論文
2-6-13

한국어 어휘 인식을 위한 혼합형 음성 인식 단위 (Monophone and Biphone Compuond Unit for Korean Vocabulary Speech Recognition)

이 기 정* 이 상 운** 홍 재 근***
(Ki-Jung Lee) (Sang-Woon Lee) (Jae-Keun Hong)

요 약

본 논문에서는 한국어의 발음 특성을 고려하여 인식시간 단축과 동시에 조음현상을 반영할 수 있는 인식단위 표현법을 제안하였다. 제안한 인식단위는 단음소(monophone)와 바이폰(biphone)의 혼합형으로서, 단음소 단위는 안정적인 특성을 나타내는 모음에 적용되고 바이폰 단위는 인접한 모음에 의해 변하는 자음에 적용된다. PBW455 데이터베이스에 대한 단어인식 실험에서 혼합형 단위표현법은 트라이폰 단위에 비해 비슷한 인식률을 나타내면서 57%의 인식시간 단축효과를 나타냈고, 음절 단위에 비해 향상된 인식률과 비슷한 인식시간을 나타내었다. 또한 트라이폰 및 음절 단위보다 적은 모델 수를 가져 메모리 양을 줄일 수 있었다.

ABSTRACT

In this paper, considering the pronunciation characteristic of Korean, recognition units which can shorten the recognition time and reflect the coarticulation effect simultaneously are suggested. These units are composed of monophone and biphone ones. Monophone units are applied to the vowels which represent stable characteristic. Biphones are used to the consonant which vary according to adjacent vowel. In the experiment of word recognition of PBW445 database, the compound units result in comparable recognition accuracy with 57% speed up compared with triphone units and better recognition accuracy with similar speed. In addition, we can reduce the memory size because of fewer units.

1. 서론

고도 정보화 시대로 변해감에 따라 각종 정보기기들과 접촉이 빈번해지고 있다. 이에 따라 인간과 기계 사이의 의사소통을 좀 더 자연스럽게 하려는 연구가 많이 이루어지고 있다. 인간과 기계 사이의 효과적인 인터페이스를 위한 도구로는 기본적이고 특별한 배움이 없이도 모든 사람들이 사용할 수 있는 통신수단인 음성을 이용하는 방법이 널리 사용되고 있다.

이러한 음성을 이용한 man-machine interface가 이루어지기 위한 핵심적인 기술 중의 하나가 음성인식이다.

음성인식 기술은 최근 음성인식에 필요한 주변기기의 발달과 음성인식의 사회적 필요성에 의해 많은 발전을 가져왔고, 또한 대용량 음성 인식기[3] 구현과 실용화를 위한 연구로 발전하고 있다. 그러나 기존의 단어 단위의 인식시스템은 대용량의 어휘를 인

* 정회원 : 포항1대학 컴퓨터응용과 부교수
** 정회원 : 포항1대학 컴퓨터응용과 전임강사
*** 정회원 : 경북대학교 공과대학 전자공학과 교수

논문접수 : 2001. 6. 18.
심사완료 : 2001. 6. 23.

식할 때 인식 어휘수의 증가에 한계가 있다. 즉 인식 어휘가 추가되면 새로운 모델을 정의하고 훈련하여야 하며, 새로운 모델을 훈련할 음성데이터가 그에 비례하여 증가하기 때문이다. 더욱이 새로운 모델의 첨가는 인식시간을 증가시킬 수밖에 없다.

이러한 문제점을 해결하는 방법은 음소나 음절과 같은 단어를 구성하는 기본 요소, 하위단어(sub-word)를 인식단위로 정하고, 그 결과를 이용하여 최종적으로 단어를 인식하는 것이다.[3,5] 이와 같은 하위단어 인식단위에는 음소, 음절, 반음소(demiphone), 반음절(demisyllable), acoustic unit 등이 있다.

음소를 인식 단위로 할 때, 훈련 및 인식에 사용하는 음소의 수는 영어의 경우에는 48개, 한국어의 경우에는 40~50개 정도이다.[6,10] 이 음소들을 인식단위로 훈련시키고 인식한 후, 이들을 연결하여 단어를 인식한다. 음소 단위는 적은 양의 음성데이터로도 잘 훈련될 수 있으나, 인접한 음소에 의해 영향을 받는 조음현상을 표현하기에는 충분하지 못하다.[10] 조음현상을 고려하기 위해 음소단위의 문맥에 따른 트라이폰 모델이 주로 사용되는데, 이 모델은 문맥의 앞뒤 음소를 고려하기 때문에 단어 내외에서 일어나는 조음현상을 잘 표현할 수 있어서 높은 인식률을 나타낼 수는 있으나 모델의 수가 많아져 인식시간이 길어진다. 음절단위[10]의 모델은 음소모델의 단점을 극복할 수가 있는데, 음절을 기본단위로 했을 때 인식모델의 수를 줄일 수 있어서 트라이폰 모델에 비해서 인식시간은 단축되지만 인식률은 떨어지게 된다.[4] 이것은 음절단위의 모델이 트라이폰 모델보다 조음현상을 잘 나타내지 못하기 때문이다.

본 논문에서는 트라이폰 모델과 같이 인접한 음소를 고려하여 조음현상을 표현할 수 있으면서, 음절모델과 같이 인식모델의 수를 줄여서 인식시간을 단축시킬 수 있는 모델을 제안한다. 우선 음성의 특성을 살펴보면 모음은 자음에 비해 안정구간이 비교적 길게 나타나고, 자음에 의한 모음의 변화 정도는 모음에 의한 자음의 변화보다 훨씬 더 적다. 따라서 제안한 방법에서는 이러한 특성을 이용하여 초성+중성+종성으로 구성된 한국어 음절에 대해, 중성 모음은 스펙트럼상의 안정한 특성을 감안하여 단음소 모델로, 초성과 종성 자음은 자음의 음소 변화를 반영하고자 왼쪽 또는 오른쪽 문맥의존 바이폰 모델로 나타내어 표현하였다. 제안한 방법에 의한 모델 수는 음절 모델 수보다 적게 된다.

기존의 인식모델과 제안한 인식모델의 비교를 위해서 인식단어 어휘수가 455이고 35,600개의 토큰을 가지는 PBW445 데이터베이스에 대해 동일한 조건에서 인식실험을 수행하였다. 인식실험을 통하여 제안한 방법이 트라이폰 모델에 가까운 인식률을 나타냄과 동시에 음절모델보다 인식시간을 단축시키는 결과를 확인하였다.

2. 음성인식 단위

단어단위 음성 인식기는 인식해야 할 어휘 수가 늘어남에 따라 새로운 어휘를 인식하기 위해서 새로 추가되는 모델을 만들어야 하는 한계점을 가지기 때문에, 대용량 어휘인식에서는 음소나 음절과 같은 하위단어를 인식대상으로 정하여 인식하고, 인식된 하위단어를 이용하여 단어를 인식하는 방법을 주로 사용하고 있다.

2.1 음소 인식

음소는 서로 구별되어 쓰이지 않는 음들의 집합을 뜻하며, 크게 자음과 모음으로 나눌 수 있고 초성, 중성 그리고 종성으로 구성된 한국어의 경우에는 초성과 종성은 자음으로, 중성은 모음으로 이루어진다. 초성자음의 개수는 19개이고 종성자음은 7개이다. 한국어 음성인식기에서 인식해야할 자음은 다음과 같이 19개이고

ㄱ ㄲ ㄴ ㄷ ㄸ ㄹ ㅁ ㅂ ㅃ
ㅅ ㅆ ㅇ ㅈ ㅉ ㅊ ㅋ ㅌ ㅍ ㅎ

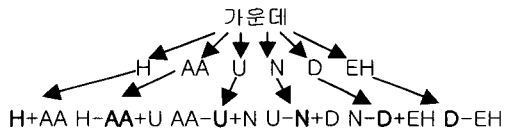
종성에 쓰이는 자음은 ㄱ ㄴ ㄷ ㄹ ㅁ ㅂ ㅇ 으로 7개이다.

또한 다음과 같이 10개의 모음과 11개의 이중모음을 인식할 수 있어야 한다.

ㅣ ㅡ ㅜ ㅐ ㅓ ㅗ ㅔ ㅏ ㅟ ㅜ
ㅠ ㅖ ㅕ ㅛ ㅒ ㅑ ㅞ ㅝ ㅙ ㅘ ㅢ

인식 단위로서의 음소는 지금까지 정의한 기본적인 음소 외에 유성음화된 자음과 초성과 종성의 위치에 따라 구분되는 음성학적인 음소를 추가할 수 있다. 이는 두드러진 변이음들을 구분함으로써 개개 음소의 변별력을 높이고자 함이 목적이며 동시에 인식률을 향상시키기 위함이다. 영어의 경우에는 48개의 음소모델을 설정하고 훈련함으로써 모든 단어나 문장을 인식할 수 있고, 한국어 인식에서는 40~50개의 음소모델이 필요하다. 본 논문에서 한국어 음성인식에 사용된 음소의 개수는 묵음을 포함하여 41개이다. 음소는 가장 기초적인 음성 단위이고 풍부한 훈련 음성데이터를 얻을 수 있기 때문에 모델의 훈련성이 좋은 장점을 가진다. 이로 인하여 대부분의 HMM 기반의 음성 인식 시스템에서 인식모델로 많이 적용되어 왔다. 그러나 음소는 수많은 문맥에 의한 전후의 환경에 따라 발생하는 다양한 음성학적 변화를 보인다. 음성은 인접한 앞뒤 음성에 영향을 많이 받기 때문에 같은 음소라도 앞뒤 음성에 영향을 받아 그 발음이 조금씩 다르게 나타난다.[1] 대용량 어휘 인식에서 단음소 모델은 이런 음성의 조음현상을 충분히 표현하지 못하는 단점을 가지고 있다.

이런 조음현상을 고려하기 위한 방법으로 문맥의존 바이폰 HMM과 트라이폰 HMM 모델을 사용할 수 있다. 바이폰 모델이란 하나의 음소를 독립적으로 훈련 및 인식시키는 것이 아니라, 인접한 앞뒤 음소 중 하나만 고려하여 훈련 및 인식하는 것이고 트라이폰은 훈련 및 인식하고자 하는 음소의 문맥상 인접한 앞뒤 음소를 모두 고려하는 것이다. 바이폰 모델이 앞뒤의 음소 중에서 더 크게 조음에 영향을 주는 하나의 음소만을 고려하는 반면 트라이폰 모델은 인접한 양쪽 음소를 모두 고려하기 때문에 조음현상을 더 잘 반영하므로 대용량 어휘인식에 주로 사용된다.



[그림 1] 단어 '가운데'의 트라이폰 표기과정

[Fig. 1] Triphone transcription procedure of /HAAUNDEH/ word

트라이폰 모델의 예로 [그림 1]의 '가운데' 단어의 인식모델을 들 수 있다. ARPABET[6]을 사용하여 단어 '가운데'를 단음소 모델로 표현한 후, 트라이폰 모델로 변환하는 과정을 나타낸 것이다. 그림에서 "-"부호는 인접한 앞 음소를, "+"부호는 인접한 뒤 음소를 고려함을 의미한다. 이와 같이 인접한 양쪽 음소를 모두 고려하는 트라이폰 HMM은 문맥상 이어지는 음성의 변화를 잘 모델링할 수 있다.

그러나, 조음현상을 잘 나타내기 위해서 문맥 종속 음소는 같은 음소라도 앞뒤의 음소가 다른 경우를 각기 다른 모델로 정의하므로, HMM 모델의 수가 많아지게 된다. 음성인식시간은 HMM 모델의 수에 비례한다. 이론상으로 음절 내에서 발생하는 트라이폰은 묵음을 포함하여 다음과 같이 약 50000개 정도의 모델을 만들어야 하고 바이폰도 1000여개의 모델을 만들어야 하기 때문에 그만큼 인식시간이 길어진다.

(자음 또는 모음 또는 묵음) − 모음 + (자음 또는 모음 또는 묵음) = (19+21+1)×21×(19+21+1) = 35301

(종성자음 또는 모음 또는 묵음) − 초성자음 + 모음 = (7+21+1)×19×21 = 11571

모음 − 종성자음 + (자음 또는 모음 또는 묵음) = 21×7×(19+21+1) = 6027

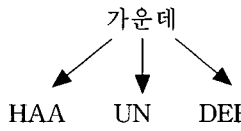
실제로 사용되는 트라이폰과 바이폰 모델의 수는 문맥의 제약으로 1000~2000개이고, 유사한 HMM 모델들의 파라미터 값들을 묶음으로써 전체 모델의 수를 줄일 수 있으나 그 한계가 있다.

2.2. 음절 인식

음성인식 시간을 단축하기 위한 방법으로 음절단위의 HMM을 만드는 방법이 있다. 음절은 더 이상 쪼갤 수 없는 최소의 발음 가능한 단위로 정의할 수 있다. 음절은 1~3음소의 결합으로 구성되고,[1,9] 음절내의 조음현상을 반영할 수 있고, 단어를 쉽게 음절 단위로 분해할 수 있어 발음사전을 작성하는데 편리하다.[2] 한국어 음성인식을 위해 이론적으로 필요한 음절의 수는 다음과 같이 3,360개이나,

초성자음×중성모음×종성자음 = 19×21×7
= 2793
초성자음×중성모음 = 19×21 = 399
중성모음×종성자음 = 21×7 = 147
중성모음 = 19×21×7 = 21

음소의 연결에 제약이 있기 때문에 실제로 쓰이는 수는 이보다 훨씬 적은 1,096개로 알려져 있다.[10] [그림 2]는 음절단위의 인식모델을 만들 때, 단어 '가운데'를 음절단위인 'HAA', 'UN'과 'DEH'로 변환되는 과정이다. 음절단위 HMM 모델을 사용한 음성인식기는 모델의 수가 적어서 인식시간을 단축시킬 수 있으나, 트라이폰 HMM을 이용한 음성 인식기보다 인식률이 저하된다.



[그림 2] 단어 '가운데'의 음절 표기과정
[Fig. 2] Syllable transcription of /HAAUNDEH/ word

3. 제안한 인식모델

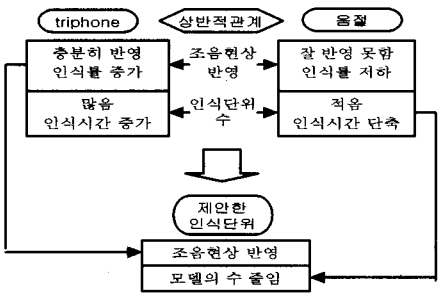
현재의 대용량 어휘의 음성인식에서 인식시간의 단축과 인식률의 향상은 중요한 요소이다. 트라이폰 HMM은 높은 인식률을 얻을 수 있으나 모델 수가 많아야 하기 때문에 인식시간이 길어지고, 음절 HMM은 인식시간은 짧으나 인식률이 트라이폰 HMM보다 못하다.

음성의 특성상, 초성과 종성 자음은 같은 음절내의 모음에 의해 받는 영향이 인접한 음절의 음소에 의한 영향보다 크기 때문에 같은 음절 내의 모음만을 고려한 바이폰 HMM으로 나타내어도 조음현상을 잘 고려할 수 있을 뿐만 아니라 인식 모델의 수를 줄일 수 있다. 중성모음은 자음보다 발음시간이 길며 스펙트럼 변화가 적고 안정하기 때문에 단음소 HMM으로도 음성특성을 잘 나타낼 수 있다.

따라서 제안한 인식시스템은 아래와 같이 자음을 나타내는 바이폰 HMM과 모음을 나타내는 단음소 HMM을 가진다.

x 문맥 독립적 모음모델
x-y 왼쪽 모음 x를 고려한 종성 자음 y 모델
y+x 오른쪽 모음 x를 고려한 초성 자음 y 모델

제안한 인식모델과 트라이폰 및 음절단위의 HMM의 비교를 <표 1>에 나타내었다. <표 1>은 PBW 445 음성데이터 중 '가운데' 단어를 각 인식모델별로 발음사전을 비교한 것이다. 이 음성데이터의 음소단위 표현법은 /H+AA+U+N+D+EH/ 이다. 제안한 단위표현법이 가장 작은 모델 수를 가지게 된다.



[그림 3] 트라이폰과 음절단위 및 제안한 인식모델 비교
[Fig. 3] Comparison among triphone, syllable, and proposed recognition model

[그림 3]에서는 트라이폰, 음절모델 및 제안한 인식모델의 특성을 비교하였으며, 제안한 인식모델을 사용한 인식기는 트라이폰 모델 인식기와 같이 조음현상을 잘 나타내며 또한, 음절 모델 인식기의 장점인 인식시간 단축을 동시에 가질 수 있음을 나타낸다.

<표 1> 단어 '가운데'의 인식단위별 표기법 비교
<Table 1> Comparison of transcription representation of /HAAUNDEH/ word

인식단위	트라이폰	음절	제안한 혼합형
ARPABET 발음열 표기법	H+AA H-AA+U AA-U+N U-N+D N-D+EH D-EH	HAA UN DEH	H+AA AA U U-N D+EH EH
모델의 수	약2000개	약1000개	500개 이하

4. 실험 및 결과

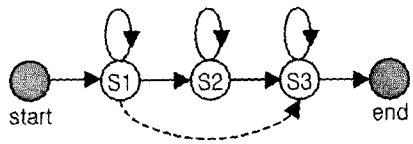
4.1 실험 데이터

인식기의 훈련 및 인식실험에 사용된 음성 데이터베이스는 한국전자통신연구소에서 음성 인식 연구를 위해 마련한 PBW445(Phonetically Balanced Words 445)이다. 이 데이터베이스는 남성화자 21명과 여성화자 19명 각각이 445개의 단어어휘를 2번씩 발음한 음성데이터이며 총 단어수는 35,600개이다. 표본화 주파수는 20kHz이고 12bit로 양자화되었다.

4.2 HMM Tool Kit

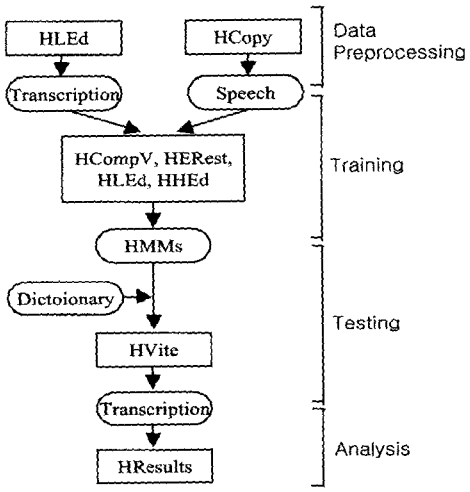
인식실험은 HMM을 기반으로 한 음성인식 시스템 구현 및 실험을 위한 상용도구인 HTK2.2[8]를 사용하였다. HTK(HMM Tool Kit)는 특별히 음성 인식 분야에서 HMM을 도구로 사용할 수 있도록 정형화된 구조를 가지고 있으며 HTK가 제공하는 기능들을 적절하게 사용할 때 대량의 음성데이터를 다루는 인식 실험을 정확하고 빠르게 수행할 수 있다는 장점을 가지고 있다. HTK는 크게 음성데이터와 관련된 각종 데이터를 처리하는 도구, 훈련을 위한 도구 그리고 인식 및 분석을 위한 도구들로 나누어져 있다. 각 도구들은 인식 시스템 구현 및 실험의 각 단계별로 사용할 수 있도록 구성되어 있다.

4.3 실험 방법 및 과정



[그림 4] 3상태 HMM 모델
[Fig. 4] 3-state HMM

인식 실험을 위한 훈련과정은 크게 특징 파라미터 추출과 HMM 모델구성, 즉 인식모델의 초기화 및 재추정으로 나눌 수 있다. 본 논문에서는 음성신호의 앞뒤 묵음구간을 제거하기 위해 끝점검출 과정을 거친 후, HTK에서 음성의 특징 파라미터를 추출하기 위한 도구인 HCopy를 사용하여 15ms 해밍 창함수를 5ms씩 중첩하여 12차 LPC 켑스트럼, 델타 켑스트럼, 델타 델타 켑스트럼 그리고 에너지, 델타 에너지와 델타 델타 에너지 파라미터들을 구하였다. 훈련 과정에서 각각의 모델은 [그림 4]와 같이 단순 좌우 3상태 HMM 모델을 사용하였다. [그림 4]에서 처음 상태(start)와 끝 상태(end)는 각각의 모델들을 연결하는 역할을 한다. 상태 1에서 상태 3으로의 천이는 묵음모델에서만 사용된다.



[그림 5] HTK 진행과정
[Fig. 5] HTK processing stages

HTK는 인식모델의 경계 정보가 없는 연속으로 발성된 음성데이터를 하위단어단위 HMM 모델의 훈련에 사용 가능케 하므로, 음성신호를 라벨링하는 단계를 거치지 않고 HCompV 도구를 이용하여 flat start procedure로 HMM 모델을 초기화하고 각 단어들을 발음 음소 단위로 기록해 놓은 발음사전과 HLEd 도구로 음성데이터를 표기한다. 초기화된 모델들은 HLEd와 HHEd 도구를 사용하여 실험에 적합한 모델로 편집 및 변환되고 Baum-Welch 알고리듬이 적용되어 재추정된다. 이 과정은 HERest 도구를 사용하여 이루어진다. 이 도구는 연속으로 발성된 음성데이터의 표기정보와 이미 초기화 또는 재추정된 하위단어단위 HMM 모델을 이용하여 여러 개의 하위단어단위 HMM 모델로 구성된 합성 HMM을 생성한다. 인식 과정은 HTK의 HVite 도구에 의해 수행된

다. HVite는 Viterbi 알고리듬을 구현한 도구로서 인식결과는 단어 사전에 정의된 심볼로 구성된 레이블링 파일이다. 이 레이블링 파일은 표기 파일의 형식과 동일하고 미리 준비된 인식용 음성데이터의 표기 파일과 인식 결과 만들어진 레이블링 파일을 동적 프로그래밍 기법을 이용해 비교하는 HResults 도구를 이용하여 인식률을 분석하게 된다. HTK를 이용한 데이터 전처리과정, 훈련 및 인식과정 그리고 분석과정의 진행을 [그림 5]에 나타내었다.

4.4 결과 및 고찰

인식단위 모델별 인식실험에는 PBW445 음성데이터 중에서 남성화자 18명과 여성화자 17명의 음성데이터를 훈련과정에 사용하였고 훈련에 사용한 35명의 음성데이터를 제외한 남성화자 3명과 여성화자 2명의 음성을 인식하였다. HMM 모델의 훈련은 각 인식모델별 HMM 훈련에서는 동일한 방법을 적용하였고, 실험은 동일한 환경에서 각각의 인식모델별(트라이폰, 음절, 바이폰 그리고 제안한 인식모델) 인식률과 인식시간을 측정하였다.

인식모델별 인식률 실험결과는 HTK의 인식률(Accuracy) 산출방법을 따른다. 인식률은 인식실험에 사용된 단어의 총수 (N), 인식된 단어 수 (H), 삭제된 단어의 수 (D), 대체된 단어의 수 (S), 삽입된 단어의 수 (I)를 이용하여 다음과 같이 계산한다.

$$Accuracy = \frac{N-D-S-I}{N} \times 100\ (\%) \qquad (1)$$

인식모델별 인식률과 인식시간 그리고 HMM 모델의 수를 <표 2>에 나타내었다. 인식모델별 인식률은 트라이폰 모델의 경우가 97.2%로 높았으며, 음절의 경우는 93.4%로 낮게 나타났다. 그리고 바이폰 인식모델과 제안한 인식모델을 사용한 인식률은 각각 96.4%, 96.2%이다. <표 2>의 음성인식 시간은 사용하는 기기의 성능에 따라 달라지기 때문에, 트라이폰 HMM을 기준하여 음절단위 HMM, 바이폰 HMM과 제안한 인식모델의 인식시간을 상대적으로 측정하였다. 음성인식 시간은 트라이폰 HMM을 기준으로 바이폰 HMM은 63%, 음절단위 HMM은 47%, 제안한 인식모델은 43%로 측정되었다. 동일한 조건하에서 인식시간은 HMM 모델의 수와 비례함을 알 수 있다.

<표 2> 인식모델별 인식률과 인식시간 비교

<Table 2> Comparison of recognition rate and time.

인식단위 / 성능	트라이폰	바이폰	음절	제안한 혼합형
단어 인식률	97.2%	96.4%	93.4%	96.2%
상대적 인식시간	1	0.63	0.47	0.43
모델의 수	1515개	629개	475개	325개

<표 2>의 모델의 수는 이론상으로 나타나는 수가 아닌, PBW445 데이터베이스에서 나타나는 문맥의존 HMM 모델의 수이다. 예를 들어 이론적 트라이폰 모델의 수는 약 3000개이나, 실제로 본 실험 데이터베이스에서 추출된 모델의 수는 묵음을 포함하여 1515개이다. 제안한 인식모델의 개수는 이론적으로 500개 이내, 실제 PBW445 데이터베이스에 적용했을 때 325개로 가장 적었다. 이로써 가장 짧은 인식시간을 가질 수 있었으며, 적은 수의 인식모델에도 불구하고, 트라이폰 인식모델이나 바이폰 인식모델에 비슷한 인식률을 나타내었다. 이것은 바이폰과 단음소의 조합으로 구성된 제안한 인식모델이 문맥의 특성을 잘 표현함과 동시에 모델의 수를 줄일 수 있었기 때문이다.

5. 결론

본 논문에서는 대용량 어휘 인식을 위하여 트라이폰 모델과 같이 문맥의 변화를 잘 나타내어 높은 인식률을 유지하면서, 음절단위보다 적은 모델 수를 가짐으로써 인식시간을 단축시킬 수 있는 모델을 제시하였다. 자음은 인접한 음절의 자음이나 모음보다 같은 음절내의 모음에 의해 받는 영향이 더 크기 때문에 자음은 바이폰 모델로 나타내었다. 바이폰으로 나타낸 자음모델이 조음현상을 잘 표현하며 트라이폰 모델보다 적은 모델 수를 가질 수 있기 때문이다. 모음은 스펙트럼 특성이 안정하므로 단음소 모델로 나타냄으로써 인식률의 변화를 일으키지 않고 음절모

델보다도 모델의 수를 줄일 수 있었다.

HTK 연속 음성인식 시스템을 사용하여 PBW445 음성데이터에 대한 인식실험 결과, 제안한 모델은 필요한 HMM의 수가 325개로서 음절모델보다 적은 모델 수를 가졌으며, 인식시간은 트라이폰 모델보다 약 57% 단축되었고 음절모델보다도 인식시간이 단축되었다. 제안한 모델은 트라이폰 모델에 가까운 96.2%의 인식률을 나타내었고 음절단위 모델에 비해 오인식률이 40% 가량 감소했다. 제안한 방법에서 단어사이의 조음현상도 고려하면 대용량 연속음성 인식시스템에 적용할 수 있을 것으로 기대된다.

※ 참고문헌

[1] A. E. Rosenberg, L. R. Rabiner, J. G. Wilpon, and D. Kahn, "DemiSyllable-based isolated word recognition system," IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. ASSP-31, no. 3, pp. 713-726, 1983.

[2] C.-H. Lin, C.-H. Wu, P.-Y. Ting, and H.-M. Wang. "Frameworks for recognition of Mandarin syllables with tones using sub-syllabicunits," Speech Communication, vol. 18, no. 2, pp. 175-190, 1996.

[3] C.-H. Lee, J.-L. Gauvain, R. Rieraccini, and L. R. Rabiner. "Large vocabulary speech recognition using subword units," Speech Communication, vol. 13, nos. 3～4, pp. 263-279, 1993.

[4] J. Hamaker, A. Ganapathiraju, J. Picone, and J. J. Godfrey, "Advances in betadigit recognition using syllables," Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 1, pp. 421-424, 1998.

[5] J. M. Kessens, M. Wester, and H. Strik, "Improving the performance of a Dutch CSR by modeling within-word and cross-word pronunciations variation," Speech Communication, vol. 29, nos. 2～4, pp. 193-207, 1999.

[6] L. R. Rabiner and B. H. Juang, Fundamentals of speech recognition, Prentice-Hall, pp. 20-42, 1998.

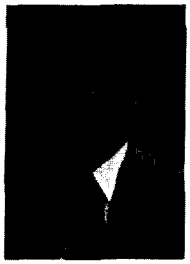
[7] L. R Rabiner and B. H. Juang, "Mixture autoregressive hidden Markov models for speech signal," IEEE Transaction on Acoustics, Speech, and Signal Processing, vol. ASSP-33, no. 6, pp. 1403-1413, 1985.

[8] S. Young, The HTK book, Entropic, 1996.

[9] T. Ji, Z. Wang, and D. Lu, "A method for chinese syllables recognition based upon sub-syllable hidden markov model," Internatinal Symposium on Speech, Image Processing and Neural Networks, vol. 2, pp. 730-733, 1994.

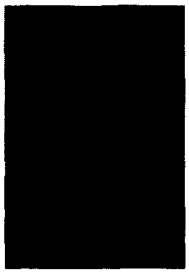
[10] 김유진, 김회린, 정재호, "인식 단위로서의 한국어 음절에 대한 연구," 한국음향학회지, vol. 16, no. 3, pp. 64-72, 1997.

이 기 정



1987 경북대학교
전자공학과 공학사
1990 경북대학교 대학원
전자공학과 공학석사
1997 경북대학교 대학원
전자공학과 박사수료
1991.3～현재 포항1대학
컴퓨터응용과 부교수
E-mail : leekj@pohang.ac.kr

이 상 운



1997 경북대학교
전자공학과 공학사
1999 경북대학교 대학원
전자공학과 공학석사
2001 경북대학교 대학원
전자공학과 박사수료
2000.3. ～ 현재 포항1대학
컴퓨터응용과 전임강사

홍 재 근

1975 경북대학교
전자공학과 공학사
1979 경북대학교
전자공학과 공학석사
1985 경북대학교
전자공학과 공학박사
1979～1982 경북산업대학
전임강사, 조교수
1983～현재 경북대학교
공과대학 전자공학과 교수
관심분야 : 음성인식,
음성코딩과 합성, 음질향상,
신호처리 시스템개발