

DSP를 이용한 홈 오토메이션용 음성인식 시스템의 실시간 구현

김이재*, 김준성**, 양성일*, 권영현**

*한양대학교 제어계측공학과, **한양대학교 물리학과

Speech Recognition System for Home Automation Using DSP

I-jae Kim*, Jun-sung Kim**,

Sung-il Yang*, Y. Kwon**

*Dept. of Control & Inst. Eng., Hanyang Univ.,

**Dept. of Physics, Hanyang Univ.

E-mail : junfrau@hanmail.net

요약

본 논문에서는 홈 오토메이션 시스템을 음성인식을 도입하여 설계하였다. 많은 계산량과 방대한 양의 데이터의 처리를 요구하는 음성인식을 DSP(Digital Signal Processor)를 통하여 구현해 보고자 본 연구를 수행하였다. 이를 위해 실시간 끝점검출기를 이용하여 추가의 입력장치가 필요하지 않도록 시스템을 구성하였다. 특징벡터로는 LPC로부터 유도한 10차의 cepstrum과 log 스케일 에너지를 이용하였고, 음소수에 따라 상태의 수를 다르게 구성한 DHMM(Discrete Hidden Markov Model)을 인식기로 사용하였다. 인식 단어는 가정 자동화를 위하여 많이 쓰일 수 있는 10개의 단어를 선택하여 화자 독립으로 인식을 수행하였다. 또한 단어가 인식이 되면 인식된 단어에 대해서 현재의 상태를 음성으로 알려주고 이에 대해 자동으로 실행하도록 시스템을 구성하였다.

1. 서론

본 연구에서는 PC와 같이 한정된 기계에서 뿐만이 아닌 우리 주변 어느 곳에서나 이용할 수 있는 장치에 접목시키기 위하여 음성인식 시스템을 DSP를 이용한 소규모 시스템 안에 구현하였다. 본 연구를 통하여 기존에 PC에서 행해지던 음성인식 기술을 DSP와 약간의 하드웨어를 통하여 구현할 수 있음을 확인하였고 그 연구 과정에서 인식 알고리즘을 더욱 작고 빠르게 최적화 시켜 나갈 수 있었다.

본 논문을 위하여 실시간 끝점 검출이 가능한 End-Point detector를 개발하였고 특징벡터로는 LPC 계수로부터 유도한 10차원 cepstrum과 log 스케일 에너지를 추가하여 11차의 특징벡터를 이용하였다. 인식기는 음소의 개수에 따라 State의 수를 달리 구성한 HMM을 이용하였다.

2. 실음성 구간의 검출

실음성 구간의 검출은 입력된 음성신호로부터 들리도록 음성이 있는 구간을 찾는 과정이다. 음성(특히 단어)의 인식에 있어서 정확한 실음성 구간의 검출은 인식률은 물론 인식속도에 있어서도 영향을 주게 되므로 정확한 실음성 구간의 검출은 인식알고리즘과 더불어 음성인식의 매우 중요한 과정의 하나이다. 이상적인 실음성구간 검출기는 주변환경에 대해 강인하고 일관적이어야 한다. 또한 응용에 있어서 실음성구간의 실시간 검출 또한 매우 중요한 과제이다. 본 논문에서는 이러한 조건을 만족시키기 위해서 에너지와 경계교차율 LCR (Level Crossing Rate)을 동시에 이용하는 함수를 이용하여 실음성 구간의 검출을 수행하였다. 여기서 경계교차율이란 인접한 신호의 값이 주어진 영역을 통과하는 횟수를 말하며, 그 신호의 주파수의 정보를 포함한다. 경계 교차율은 무성음 구간에서는 값이 크게 나타나고, 주파수가 낮은 유성음 구간에서는 작게 나타난다. 음성신호의 단구간 에너지 E와 LCR값 L을 각각 구하여 일정구간에서의 곱으로 실음성 구간 여부를 판단하며 그 함수는 다음식과 같다.

여기서 N 은 분석하려는 한 프레임의 길이이며 n 은 현재 샘플의 인덱스이다.

$$E(n) = \sum_{m=n}^{N+1} |x(m)|$$

$$L(n) = \sum_{m=n}^{N+1} |sgn[x(m)] - sgn[x(m-1)]|$$

$$F(n) = E^2(n)L^2(n)$$

$$sgn[x(m)] = \begin{cases} 1, & x(m) > \text{threshold 1} \\ 0, & x(m) \leq \text{threshold 2} \end{cases}$$

3. 특징벡터

음성을 인식하기 위해서는 음성이 가지는 특징을 나타내는 파라미터를 추출하여 이 특징 파라미터로부터 학습과 비교과정을 거쳐서 인식을 수행하게 된다. 따라서 음성인식에 관한 연구는 효과적인 음성 파라미터를 추출하는 연구와 음성특징 파라미터로부터 효과적으로 학습, 비교하는 연구로 나뉘어 질 수 있다. 그 중에서 음성특징 파라미터의 추출은 음성을 되도록 적은 양의 수치데이터를 가지고 효과적으로 묘사하는 것이 매우 중요하다.

본 논문에서는 10차원의 LPC계수를 이용하여 Cepstrum계수를 유도하여 인식 파라미터로 이용하였다.

4. HMM을 이용한 음성인식

HMM을 훈련시키기 위해서 훈련 데이터로부터 특징 파라미터를 구해야하며, 심볼의 집합을 만들기 위해서 특징 파라미터들로부터 벡터 양자화를 설계해야 한다. 이 벡터 양자화기를 이용하여 훈련을 위해 입력된 음성에서 관측열을 생성하고 모델 파라미터 λ 가 수렴할 때까지 Baum-Welch 알고리즘을 반복해 재추정하여야 한다.

각 음성에 대해 만들어진 모델 파라미터들은 인식할 때 사용된다. 인식하고자 하는 음성이 들어오면 벡터 양자화를 하여 관측열을 생성하며, 이 관측열을 이용하여 각 음성의 모델에 대해 Viterbi 알고리즘을 통해 확률을 계산하며, 가장 큰 확률값이 나온 모델의 인덱스로 인식하게 된다.

훈련과 인식에 대한 구성의 순서도는 <그림 1>과 <그림 2>에 도시하였다.

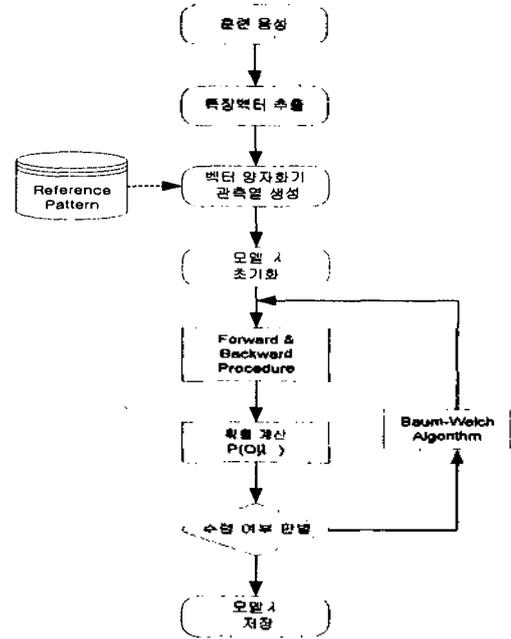


그림 1 HMM 음성인식기의 훈련과정

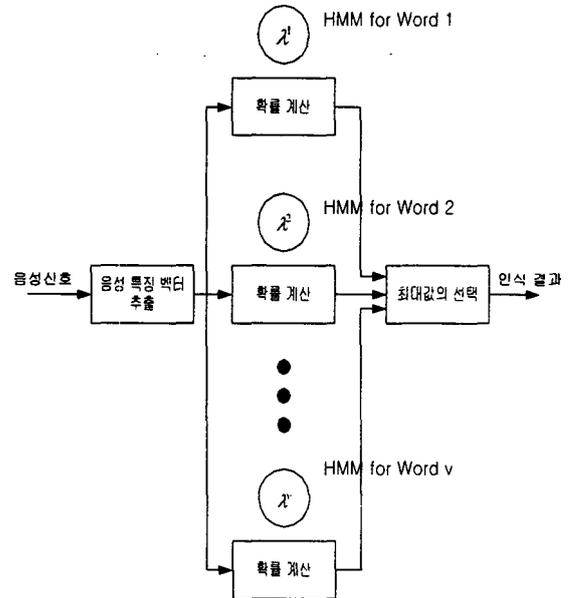


그림 2 HMM 음성인식기의 인식 과정

5. DSP

TMS320C3X 시리즈의 DSP는 TI사에서 처음으로 발매한 32비트 부동소수점 방식의 범용 DSP로서, 수정된 하바드 구조 및 내부 병렬처리로 고속의 연산기능을 가진다. 클럭 주파수 40MHz에서 20MIPS의 처리속도를 가진다. CPU 부분은 ALU, 곱셈기, 32비트 barrel shifter, 보조 레지스터 연산장치, 레지스터 파일, 그리고 이들을 연결하는 내부 버스로 구성된다. 내부 버스는 CPU1/CPU2 및 REG1/REG2가 있어서 메모리로부터 2개의 오퍼랜드와 레지스터로부터의 2개의 오퍼랜드를 동시에 액세스할 수 있으며, 1사이클에 ALU에서의 연산과 곱셈기에서의 연산을 동시에 수행할 수 있다. 바로 이러한 구조가 DSP가 빠른 연산을 할 수 있게 해준다. 메모리 영역은 2개의 1Kx32 비트 내부 RAM 블록(총 2K 워드), 64x32 바트의 외부 확장 메모리 영역, 16Mx32 비트의 외부 확장 메모리 영역(24 바트 어드레스) 등의 특징이 있다.

본 연구에서는 음성인식의 빠른 처리 위해 20ns이하의 지연시간을 갖는 SRAM장착하였다. 20ns이하의 SRAM을 외부 메모리로 사용할 경우 50MHz의 메인 클럭에서 wait 사이클이 필요하지 않으므로 더욱 빠른 음성과 코드의 처리가 가능하다. 외부램은 128k Word SRAM(4×1M 비트)을 장착하였고 프로그램을 RAM으로 loading 하여 인식과정을 수행할 수 있게 하였다.

본 연구를 위한 개발환경으로는 DSP Emulator와 Code Composer를 이용하였으며 C와 Assembler를 혼용하여 프로그래밍을 하였다.

6. 실험 및 결과

전체 시스템의 구성도는 <그림 3>과 같다.

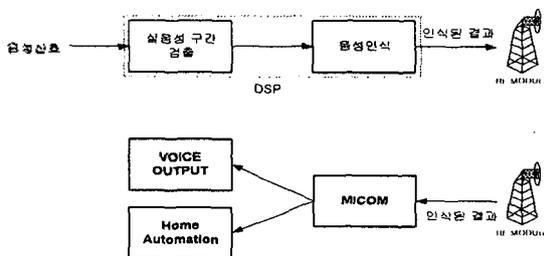


그림 3. 전체 시스템 블록 다이어그램

인식에 쓰인 어휘는 가정에서 많이 쓰일 수 있는 / 텔레(TV)/, /비디오/, /에어컨/, /난방/, /현관/, /세탁기/, /취사/, /가스/, /예/, /아니오/의 10가지 단어로

구성하였으며 10개의 모델에 대해 Baum Welch 알고리즘으로 훈련시킨 후 인식실험을 하였다.

하나의 음성을 인식하는데 약 15MIPS정도 걸렸으며 인식결과는 바로 홈오토메이션 시스템으로 보내져 인식된 결과의 실행과 이에 대한 안내음성이 나오게 하였고 전체 인식률은 10이 인식 실험에 참여한 결과 약 94%를 얻었다.

7. 결론

본 논문에서는 가정 자동화를 실현하기 위한 음성인식 시스템을 구현하였다. 입력된 음성은 DSP를 통하여 인식과정이 수행되고 인식된 결과는 다시 가정 자동화를 위한 제어기로 입력이 된다. 여기서 입력된 데이터를 가지고 우리가 원하는 가전기기나 현관의 상태를 음성으로 알려주고 이에 대해 자동으로 실행을 하게 시스템을 구성하였다.

8. 참고논문

- [1] L. Rabiner and B. Juang, *Fundamental of Speech Recognition*, Prentice Hall.
- [2] Texas Instruments, *TMS320C3X User's Guide*
- [3] Texas Instruments, *TMS320C3X Optimizing C Compiler User's Guide*.
- [4] Texas Instruments, *TMS320 Floating Point DSP Assembly Language Tools User's Guide*.
- [5] 박은정, 감지은, 백인찬, 권영현, 이진상, 양성일, "에너지와 영교차율의 새로운 함수를 이용한 실음성 구간 검출법", 한국음향학회 학술발표대회, 제 15권, 1(s)호, 1996.
- [6] John G. Proakis, Dimitris G. Manolakis, *Digital Signal Processing*, Prentice Hall, 1996.
- [7] J. F. Kaiser, "On Teager's Energy Algorithm and its generalization to continuous signals", In *Proc. 4th IEEE Digital Processing Workshop*, Mohonk, NY, Sept. 1990.
- [8] G. S. Ying, C. D. Mitchell and L. H. Jamieson, "Endpoint Detection of Isolated Utterance Based on a Modified Teager Energy Measurement", In *Proc. IEEE ICASSP-93*, pp.732-735, 1993.
- [9] L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoints of Isolated Utterances", *Bell System Tech. J.*,