

# TMS DSP 칩을 이용한 음성 특징 벡터 추출기 설계

예병대, \*이광명, \*성광수

(주)디엠비테크놀로지, \*영남대학교 전자 정보 공학부

전화 : 042-828-7110, 핸드폰 : 016-560-9401

## A Design of Speech Feature Vector Extractor using TMS320C31 DSP Chip

Byoung-Dae Ye, Kwang-Myoung Lee, and Kwang-Soo Sung

DMB Technology CO.,LTD.

E-mail : ybdae@dmbtech.com

### Abstract

In this paper, we proposed speech feature vector extractor for embedded system using TMS 320C31 DSP chip. For this extractor, we used algorithm using cepstrum coefficient based on LPC(Linear Predictive Coding) that is reliable algorithm to be is widely used for speech recognition. This system extract the speech feature vector in real time, so is used the mobile system, such as cellular phones, PDA, electronic note, and so on, implemented speech recognition.

### 1. 서론

우리가 일상생활에서 가장 보편적이고, 편리한 정보 수단을 얘기한다면, 그건 아마 음성일 것이다. 이 음성을 통해 기계와 대화를 한다면, 즉 기계 및 사용 장치에 인식시켜서 동작을 하게 되면, 더욱더 편리하게 이용할 수 있을 것이다. 최근 음성인식의 분야는 각종 가전 제품, 자동차등 여러 제어 분야에 활발히 연구되어지고 있다. [1][2] 특히 최근 PC와 같은 고정된 환경이 아닌 OS 기능을 내장한 셀룰러폰, PDA, 전자수첩, 임베디드 시스템 등과 같은 휴대용 시스템의 음성 인식 기술이 활발하게 연구되어지고 있다.

음성 인식을 크게 두 부분으로 나누어진다. 먼저 음성의 특징 벡터를 추출하는 부분이다. 다음은 추출된

특징 벡터를 이용하여 음성을 인식하는 부분이다. 사실 특징 벡터 추출은 이상적인 정답이 없기 때문에 음성인식을 위한 특징의 좋고 나쁨은 음성인식률로 판단된다. 따라서 정확한 음성 특징 벡터 추출만이 음성인식의 성능을 좌우하는 것이다.

이에 본 논문에서는 임베디드 시스템을 위한 높은 인식률과 실시간 음성 인식을 위한 특징 벡터 추출기를 설계하고자 한다.

본 논문에서 같은 여타의 음성 특징 계수보다 신뢰성이 높고 음성 인식에서 가장 널리 사용되고 있는 LPC(Linear Predictive Coding)에 기반을 둔 Cepstrum 계수를 이용한 알고리즘을 이용하였다. [3][4][5][6]

이러한 LPC cepstrum 계수를 계산하는 알고리즘을 하드웨어로 구현하기 위해 IT사의 부동 소수점 범용 DSP인 TMS320C31를 사용하였다.

### 2. 음성 특징 추출 알고리즘[3]

LPC 계수를 계산하는 방법에는 자기 상관법(Auto correlation)과 covariance 방법이 있는데, 자기 상관법이 더욱 안정된 성능을 제공하기 때문에 음성 인식에서는 일반적으로 사용된다. [5][6]

LPC cepstrum 프로세서는 한 프레임(200, 샘플, 20ms)의 음성 표본치  $s(n)$ 으로부터 5단계의 연산과정을 거쳐 cepstrum 계수  $C_n$ 을 추출하게된다. 점선으로 표시된 부분은 자기 상관법을 이용한 LPC 계수 연산

절차의 일반적인 다이어그램이다.

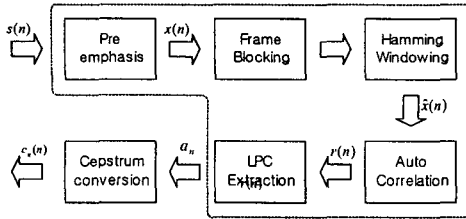


Fig. 1. The procedure of calculating the LPC cepstrum.

LPC cepstrum을 계산하는 과정을 살펴보면 먼저 입력된 음성 표본지  $s(n)$ 에 대해 전치 증폭기(pre emphasis)를 통해 실음성 구간에서 자음의 특징을 확실하게 해준다.

자음에 해당하는 스펙트럼은 고주파 성분을 포함하고 모음에 해당하는 스펙트럼은 저주파 성분을 포함하고 있다. 하지만 음성의 경우 고주파 대역의 에너지는 저주파 대역의 에너지에 비하여 상당히 작다. 따라서 그냥 음성 샘플로 인식을 수행한다면 무성음에 의하여 왜곡되는 값은 전체 왜곡 값에 비해 상당히 작기 때문에 영향을 줄 수 있다. 음성신호의 DC성분을 제거할 수 있을 뿐만 아니라 고주파 성분을 강조하기 위해 식 (1)과 같이 1차 고대역 필터를 하게 된다. 여기서  $a$ 는 필터계수로 0.95-0.98 정도의 값을 가진다. 또한  $s(n)$ 는 입력 음성 샘플이다.

$$x(n) = s(n) - as(n-1) \quad (1)$$

전치증폭기를 통과한 신호는 20 ~ 30msec 단위로 음성신호를 블록화 하여 각 frame 별로 특징 벡터를 계산하게 되는데, 이렇게 음성 샘플을 200 샘플의 분석구간과 100 샘플의 중첩구간으로 나누는 과정을 frame blocking이라 한다.

이렇게 나누어지는 frame은 처음과 끝에서 불연속 부분이 발생하는데, 이러한 불연속성을 보정하기 위해, 식 (2)과 같이 해밍 윈도우를 이용한다. 여기서  $w(n)$ 은 해밍 윈도우의 계수이다.

$$\hat{x}(n) = x(n)w(n) \quad n=0, 1, \dots, 100 \quad (2)$$

음성신호는 모음의 경우 준 주기적인 성질이 있으므로 분석 구간 내에 주기적으로 2~3개의 기본주파수가 나타나며 무성음은 최대값이 상대적으로 작고 비주기적이다. 유성음의 파형을 살펴보면 대체로 비슷한

파형이 반복되어지는 것을 볼 수 있는데 이 준 주기적인 파형의 피치라고 한다. 이것은 음성의 기본 주파수를 의미한다.

이러한 피치는 음성의 주기적 특성을 나타내는 특성으로 음성으로부터 여러 가지 의미를 추출해 내는 데 중요한 요소로 작용하고 있다. 이 알고리즘에서는 Auto correlation을 이용하여 식 (3)과 같이 11개의 자기 상관계수  $r(m)$ 을 구함으로써 피치 검출을 하게 된다.

$$r(m) = \sum_{n=0}^{100-m} \hat{x}(n)\hat{x}(n+m), \quad m=0, \dots, 99 \quad (3)$$

LPC 알고리즘에서는 음원을 백색잡음(무성음의 경우) 또는 주기적 펄스열(유성음의 경우)로 가정하고 음성신호로부터 이에 따른 특성함수를 구하여 LPC 계수를 구한다.

위의 알고리즘 각 단계에서 사용된 연산자를 덧셈과 곱셈으로 나누어 각각의 연산 횟수를 계산해보면 대부분이 자기 상관함수 연산 단계라는 것을 알 수 있다. 그러므로 전체 연산과정을 빠르게 수행하기 위해서는 프로세서가 자기 상관 연산을 빠르게 처리 할 수 있어야 한다. 특히 소수점 곱셈 연산이 많은 것을 고려하면 floating 곱셈 연산을 잘 수행 할 수 있는 DSP 전용 프로세서를 이용하여야 할 것이다.

### 3. 음성 특징 벡터 추출을 위한 하드웨어 구현

#### 3.1 TMS320C31 DSP

Texas Instrument사의 TMS320C31 DSP 칩은 실시간 디지털 신호 처리를 위해 32bit의 floating-point 연산 방식 곱셈기를 내장하고 있다. 또한 하버드 구조에 의한 3라인 파이프라인 구조를 가지고 명령어의 패치, 디코더, 실행을 동시에 처리하므로써 명령의 처리 및 연산에서 최대한의 효율을 가진다. CPU는 60MHz에서 30MIPS 속도의 처리 능력을 가졌다.

TMS320C31 DSP 칩의 내부구조를 보면 Floating-point 연산을 위한 40bit 확장 정밀 레지스터와 I/O와 CPU의 동작을 동시에 수행하기 위한 DMA 제어기, 2개의 32bit 타이머 등이 존재한다.

TMS320C31의 전체 메모리 영역은 16M X 32bit 워드에 이르며 프로그램, 데이터 I/O영역을 포함하는 memory-mapped I/O로 되어 있다.

앞에서 소개된 TMS320C31 DSP 칩은 많은 곱셈 연산을 수행하는 음성 특징 벡터를 추출하기 위한 하

드웨어 구현에 아주 적합함을 알 수 있다.

### 3.2 설계된 특징 벡터 추출기의 구성

그림 2는 TMS320C31을 이용하여 설계된 음성 특징 벡터 추출기의 블럭다이어그램을 나타내고 있다.

마이크를 통해 입력된 음성 신호는 ADC를 거쳐서 128K Word SRAM에 저장된다. 32K Byte ROM에 저장된 벡터 추출 프로그램에 의해, 디지털 신호로 변환된 음성 신호는 연산되어 다시 SRAM에 추출 벡터를 저장하게 된다. 모든 처리는 입력 스위치로 제어되어지며, 수행 과정은 LED 및 LCD에 의해 표시되도록 하였다. TMS320C31 DSP 칩에서 전송되는 어드레스와 각 명령을 디코딩하여 각 주변 장치로 연결하기 위해 디코드 블록을 VHDL로 직접 설계하여 FPGA로 구현하였다. 또한 수행된 결과 데이터를 PC로 전송하기 위해 직렬 통신 블록을 설계하였다.

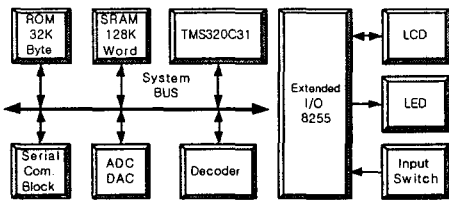


Fig 2. Speech Feature Vector Extractor using TMS320C31 DSP Chip

먼저 설계된 보드와 모든 메모리를 초기화 한 후, 앞에서 소개된 알고리즘에 의해 LPC cepstrum 계수를 추출하게 된다. 또한 특징 벡터 추출 시간을 측정하기 위해 시간 출력이 가능하도록 프로그램 되어 졌다.

## 4. 실험 및 결과

설계된 벡터 추출기의 성능을 측정하기 위해 그림 3과 같은 음성 신호를 사용하였다. 이 음성 신호는 '가운데'라는 음성이다.

사용된 음성 샘플의 프레임 에너지는 그림 4와 같이 무음 부분과 실제 입력 음성 부분으로 명확하게 구분되어짐을 알 수 있다.

그림 5는 설계된 음성 특징 벡터 추출기에 의해 추출된 LPC Cepstrum 0차 계수를 보여 주고 있다.

설계된 음성 특징 추출기에서 추출된 LPC Cepstrum 계수는 PC에서 똑같이 추출한 LPC Cepstrum 계수와 비교 결과 같음을 확인할 수 있었다.

Pentium III 800MHz, 메모리 256MB인 PC의 경우 벡터 추출 시간이 약 0.04s 이었다. 본 실험에서의 벡

터 추출 시간은 0.462s였다.

일반적으로 PDA와 같은 휴대 시스템의 CPU에 널리 사용되는 206MHz, 메모리 64MB의 사양을 가지는 Strong ARM의 경우 14초 이상 추출 시간이 소요된 것과 비교 해볼 때, TMS320C31을 이용하여 설계된 음성 특징 벡터 추출기는 실시간으로 충분히 사용될 수 있으리라 본다.

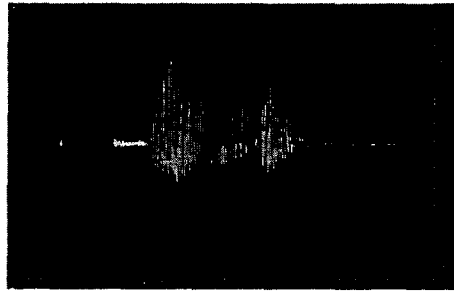


Fig 3. Speech sample signal.

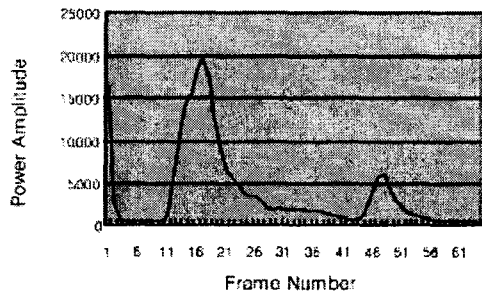


Fig 4. Frame energy of speech sample signal.

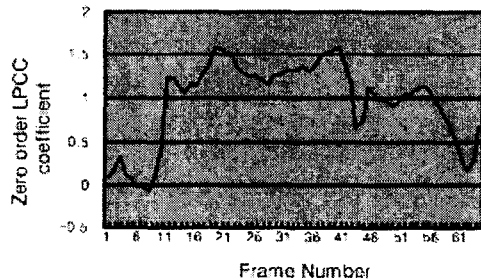


Fig 5. LPC Cepstrum zero order coefficient of speech sample signal.

## 5. 결론

본 논문에서는 임베디드 시스템을 위한 특징 벡터 추출기를 TMS320C31 DSP 칩을 이용해 구현하였다.

또한 특징 벡터 추출을 위해 신뢰성이 높고 음성 인식에서 가장 널리 사용되고 있는 LPC(Linear Predictive Coding)에 기반을 둔 Cepstrum 계수를 이용한 알고리즘을 이용하였다.

실험 결과 벡터 추출 시간은 0.462s로 실시간 음성 인식 가능할 것으로 본다.

즉 복잡한 연산과정을 가지는 특징 벡터 추출은 전용 벡터 추출기를 통해서 연산하고, 그 벡터를 이용한 실제 음성 인식은 임베디드 시스템의 CPU가 담당함으로써, 실시간 음성 인식이 가능할 것으로 본다.

이는 셀룰러폰, PDA, 전자수첩 등과 같은 휴대용 시스템의 음성 인식 분야에 응용될 수 있으리라 본다.

## 참고문헌

- [1] 오영환, "음성언어정보처리", 홍릉 과학 출판사, 1998.
- [2] 류홍석, 김정훈, 강성인, 강재면, 이상배, "다기능 진동원체어의 음성인식 모듈에 관한 연구", 대한 전자공학회 하계종합학술대회 논문집, 제 3 권, pp.83-86, 2002년 6월.
- [3] Shuzo Saito, "Fundamental of Speech Signal Processing", Academic Press, 1985.
- [4] J.R. Deller, J.G. Proakis and J.H.L. Hansen, "Discrete-Time Processing of Speech Signals", Macmillan Publishing Company, 1993.
- [5] 황인철, 김성남, 김용우, 김태균, 김수원, "LPC-CEPSTRUM 추출을 위한 전용 프로세서의 설계", 전자공학회 논문지, 제 34 권, C 편, 제 8 호, pp 611-618, 1997년 8월.
- [6] 윤덕용, "TMS320C31 마스터(DSP마스터시리즈 1)", OHM사, 1998.