

TMS320C6711을 이용한 어휘 인식기

최지혁, 김상준, 홍광석
성균관대학교 정보통신공학부 휴먼컴퓨터연구실

전화 : 031-290-7208 / 핸드폰 : 016-368-6169

Word Speech Recognition System by Using TMS320C6711

Ji-Hyouk Choi, Sang-Jun Kim, Kwang-Seok Hong

HCI Lab., School of Information and Communication Engineering, Sungkyunkwan Univ.

E-mail : sbncap@hanmail.net, kshong@yurim.skku.ac.kr, markjun@empal.com

Abstract

In this paper, we present a new speech recognition system using DSP chip. DSP chip used TMS320c6711 of TI. We designed hardware system including acoustic model, word list and code book in flash memory. The word candidates are recognized based on CV, VCCV, and VC units HMM. This system can be applied to various electric & electronic devices: home automation, robotics etc.

리를 요구하기 때문에 개발되어지고 있는 음성인식 제품들은 PC환경에서의 제품들이 대부분이다. [1][2]이러한 제품들은 일반적인 전자 제품에 음성인식의 기능을 구현하기에는 많은 어려움이 있다.

본 논문은 이러한 문제점을 해결하고자 고성능의 DSP 칩을 이용하여 임베디드 음성인식 시스템을 설계하고 구현하는 것을 목적으로 한다. 본 시스템은 다양한 적용분야(Robot, Telematics, Homeautomation)에 별도의 외부설계 없이 적용이 가능한 시스템을 설계하는 것을 목적으로 한다.

I. 서론

음성이 인간의 가장 자연스러운 의사전달 수단이라는 데에는 반론의 여지가 없을 것이다. 이러한 음성인식 기술은 키보드나 마우스와 같이 번거로운 입력수단을 사용하지 않고도 컴퓨터 및 기계를 조작 할수 있게 해 주고, 멀리 떨어진 곳에서 전화를 이용하여 컴퓨터에 수록된 정보를 꺼낸다든지, 각종 작업지시를 명령한다든지 하는 일들을 가능케 해준다. 전자 산업의 발달로 일반 가전에서부터 로봇, Telematics, Homeautomation 등과 같은 수많은 분야에서 음성인식 기술을 접목하기 위하여 많은 시도를 하고 있다. 그러나 음성인식의 구현을 위한 알고리즘들이 높은 계산 능력과 많은 메모

II. 음성인식 시스템

2.1 음성인식 시스템의 구조

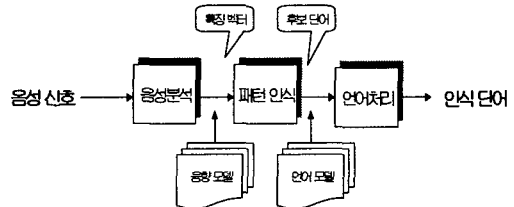


그림 1. 음성인식 시스템의 구조

음성인식 시스템은 크게 전처리부와 인식부로 나눌

수 있으며 인식성능 개선을 위해 언어모델을 추가할 수가 있다. 전처리부에서는 사용자가 발성한 음성으로부터 인식에 필요한 특징 벡터를 추출하고, 인식부에서는 음성 DB부터 혼란한 기본 패턴과의 비교를 통해서 인식 결과를 얻게되는데 실제로 사람이 발성한 음성은 표기와는 다른 발음이 존재하게 된다. 이러한 문제점을 극복하기위해 언어모델을 추가하게된다. 그림 1은 이러한 일반적인 음성인식 시스템의 구성도를 나타낸다.[1][2]

2.2 특징 추출

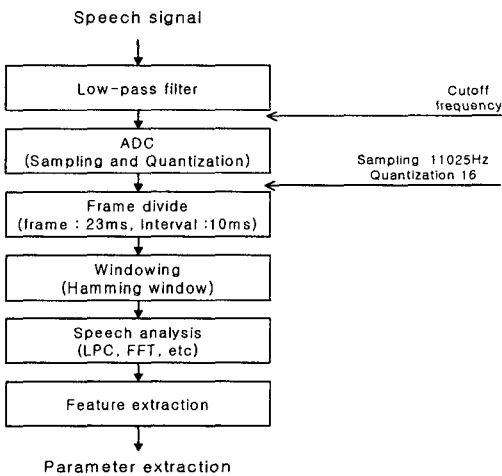


그림 2. 특징 추출 순서도

입력음성에 대한 전처리 과정에서 음성특징을 분석하기 위한 적절한 구간을 프레임별로 받아 프레임간의 불연속성을 최소화하기 위하여 적절한 윈도우 함수 $W(n)$ 을 곱해주게 된다. [3][4] 윈도우 함수의 선택에서 중요한 변수는 Main-lobe의 Bandwidth가 커지게 되면 좋지 못한 Frequency resolution 가지게되고 또한 Side-lobe peak의 dB가 커질수록 주파수 축에서 leakage가 커지게 되어 분석 시 좋지 않게 된다. 이런 모든점을 고려하여 Hamming window를 사용하게되었다.[3][4]

본 논문에서는 MFCC(Mel-Frequency Cepstral Coefficient)와 LPCC(Linear Prediction Coefficient Cepstrum)을 이용하여 실험을 하였고 시스템의 성능에 따라 적용을 하였다. 위에서 설명한 방법을 통하여 음성신호의 특징추출방법을 나타낸 그림이다. 각 특징 벡터는 13차로 계산하였다.[2][3]

2.3 HMM을 이용한 음성인식

DHMM은 관측 벡터를 VQ작업을 통하여 언어인 codebook의 index로 양자화하여 모델링하는 방법이다. 음성 구간 검출 알고리즘에 의해 검출된 음성신호는 매 프레임마다 특징 파라미터로 변환되는 vector열을 미리 준비된 codebook에 의해 벡터 양자화되는데, 이때 생기는 codeword index 열을 DHMM에서는 입력된 음성신호의 관측열로 사용한다. 따라서 DHMM의 관측열은 불연속적인 정수로 표시된다.

그림 3은 DHMM의 처리 과정을 나타낸것이다.[6]

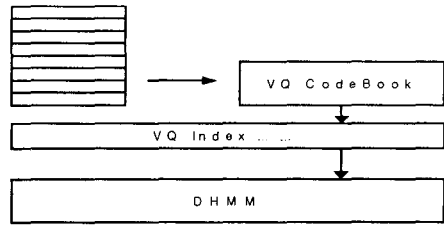


그림 3. DHMM의 처리과정

K-means 알고리즘을 사용하여 음성신호의 특징의 양자화 하였고 B-W 알고리즘을 이용하여 모델을 트레이닝하였다.[1][2]

III. TMS320C6711을 이용한 하드웨어 시스템의 설계

3.1 TMS320C6711의 장점

TMS320C6712 DSP는 여러 분야에서 고성능 / 고효율을 제공하는 최신의 부동소수점 DSP 솔루션이다. 코어 전압 1.8V, I/O 전압 3.3V에서 100MHz로 동작하는 이 제품은600MFLOPS의 고성능을 자랑한다. C6712 DSP는 C6000 TM DSP 플랫폼 기반의 다른 모든 제품군 멤버들과 코드 호환성을 갖고있어 개발자들이 소프트웨어에 투자한 부분을 그대로 이용할 수 있으며 향후 제품에서 최대 3GFLOPS까지 확장 업그레이드가 가능하다. 이러한 DSP에는 검증을 거친 기술에 기반을 둔 새로운 부동소수점 DSP 옵션이 널리 쓰이고 있는 TI의 TMS320C6711 DSP과 같이,C6712 DSP에는 레벨2 캐쉬,직접 매핑 메모리가 있는 레벨 1 캐쉬,또는 레벨 2 캐쉬와 직접 매핑 메모리의 조합 등으로 시스템의 필요사항에 따라 바꾸어 쓸 수 있는 다목적 메모리 아키텍처를 내장하고 있다. 주변기기 셋에는 16채널 EDMA(enhanced direct memory access; 고급 직접 메모리 액세스), 16비트 EMIF(external memory interface;외부 메모리 인터페이스),2개의

McBSP (multi-channel buffered serial port; 다채널 버퍼 직렬 포트) 및 32비트 타이머가 있어 고속 데이터 입출력과 유연성있는 설계가 가능하다. 또한 업계에서 가장 효율이 높은 DSP C 컴파일러의 지원도 받고 있는데, 이 컴파일러는 부동 소수점 성능을 최대 30%까지 높여 준다. [7][8][9]

3.2 음성인식 시스템 설계

1) 코덱 설계

AD535 Codec의 특징으로는 두개의 voice/data channel을 가지고 있고 이것은 analog hybrid circuitry 구조로 되어 있으며 서로 독립적으로 serial data 라인을 가지고 있어 양방향 음성 코덱을 구현할 수가 있다. AD535에서 보내지는 직렬 음성 데이터는 TMS320C6711내부의 EDMA(External Direct Memory Access)를 이용한다. EDMA는 음성신호의 입력과 출력에 각각 하나의 channel을 할당하고 Real-Time으로 데이터의 처리를 하기위하여 Interrupt를 이용하게 된다. 그림 4는 EDMA에서의 음성데이터 처리 과정을 보여준다.[7][8][9]

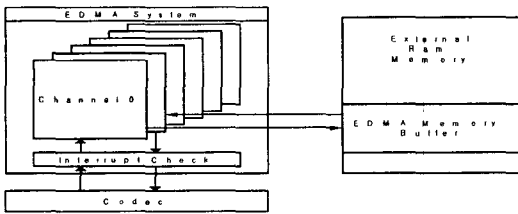


그림 4. 음성 코덱

2) External Memory 설계

모델의 저장을 위한 영역으로는 0XA0000000번지에 8MBYTE의 FLASH ROM을 설치하였다. 현재 시스템의 설계는 8BIT, 8MBYT형태로 되어 있지만 인식 알고리즘 수행시간을 단축을 위하여 16BIT 또는 32BIT 형태의 FLASH ROM을 설계 할 수도 있다.[8]

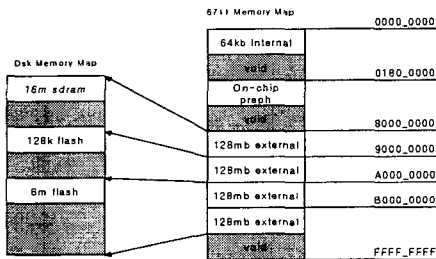


그림 5. 시스템 메모리 맵

3) LCD Display 및 rs-232통신

display부는 한글과 영문을 동시에 처리 할수 있도록 LCD를 사용하였다. LCD display는 compile사에서 제공하는 HD320이란 제품을 사용하였는데 이 제품의 특징으로는 rs-232통신을 이용하여 4*8의 한글을 지원하며 4*16의 영문을 지원한다. 본 6711에서는 McBSp라는 직렬포트를 2개 제공하고 있다. McBSp는 다양한 직렬 어플리케이션의 적용이 가능하다. 적용 가능한 디바이스들은 rs-232, Iic, T1-E1 framers, SPI등의 설계가 가능하게 되어 있다.

McBSp channel1은 codec과 연결이 되어 있고 남은 channel2가 LCD display와 연결이 되어 있는데 이것은 PC와의 통신을 위한 rs-232포트와 같이 연결을 하였는데, 이렇게 설계된 이유는 Dsk가 지원해주는 외부 포트의 제약으로 인해 동일한 channel을 사용 할 수 없었었다.[9]

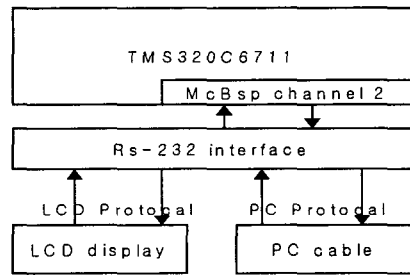


그림 6. 디스플레이부와 통신부 설계

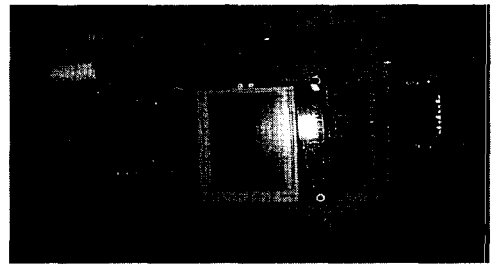


그림 7. 음성인식 시스템

V. 실험

4.1 모델 제작

본 논문에서 사용된 DB는 연구실에 제작한 남성 70명이 발성한 사연속 숫자음성 DB이다. 이 음성 DB를 가변어휘 음성인식기의 구현을 위해

부단어 단위 CV,VCCV,VC로 세그먼트하여 사용하였다 사연속 숫자음에서의 부단어 단위의 종류별수는 다음 표와 같다.

표 1. 사용 모델의 수

단위	갯수	비고
CV	8	고, 구, 사, 오.....
VCCV	72	아고, 아구, 아사....
VC	9	일, 아, 압....

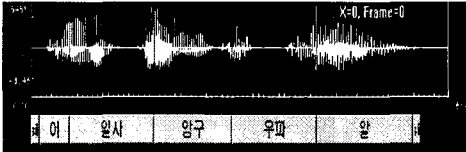


그림 7. 기준 모델 세그먼트

4.2 TMS320C6711 보드 실험

실험은 자체 제작한 TMS320C6711 보드를 이용하여 사음절 숫자음 실험 (35단어)를 실험 하였다. 실험 화자가는 각각 3회 발성을 하여 평균 인식률을 표시하였다. 음성신호의 샘플링은 8000Hz 이고 특징 추출은 MFCC, LPCC두가지로 하여 각각의 성능을 비교하였다.

표 2. MFCC와 LPCC의 인식률

	MFCC, 8000	LPCC, 8000
화자 1	88.57	85.71
화자 2	85.71	88.57
화자 3	88.57	82.28
화자 4	85.57	85.71
화자 5	91.42	88.57
평균	87.968	86.168

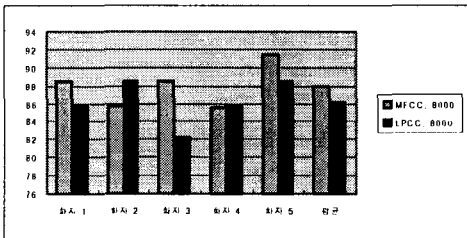


그림 8 MFCC와 LPCC의 성능 비교

V. 결론

본 시스템에서 고려한 가변어휘 인식기는 사용자의

입장에서 인식하고자 하는 단어를 자유로이 선택하여 인식시킬 수 있다는 장점이 있다. 수행 시간의 문제점은 Ti TMS320C6711 Ams을 이용한 파이프라인 코드 최적화 기법을 이용하면 충분히 만족할만한 시간을 보장 받을 수 있을 것이라 생각된다.

실제 가변어휘 음성인식기를 구현하기 위해서는 다양한 DB를 통해 많은 CV, VCCV, VC모델을 제작하여 실험을 해야 하나 모델 제작 시간 및 환경적 요인으로 인해 사연속 숫자 모델을 이용하여 기본적인 가변어휘 모델을 구성하여 보았다. 앞으로 더 진행해야 할 연구 방향으로는 일반 단어까지 처리 할 수 있는 다양한 DB를 이용하여 가변어휘 음성인식기를 구현하고 실제 구현에서 빠져 있는 언어 모델 처리부를 구현한다면 더 좋은 음성인식 성능을 보장 할 수 있을 것이라고 생각한다.

< acknowledgement >

본 연구는 2002(10차)년도 산학연 컨소시엄 지원으로 수행되었음.

Reference

- [1] 윤재선, 홍광석, "VCCV 단위를 이용한 어휘 독립 음성인식 시스템의 구현", 한국 음향학회지 21권 2호, 2002.
- [2] 윤재선 "한국어 음성인식 Dictation System의 구현", 박사학위 논문 2001. 10
- [3] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon, "Spoken Language Processing", Prentice Hall PTR 2001
- [4] Lawrence Rainer, Biing-Hwang Juang, "Fundamentals of Speech Recognition", Prentice Hall 1993
- [5] S. Michaelson, M. Steedman "HIDDEN MARKOV MODELS FOR SPEECH RECHGNITION"
- [6] Ti Application Report "Flashburn : A DSK Flash memory Programmer" Texas Instruments <http://www.ti.com>
- [7] Ti Application Report "TMS320C6000 EMIF to External Flash Memory" Texas Instruments <http://www.ti.com>
- [8] Ti Reference Guide "TMS320C6000 Peripherals" Texas Instruments <http://www.ti.com>
- [9] Ti Technical Training "TI DSP WorkShop", Texas Instruments <http://www.ti.com>
- [10] <http://www.dnmtech.co.kr/>