 등의 기능을 수행할 수 있다 [4]-[8].

Microsoft Windows95 환경에서 동작하는 호스트 프로그램을 이용하여 전체 시스템의 관리, 저장된 음성 데이터 검색, 압축 음성 데이터의 재생 등을 편리하게 할 수 있다.

본 논문의 구성은 다음과 같다. 다음 II절에서 음성

압축 알고리즘으로 사용한 VSELP 음성 부호화기 알고리즘의 실시간 구현에 대해 언급 하였다. III 절에서는 구현된 시스템의 구성에 대해 언급하고 끝으로 IV 절에서 결론을 맺는다.

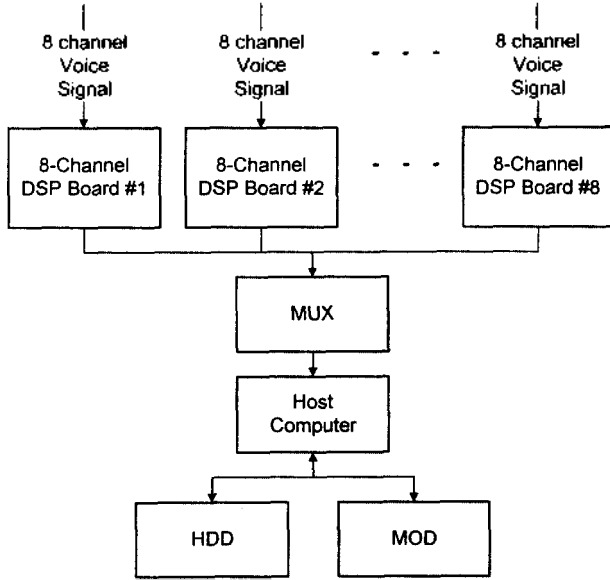


그림 1. 전체 시스템 블록도

II. 음성 부호화 알고리즘의 실시간 구현

본 연구에서는 음성 신호의 압축 알고리즘으로 VSELP 음성 부호화기를 이용하였다. 8 Kbps VSELP 음성 부호화기 [2] 는 미국의 TIA (Telecommunications Industry Association)와 EIA (Electronic Industries Association) 에 의해 북미 디지털 셀룰러 전화의 공식 규격으로 채택된 바 있다. 또한 6.7 Kbps VSELP 음성 부호화기가 일본 디지털 셀룰러 표준으로 사용되고 있다 [3]. VSELP 방식은 CELP (Code Excited Linear Prediction) 음성 부호화기의 변형으로 8Kbps 의 비교적 낮은 전송률 (bit rate)에서 양호한 음질을 유지하며 CELP 부호화기가 지닌 코드북 탐색에 많은 계산량을 요구하는 단점을 보완한 독특한 구조의 코드북을 사용 및 40 개 7조의 기저 벡터(basis vector)를 포함하는 여기 코드북(excitation codebook)을 이용하므로 채널 에러에 강한 특성을 갖는다. 음성 처리는 프레임(frame) 단위 혹은 프레임을 4 개로 나눈 서브 프레임(subframe) 단위로 행해진다.

코드북을 위한 매개 변수(parameter)들은 매 서브프레임마다 구해지고, 반사계수와 프레임 에너지는 매 프레임마다 구해진다. 모든 매개 변수들은 매 프레임마다 8Kbps 의 전송속도로 부호화되어 전송된다. 8 Kbps VSELP 에서 사용하는 기본 매개변수와 비트할당(bit allocation) 정보를 요약하면 표 1 과 같다.

VSELP 알고리즘은 매 서브프레임마다 Lag search 를 수행한다. 그러나 연산량을 줄이기 위해서 1 번째와 3 번째 서브프레임만 full Lag search 를 수행하고 2 번째와 4 번째 서브프레임은 1,3 번째 서브프레임에서 구해진 Lag 의 주변만 탐색하는 알고리즘을 적용하였다.

표 1. 8Kbps VSELP 부호화기의 비트 할당

Parameter	Bits / 5 msec	Bits / 20 msec
10 LPC coefficients	-	38
Average speech energy	-	5
Excitation codewords from two VSELP codebooks	14	56
Lag of pitch filter	7	28
Gain parameters	8	32
unused	-	1
Total	29	160

VSELP 음성 부호화기의 실시간 동작을 위해 우선 VSELP 음성 부호화기의 부호화 알고리즘을 C 언어로 프로그램하여 동작을 확인하였다. 실시간 동작을 위해서는 어셈블리 언어로 코딩하는 것이 필수적이다. 이를 위해서 일단 C 언어 상에서 어셈블리 언어의 특징을 염두에 두고 일차적인 최적화를 하였다. 그리고 각 모듈별로 필요한 처리 시간이 큰 것부터 어셈블리 언어로 다시 프로그램 하였다. 효과적인 어셈블리 코딩을 위해서는 TMS320C3x 의 구조를 파악하는 것이 반드시 필요하다. 칩 구조의 특징을 효과적으로 이용하기 위해서 어셈블리 언어로 코딩할 때 가장 주안점을 둔 것은 1) 병렬 처리문의 사용, 2) 내부 메모리 사용의 극대화, 3) 파이프라인 컨플릭트 최소화, 4) 효과적인 loop 구성, 5) 사용 메모리 공간 절약 등이다 [1], [9]-[11].

병렬 처리문은 한 머신 사이클에 두 번의 연산이 가능하므로 처리 시간 절감에 가장 큰 비중을 차지한다. 그러나 병렬 처리문 사용할 때는 TMS320C31 의 레지스터 사용에 제약이 있으므로 병렬 처리문을 잘 사용하려면 레지스터의 효율적인 사용이 필요하다. 본 연구에서 사용한 TMS320C31 은 2Kword 의 내부 메모리 공간이 있다. 내부 메모리를 사용하면 외부 메모리 사용할 때 보다 access time 이 줄어들어 실시간 동작에 유용하다. 그리고 연산량을 줄이기 위해서 자주 사용되는 table 이나 변수들을 내부 메모리에 위치시키면 연산 시간을 많이 줄일 수 있다. TMS320C3x 는 명령어를 고속으로 처리하기 위해 파이프라인 구조로 되어 있다. 파이프라인이 제대로 활성화 되면 1 머신 사이클의 실행속도로 프로그램을 실행하게 된다. 파이프라인의 상태가 여러가지 이유로 1 명령 1 머신 사이클이 불가능 해진다. 이와 같은 현상을 것을 파이프라인의 컨플릭트(conflict)라 부르며 분기 컨플릭트, 레지스터 컨플릭트, 메모리 컨플릭트 등이 있다. 메모리와 레지스터 컨플릭트는 코딩시에 미연에 방지하기가 사실상 어렵다. 그러나 분기 컨플릭트는 어셈블리 언어로 코딩시에 조금만 신경을 써주면 상당히 예방할 수 있다. 지연 분기 명령은 분기 명령 실행 후에 후속의 3 가지 명령을 폐차하고, 파이프라인을 만족하면서 분기선의 어드레스로 새로운 명령을 폐차한다. 따라서 지연분기의 경우는 어디까지나 1 명령 1 머신 사이클의 실행속도가 유지된다. 지연 분기 명령의 삽입 장소는 실제의 분기 포인트보다 3 명령 앞이 되는 지점이다. TMS320C31 은 하드웨어 루프를 제어할 수 있다. CPU 레지스터의 RS (리퍼트 개시 레지스터), RE (리퍼트 종료 레지스터), RC (리퍼트 카운터)에 각각 파라미터를 세트하고 블록 리퍼트 명령 (RPTB)을 실행함으로써 오버헤드없이 명령 블록을 반복할 수 있게 된다. 단일 리퍼트 명령(RPTS)은 하나의 단일 명령이나 하나의 병렬 명령을 반복하는 경우에 사용한다. 다음의 소스는 VSELP 음성 부호화 알고리즘을 어셈블리 언어로 코딩한 것 중 일부분이다

다채널 음성 감시 시스템 개발

```

;
;   for (i = 0; i < 170; i++, s++, d++)
;       *d = *s;
;
LDF    *AR6++,R0
RPTS  168
LDF    *AR6++, R0           ;R0 = *s++
STF    R0, *AR7++         ;*sd++ = R0
STF    R0, *AR7++
    
```

소스 윗 부분의 C 루틴을 RPTS 와 병렬 처리문을 이용하여 구현한 것이다. 위와 같은 기법들을 적절히 사용하면 TMS320C31 의 아키텍처에서 고속으로 실행되는 프로그램을 구현할 수 있다.

구현된 VSELP 음성 부호화 알고리즘의 연산량은 약 19 MIPS 정도로, 본 연구에 사용된 25 MIPS TMS320C31 최대 연산량의 약 76% 정도이다. 그림 2에 어셈블리어로 구현된 VSELP 음성 부호화 알고리즘의 각 기능별 연산량을 나타내었다.

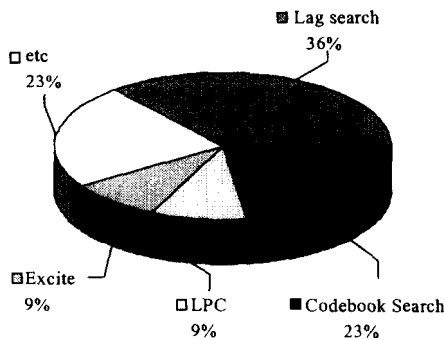


그림 2. 구현된 VSELP 음성 부호화기의 기능별 연산량

III. 시스템 구성

8 채널 DSP 보드는 1 채널 DSP 모듈 8 개로 구성되어 있고, 8 채널의 음성 신호를 실시간으로 압축하기 위해서 설계되었다. 1 채널 DSP 회로의 구성은 그림 3 과 같다.

hook-off detector 는 전화선의 상태를 감시하여 통화가 이루어지면 TMS320C31 에 인터럽트 신호를 보내게 된다. TMS320C31 이 인터럽트 신호를 받으면 A/D 변환기를 동작시켜서 A/D 변환을 시작 시키고 버퍼의 상태를 검사하게 된다. 입력되는 음성 신호는 AGC 회로와 저역 통과 필터를 거쳐 A/D 변환기를 기친 다음 디지털 신호로 바뀌어 입력 buffer 에 저장되어진다. TMS320C31 은 입력 버퍼의 상태를 감시하고 있다가 버퍼에 일정한 양의 데이터가 차게 되면 저장되어진 디지털 음성 데이터를 읽어와 VSELP 음성 부호화 알고리즘을 이용해 부호화하고 출력 버퍼에 저장한다. hook-off 검출기가 통화의 끝을 검출하면 다시 TMS320C31 에 인터럽트 신호를 보내어 통화의 끝을 알린다. 그러면 TMS320C31 은 A/D converter 를 중지시키고 입력 buffer 에 있는 디지털 음성

데이터들을 모두 읽어서 마지막 부호화를 하고 출력 buffer 에 데이터를 저장한다. VSELP 음성 부호화 알고리즘의 전송속도는 8Kbps, 즉 한 프레임(20ms)당 출력되는 데이터양은 사용되지 않는 1 bit 을 포함하여 160 bit 이다. 이를 8 bit 단위로 나누고 여기에다가 각 채널의 정보등을 나타내는 8 bit 의 데이터가 첨가되어 인데 여기에다가 각 채널의 상태를 나타내는 8bit 의 데이터가 첨가되어 16 bit 단위로 1 채널에서 출력된다. 8 bit 의 정보 데이터를 추가하는 이유는 64 채널의 부호화된 음성 데이터가 MUX 보드에 의해서 PC 로 전송될 때 각 채널의 데이터를 구분하기 위해서이다. PC 에서 동작하는 host program 은 부호화된 음성 데이터를 하드 디스크나 광자기 디스크에 저장할 때 정보 데이터를 보고 데이터 배이스화 하여 저장하기 때문에 실제로 저장할 때는 부호화된 음성 데이터만 하게 된다. 이러한 1 채널 DSP 블록으로 8 채널 DSP 보드를 구성하였다. 8 채널 DSP 보드는 8 채널의 음성 데이터를 실시간으로 VSELP 음성 부호화를 할 수 있고, 8 채널 DSP 보드를 최대 8 장, 즉 64 채널까지 확장이 가능하다. 64 채널의 부호화된 음성 데이터들은 PC interface 형태로 제작된 MUX 보드의 FPGA 에 의해 읽혀지고, 16 bit 단위로 PC 에 전송이 된 후, PC 의 하드 디스크(HDD)나 광자기 디스크(MOD)에 저장되어진다.

MUX 보드의 구성은 그림 4 와 같다. MUX 보드는 FPGA 와 buffer 로 구성되어 있다. FPGA 의 Address Decoder 부분은 PC 에서 전송되는 어드레스 신호를 가지고 각 블록, 즉 Channel ON/OFF Control 블록과 Buffer Control 블록 중 해당되는 블록에 신호를 보내준다. Channel ON/OFF Control 블록은 PC 로부터 받은 데이터를 가지고 64 채널의 DSP 를 각각 ON/OFF 할 수 있도록 디자인 되어 있다. Buffer Control 블록은 8 채널 DSP 보드의 출력 버퍼들의 상태를 조사한다. 만약 어떠한 채널의 출력 buffer 에 부호화된 음성 데이터가 저장되면 2Kword(1word = 16 bit) 단위로 읽어서 MUX 보드에 있는 buffer 에 저장한다. 이렇게 해서 64 채널의 부호 음성 데이터를 읽어와서 MUX 보드의 buffer 에 16bit 단위로 저장한다. 만약 MUX 보드의 buffer 가 차면 읽어오는 일을 멈추고 PC 에 인터럽트 신호를 보낸다. PC 에서 수행되는 host program 은 해당하는 interrupt 가 들어오면 MUX 보드의 buffer 로부터 부호 음성 데이터를 모두 읽어와서 HDD 나 MOD 에 저장한다. PC 가 데이터 읽기를 끝내면 다시 8 장의 8 채널 DSP 보드의 buffer 에서 부호 음성 데이터를 읽고 MUX 보드의 buffer 에 저장한다. MUX 보드의 buffer 의 크기는 16 Kword (1 word = 16 bit) 이다.

Microsoft Windows95 환경에서 수행되는 host program 은 MUX 보드로부터 인터럽트 신호를 받으면 MUX 보드의 버퍼에서 부호화된 음성 데이터를 읽어 온다. 8 bit 의 헤더를 이용하여 채널별로 구분하여 하드 디스크나 광자기 디스크에 저장한다. VSELP 복호 알고리즘을 내장하여 하드디스크나 광자기 디스크에 저장되어 있는 부호화 음성 데이터들을 다시 복호하여 재생할 수 있다. 또한 사용자의 레벨에 따라 제공하는 기능이 다르게 하여 시스템 보안을 한층 강화 하였고, 네트워크 관리기능이 있어 네트워크상의 다른 PC 에서 다채널 음성 압축/저장 및 검색 시스템을 관리할 수 있다.

IV. 결론 및 추후과제

본 연구에서는 전화선상의 음성 신호를 최대 64 채널까지 동시에 실시간으로 압축, 저장하고 재생할 수 있는 시스템을 구현하였다. VSELP 음성 부호화 알고리즘을 이용하여 음성 신호를 실시간으로 압축하였다. TMS320C31을 탑재한 8 채널 DSP 보드를 설계하였다. 또한 64 채널의 압축 음성 데이터를 PC로 전송하기 위해 FPGA를 내장한 MUX 보드를 설계하였다. 구현된 시스템은 음질저하 없이 최대한의 음성 데이터를 저장하도록 설계되었다.

현재 1개의 TMS320C6201 DSP를 이용하여 CS-ACELP 음성 부호화 알고리즘으로 16 채널의 음성 신호를 실시간 압축할 수 있는 시스템이 연구되고 있다.

참고문헌

- [1] Texas Instruments, *TMS320C3x User's Guide*, 1993.
- [2] I. A. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at 8Kbps," *Proc. Int. Conf. on Acoust., Speech, and Signal Processing*, pp.461-464, Apr. 1990.
- [3] I. A. Gerson, "Vector sum excited linear prediction (VSELP) speech coding for Japan digital cellular," *IEICE*, pp. 35-40, November 1990.
- [4] D. L. Perry, *VHDL 2nd Ed.*, R. R. Donnelley & Sons Company, 1993.
- [5] J. R. Armstrong and F. G. Gray, *Structured logic design with VHDL*, Prentice-Hall, 1993.
- [6] Zainalabedin Navabi, *VHDL - Analysis and modeling of digital systems*, McGraw-Hill, 1993.
- [7] Pran Kurup, Taher Abbasi, *Logic synthesis using synopsis*, Kluwer Academic Publishers, 1995.
- [8] David W. Knapp, *Behavioral synthesis - digital system design using the synopsis behavioral compiler*, Prentice-Hall, 1996.
- [9] R. Chassaing, *Digital signal processing with C and the TMS320C30*, John Wiley & Sons, Inc, 1992.
- [10] M. H. Sunwoo and S. I. Park, "Real-time implementation of the VSELP on a 16-bit DSP chip," *Speech Coding Workshop*, Whisler, Canada, 1991.
- [11] Motorola Inc., *Principles of vector-sum excited linear predictive (VSELP) speech coder and its implementation on the DSP56156*, Austin, Texas, 1991.

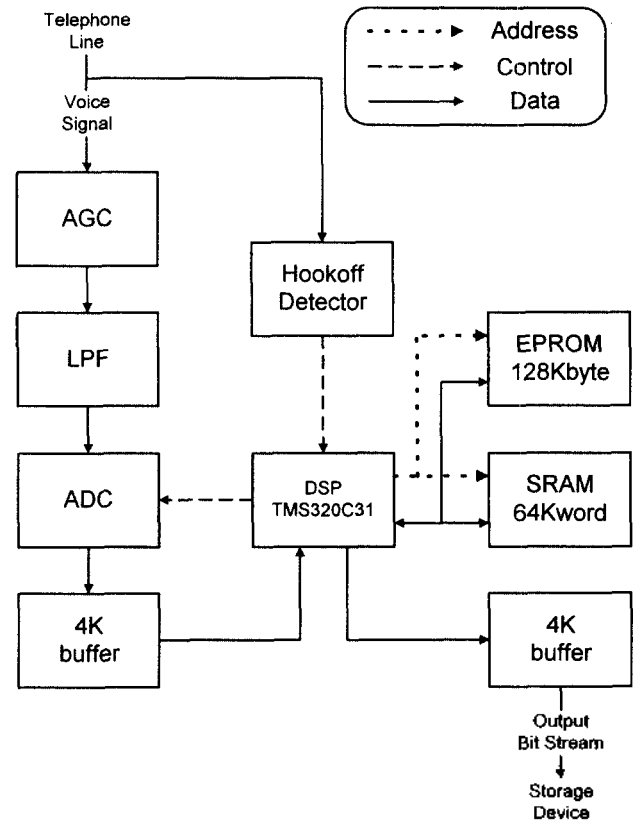


그림 3. 1 채널 DSP 모듈의 구성도

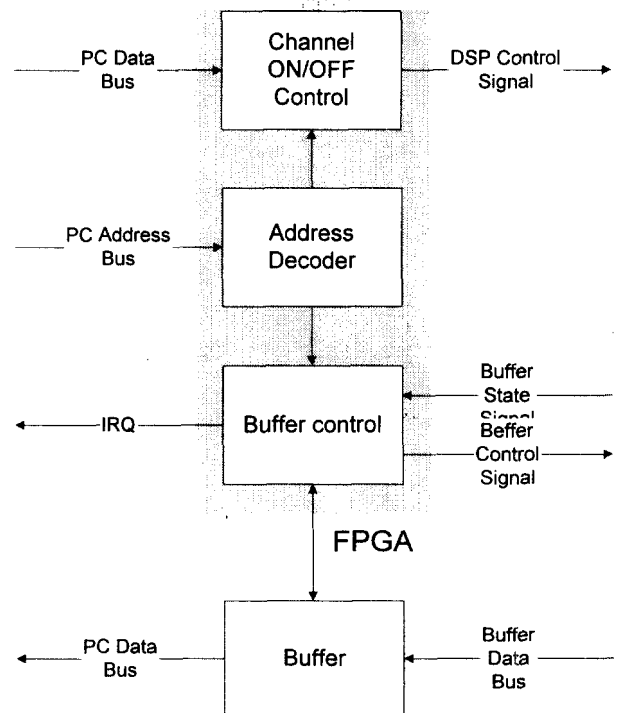


그림 4. MUX 보드의 기능별 블록도