심볼을 이용한 한국어 숫자음의 광역 음소군 분류에 관한 연구

이 봉규° ,이 극 , 황 희웅

서울대학교 컴퓨터 공학과

(A study of broad board classification of korean digits using symbol processing)

Lee Bong Gu°,Lee Guk , Hhwang Hee Yoong

Seoul Nat'l Univ. Dep. of Comp. Eng.

abstract

The object of this paper is on the design of an broad board classifier for connected. Korean digit. Many approaches have been applied in speech recognition systems: parametric vector quantization, dynamic programming and hiden Markov model. In the 80's the neural network method, which is expected to solve complex speech recognition problems, came back. We have chosen the rule based system for our model. The phoneme-groups that we wish to classify are vowel_like, plosive_like fricative_like, and stop_like.The data used are 1380 connected digits spoken by three untrained male speakers. We have seen 91.5% classification rate.

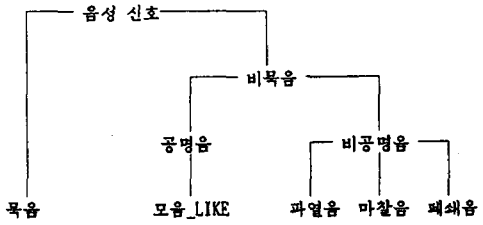
1. 서론

21세기를 지향하고 있는 모든 인간들이라면 생각할 수 있는 것 중에 하나는 "인간이 하고 있는 모든 작업을 기계가 대신할 수 있게 할 수 없을까?"라는 것이다. 물론 여기서 언급하고 있는 모든 작업이란 현존하는 인간이 할 수 있고, 생각이 가능한 영역에 제한되는 개념이라고 볼 수 있다. 이미 4세대 컴퓨터를 연구,개발하고있는 컴퓨터 공학자들은 차세대 컴퓨터로써 제 5세대 컴퓨터에 대한 연구 전략을 수립하여 미국방성, 일본등지에서는 의욕적인 출발을 하여 상당 부분을 이룩하고 있는 추세이다. 이러한 제 5세대 컴퓨터의 주요한 사양중에는 인간과의 인터페이스(Interface)를 위해 음성 인식 시스템의 내장을 필요 조건으로 하고 있는 바, 이러한 기능은 5세대 컴퓨터가 지향하고 있는 것이 무엇인 가를 단편적으로 보여주고 있다고 할 것이다. 이렇게, 음성 인식이 차지하는 비중은 나날이 커가고 있는 추세 속에서 한국어에 대한 인식 수준은 미국, 일본등에 비한다면 아직 초보적인 단계에 불과하다고 하지 않을 수 없다. 그러므로, 현실적으로 필요한 것은 한국어 음성에 대한 인식인데 이를 위해서는 한국어 음성 자체에 대한 기초 연구와 더불어 ,인식 단위를 음소라고 했을 때 음소군에

대한 분류가 이루어지지 않는 상태에서 음성 인식의 수준을 향상 시킨다는 것은 해당 시스템의 신뢰성에 있어서 문제가 발생할 수 있다는 것이다. 결국 숫어의 인식에 있어서 정확한 음소군의 분류는 다음에 진행될 음소별 분류에 있어서 탐색 공간(Search Space)을 줄일 수 있다는 점에서 뿐만 아니라 시스템의 유지,보수 측면에서도 유리할 수 있다는 것이다.

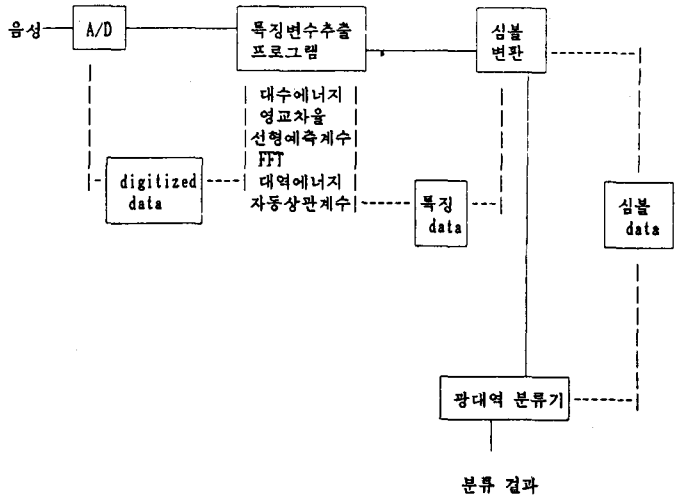
본 연구에서는 위와같은 배경하에서 한국어 음성상에서 특별히 숫자음에 대하여 광대역 음소군인 파열음, 마찰음,폐쇄음,모음으로 분류하는 것이다. 이의 구현을 위한 방법에는 여러가지 방법이 소개,구현되어 왔는데, 디지탈 신호 처리 측면에서의 filter bank 방법이나 패턴 분류(classification) 측면에서의 parametric & statistical 패턴 인식, 그리고 80년대 들어 각광 받고 있는 전문가 시스템 이론에 근거한 음성 신호의 지식 표현과 조작에 의한 지식을 기반으로한 방법, 최근 들어 다시 흥미를 끌고 있는 신경 회로망(Neural-Network)방법등이 있을 수 있으나 본 연구에서는 지식을 기반으로 한 접근 방법을 택하여 한국어 숫자음에 있어서의 광대역 음소군 분류를 시도하고자 한다.

발견된 음성 신호 부분에 대하여 광대역 음소군으로 분류하는 전략은 그림과 같다.



2. 시스템 구성

그림 1.1에서 볼 수 있는 것과 같이 본 연구에서는 추출된 심볼이 가지는 음성적 특징을 궁극적으로는 적절한 프롤로그 규칙(prolog rule)으로 구성하여 발음된 음성에 대하여 광대역 음소군으로 분류하는 시스템이라 할 수있다.



remark) 위 그림의 실선은 제어의 흐름, 점선은 데이타의 흐름을 표현

그림 1.1 전체 시스템 구성도

2.1) 음성 시료 (digitizing data)

실험에 사용한 숫자음은 0-9 까지의 한 자리수, 10-99 까지의 두 자리수는 모두 포함시키고, 100 - 999 까지의 세자리와 1000 - 9999 까지의 네자리수는 모두 포함시키지 않고 RGF(random generatP function)을 사용하여 세자리수 100개, 네자리수 20개를 선택하였다. 이렇게해서 얻어진 숫자음은 230개(1자리수 10개, 2자리수 90개, 3자리수 100개, 네자리수 20개)가 되었다.

이들 음성 데이타는 임시로 테이프에 저장되었다가 아나로그/디지탈 변환기에 의해서 디지탈화 되어서 디스켓에 저장된다. 아날로그/디지탈 변환기로는 ILS (Interactive laboratory system)을 사용하였다. ILS는 아날로그/디지탈, 디지탈/아날로그 변환 및 기본 통계자료 분석, 스펙트럼 분석, 필터링, 데이타 화일 조작들을 간단한 명령어에 의해 수행할 수 있는 패키지이다. 이 ILS에 의해 음성을 포함한 충분한 구간이 1.2 M 플로피 디스켓에 저장된다. 이 때 데이타의 크기는 30-100K바이트이며 한 프레임의 포인트수는 128 포인트로 하였다.

2.2) 특징 변수 추출

음성을 심볼로 표현하기 위해서는 음성의 특징 변수를 사용하여야한다. 여기에서는 음성의 특성을 비교적 잘 표현한다고 생각되는 대수 에너지와 영교차율, 대역별 에너지, 포르만트를 음성 신호의 특징변수로 택했다.

2.3) 심볼 변환

2.3.1) 심볼의 정의

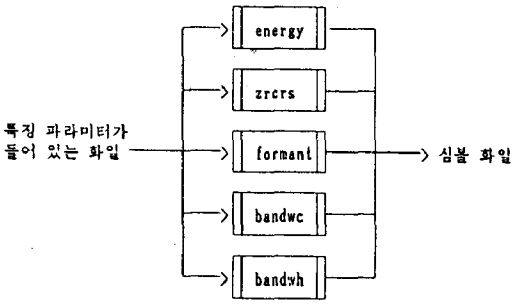
음성의 특징 변수들은 2차원 평면 상에 스칼라(scalar)값으로 표현하게 되면 인간은 시각적으로 음성 신호의 모양을 느낄수 있는데 이때 느끼는 모양을 묘사한 것을 심볼이라 할수있다.

2.3.2) 심볼의 역할

본 연구에서는 추출된 심볼이 가지는 음성적 특징을 궁극적으로는 적절한 프롤로그 규칙(prolog rule)으로 구성하여 발음된 음성에 대하여 광대역 음소군으로 분류하는 시스템이므로 심볼 기술 시스템(symbolic description system)은 최초의 음성 신호(Raw Speech Signal)와 광대역 음소군 분류기(Broad Board Classifier)의 중간적인 역할을 담당하고 있다고 볼 수 있다.

2.3.3) 심볼 추출

심볼 추출 프로그램의 구성도는 다음과 같다.



위의 구성도의 각 특징 파라미터에 대한 심볼을 추출하는 function은 순차적인 수행을 한다. 본 심볼 추출 프로그램에서 사용된 function을 자세히 기술하면 아래와 같다.

- energy() : 음성 신호의 프레임간의 차의 상대적인 값인 대수 에너지에 대한 심볼을 추출.
- zerocr() : 영교차율에 관한 심볼을 추출.
- formant() : 제 1 포르만트와 제 2 포르만트의 심볼을 추출.
- bandwc() : 60 - 3000 Hz 사이 주파수 성분의 크기에 관한 심볼을 추출.
- bandwh() : 3000 - 5000 Hz 사이 주파수 성분의 크기에 관한 심볼을 추출.

위에서 추출한 모든 심볼은 d("특징 파라미터","추출된 심볼","묘사의 시작", "묘사의 끝") 같은 형태의 심볼묘사로 변형되는데 그 의미는 "어떤 특징 파라미터가 일정 구간에서 추출된 심볼로 대표되는 모양적 특징을 갖는다."라고 할 수 있다. 이러한 심볼묘사는 프롤로그의 데이타베이스로써 사용된다. 본 연구에서 사용하는 모든 심볼묘사를 아래에 나타내었다.

가. 대수 에너지.

d("energy","narrowpeak",begin,end)

d("energy","widepeak",begin,end)

d("energy","narrowdip",begin,end)

d("energy","widedip",begin,end)

d("energy","stable",begin,end).

나. 영교차율.

d("zcr","high",begin,end)

d("zcr","middle",begin,end)

d("zcr","low",begin,end).

다. 포르만트(1 & 2)

d("formant1","stable",begin,end)

d("formant1","transient",begin,end

d("formant1","decreasing",begin,end).

d("formant2","stable",begin,end)

d("formant2","transient",begin,end)

d("formant2","decreasing",begin,end).

라. 대역별 에너지. (c & h)

d("cgraph","highpeak",begin,end)

d("cgraph","lowpeak",begin,end)

d("cgraph","deepdip",begin,end) ,

d("cgraph","shallowdip",begin,end).

d("hgraph","highpeak",begin,end)

d("hgraph","lowpeak",begin,end)

d("hgraph","deepdip",begin,end)

d("hgraph","shallowdip",begin,end).

d("hgraph","deepdip",begin,end)

d("hgraph","shallowdip",begin,end).

2.4) 광대역 분류기

2.4.1) 개요

"C" 프로그램에 의해서 판찰된 음성의 특징에 따라서 각 구간을 심볼화 하였다. 이렇게 추출되어진 심볼을 이용하여발음되어진음성구간을4가지의범주(category)로 나누는 작업이 필요하다. 이러한 광역 분류기의 설계를 여기에서는 규칙을 이용하여 설계할 목적이였기에 프롤로그를 이용하였다. 전체적인 광역 분류기의 흐름은 그림 1.2과 같다.

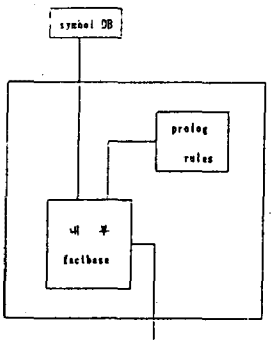


그림 1.2

2.4.2)프롤로그 규칙들

동적 factbase에 저장되어진 심볼을 사용하여 음성구간을 분류해내는 여러가지의 규칙들을 프롤로그로 구현한다. 여기에서는 음성을 (모음,유성자음),마찰음,폐쇄음, 파열음의 4가지 범주로 구별하고자 하기때문에 규칙들도 4개의 집합으로 나뉘어진다. 각각의 규칙 집합들은 4개의 음성 범주에 해당되는 고유한 특징들을 의미하며 이들 고유의 특징이 나타나는 음성의 구간에 적용되는 조건들을 가지고 있다. 만약 어느 음성구간의 에너지가 안정적이고, 포르만트가 역시 안정적이면 이 구간은 (모음, 유성자음)의 구간일 가능성이 매우 높고,반대로 마찰음이나,파열음에서는 이러한 특징이 나타나지 않기 때문에 이 구간은 규칙의 조건 중에서 (모음,유성자음)을 나타내는 것에 의해 음성구간이 나누어진다. 각각의 음성구간 범주와 해당 규칙을 도식화한 것이 그림 1.3이다.

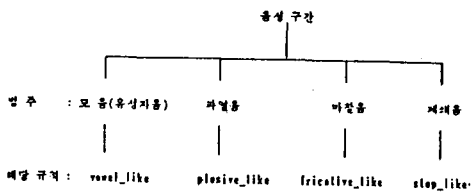


그림 1.3

3. 실험 및 결과

광역 분류기(broad board classifier) 의 실제 구현은 IBM-PC AT 상에서 운영되는 볼랜드사의 터보 프롤로그 VER. 1.1 을 이용하였다. 터보 프롤로그는 기존의 프롤로그가 interpreting형식 이었던 반면에 compile형식의 시스템이기 때문에 다른 프롤로그에 아주 적합하다고 생각되어진다. 프롤로그에 의해서 설계된 광역 분류기(broad board classifier)는 실험에 사용되는 실험 데이타에서 추출되어진 심볼을 이용하여 음소군 분류를 행하게 되는데 우리의 실험 결과는 3 명의 독립화자에 대해서 91% 정도의 인식율을 나타내었다. 실험중에 나타나는 에러는 주로 데이타에서의 모호성, 음의 중복 등에서 기인되는 것으로 생각되는데 , 이러한 사실을 뒷받침하는 것으로 한자리에서는 거의 완벽(100%)하게 음소군을 분류하는데 비해서 이를 음이 연속될 때에는 완벽하게 인식하지 못하고 오인식을 한다는 것이다.

4. 결론

본 연구에서는 한국어 연속 숫자음에 대하여 효과적인 음소군 분류를 위한 방법으로 심볼로 음성의 음향학적인 특성을 표현하고, 이렇게 추출된 심볼을 이용하여 음소군 분류 시스템을 설계하여 실제 음성 데이타에 대해 실험해 보았다. 지금까지 음성을 컴퓨터로 하여금 인식케 하려는 노력이 많이 진행되어 여러가지 패턴 인식 방법들이 논의 되었으나 음성자체의 모호성 때문에 이들 방법으로는 효과적인 인식 시스템의 구현에 어려움이 많았다. 이러한 이유로 지금 많은 다양한 방법들이 진행되고 있으며 그중의 하나가 바로 이러한 심볼을 이용한 인식 시스템 이다. 인식하려는 음성의 구간을 각각의 음향학적인 특성에 따라서 적절한 심볼로써 표현하고 이들간의 상호 연관된 규칙을 이용하여 실제 인식을 하는 시스템이다. 이러한 시스템의 장점은 확장이 용이하다는 것이다. 즉 새로운 음향학적인 특성이 발견되면 기존의 특성에 무관하게 새로운 규칙을 첨가함으로서 이러한 새로운 음향 특성을 분류에 적용시킬수 있다. 실제의 실험에서도 음소 상호간에 간섭으로 인하여 일반적 음향 특성을 벗어나는 음소에 대해서 새로운 규칙을 첨가함으로써 인식이 가능했다.

결론적으로, 본 연구에서는 이와같은 음성의 특성을 심볼로 표현하는 심볼 표현 시스템의 효과적인 구현을 위한 기초 모델을 제공하고 있다고 할수있다. 그러나, 본 연구에서는 단지 음향학적인 특성만을 고려하여 심볼을 구성하였기 때문에 음소간의 전이부분,상호 간섭과 같은 예외성들에 대해서 근본적인 해결책이 될 수는 없었다. 따라서, 보다 효과적인 분류를 위해서는 심볼을 표현하는데 있어서, 음향학적, 음운론적,의미론적인 특성도 고려하는 방법에 대한 연구가 많이 진행되어야 할것이며 이러한 연구가 효과적인 음성인식 시스템에로의 새로운 접근 방법의 하나인 심볼 표현 시스템(symbolic description system)에 대한 연구과제라 할것이다.

5. Reference

[1] A. M. Noll, "Cepstrum Pitch Determination," Jounal of Acoustici, Lily Lam, and Michel Gilloux, "Learnig and Plan Refinement in a Knowledge-Based System for Automatic Speech Recognition"

[2] B. S. Atal and L. R. Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition," IEEE trans. ASSP-24, No. 3, pp.201-212, Jun. 1976

[3] F. Fallside and W. A. Woods, Computer Speech Processing, Prentice-Hall, 1983.

[4] F. Itakura, " Minimum prediction residual principle applied to speech recognition", IEEE trans. Acoust., Speech Signal Processing, Vol. 23, pp.67-72, Feb. 1975.

[5] G. Mercier, A. Cauec, J. Monne, M. Querre and O. Trevarain, "Automatic Segmantation, Recognition of Phonetic Units and Training in the Keel Speech Recognition System", IEEE int.Conf. ASSP, pp. 2000-2003, 1982.

[6] James L. McClelland & David E. Rummelhart, " Parallel distributed processing", The MIT press , 1986.

[7] L. R. Rabiner and R. W. Shafer,"Digital Processing of Speech signal", Prentice-Hall, 1978.

[8] Renato De Mon, A. Giordana, P. Laface, L. Saitta, "Parallel algorithms for Syllable Recognition in Continuous speech", IEEE trans. ON PAMI.,PAMI-7 no.1, pp. 56-69, 1985.