

음성학적 특징을 이용한 연속 숫자음인식

김민정*, 황철준*, 김주곤*, 정현열*

* 영남대학교 정보통신공학과

Connected Digit Recognition Using Phonetical Features

Min-Jung Kim*, Cheol-Jun Hwang*, Joo-Kon Kim*, Hyun-Yeol Chung*

* Department of Information and Communication Eng., Yeungnam University

{knj, hcj, kjk, chy}@speech.yeungnam.ac.kr

요약

본 논문에서는 숫자음 인식시스템의 인식을 향상시키기 위한 연구로서 4연속 숫자음을 대상으로 연음 현상 및 경음화 현상 등과 같은 음성학적 특징을 고려하여 숫자음에 강건한 모델을 작성하는 방법을 제안하고 인식실험을 통하여 그 유효성을 확인하고자 한다.

이를 위하여 음성자료로서는 국어공학센터(KLE)에서 채록한 4연속 숫자음을 사용하며 인식의 기본단위로서는 음향학적 특징을 고려한 19개의 연속분포 HMM을 유사음소 단위(Phoneme Like Units; PLUs)로 사용한다.

또한, 인식실험에 있어서는 기존의 방법으로 모델을 작성한 경우와 연음 현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 모델을 작성한 경우에 대해서 유한상태 오토마타(Finite State Automata ; FSA)에 의한 구문제어를 통한 OPDP(One Pass Dynamic Programming) 법으로 인식실험을 수행하여 그 결과를 비교 검토하였다.

그 결과, 기존의 방법의 경우 64.6%, 음성학적 특징을 고려한 경우 68.6%의 인식률을 보여, 음성학적 특징을 고려한 경우가 4.0% 향상된 인식률을 얻어 제안한 방법의 유효성을 확인하였다.

1. 서론

최근에는 개인용 컴퓨터의 보급의 가속화와 컴퓨터를 이용한 신호처리기술 및 정보처리기술의 급속한 발전으로 인하여 인간의 가장 편리한 정보 교환 수단인 음성음 악용하고자 하는 Man-machine Interface 기술의 중요성이 강조되고 있으며, 현재 국외의 경우에 있어서는 단어 및 대어휘 연속음성인식 시스템에 대한 연구가 활발하게 수행되어 일부 한정된 태스크를 대상으로한 상용화 시스템이 개발되고 있다. 또한, 국내의 경우에 있어서는 단어와 연속음성을 대상으로한 인식시스템에 있어서는 비

교적 높은 인식률을 얻고 있으며, 일부 상용화를 위한 필드 테스트 중에 있다. 그러나 외국의 경우와는 달리 한국어 특성을 고려할 때 연속 숫자음에 있어서는 아직까지 인식률이 비교적 저조한 실정이며, 숫자음에 대한 정확한 분석과 인식에 관한 많은 연구가 요구되고 있다.

특히, 숫자음 인식 시스템이나 숫자음을 포함한 음성 인식 시스템의 실용화를 위해서는 발성화자의 개인성, 환경 잡음, 인식의 기본단위 등에 관한 연구와 더불어 숫자음의 음성학적 특징을 고려한 정확한 분석을 통한 모델 작성과 인식에 대한 연구가 요구되고 있지만 현재 대부분의 숫자음을 대상으로한 인식 시스템의 경우 이에 대한 충분한 검토없이 구현되고 있다.

따라서 본 연구에서는 숫자음 인식시스템의 인식을 향상시키기 위한 연구로서 4연속 숫자음을 대상으로 연음 현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 숫자음에 강건한 모델을 작성하는 방법을 제안하고 인식실험을 통하여 그 유효성을 확인하고자 한다.

이를 위하여 음성자료는 국어공학연구소(KLE)에서 채록한 4연속 숫자음을 사용하였으며, 이로부터 정적 특징으로서 멜-코efficient(Mel-Frequency Coefficient; MFC)과 동적 특징으로서 회귀계수(Regressive Coefficient; RGC)를 추출하여 연속 숫자음 인식실험을 위한 특징파라미터로 사용한다.

이때, 인식의 기본단위로서는 음향학적 특징을 고려하여 19개의 연속분포 HMM 유사음소 단위(PLUs)를 사용한다.

또한, 인식실험에 있어서는 기존의 방법으로 소모델을 작성한 경우와 연음 현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 모델을 작성한 경우에 대해서 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP [1,2,3] 법으로 인식실험을 수행하여 그 결과를 비교 검토

도하고 제안한 방법의 유효성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. II장에서는 본 연구에 사용된 음성자료 및 분석방법, III장에서 연속 숫자음의 음성학적 특징에 대하여 설명하고, IV장에서 음성학적 특징을 고려한 모델 작성 및 인식방법에 대하여 기술하고, V장에서는 인식 실험 및 고찰, 마지막으로 VI장에서 결론을 맺는다.

II. 음성자료 및 분석방법

연속 숫자음의 인식 실험을 위한 음성자료는 국어공화 센터(KLE)에서 구축한 한국인 남·여 72인의 4회 발생한 4연속 숫자음 중에서 남성 5인이 발생한 4연속 숫자음을 모델학습에 사용하고, 학습에 참여하지 않은 5인의 남성화자가 발생한 4연속 숫자음을 평가용 자료로 사용한다.

음성자료의 분석은 표 1과 같이 음성 데이터물 7KHz의 LPF를 통과시킨 후 샘플링 주파수 16KHz, 양자화 정도 16Bits A/D 변환기를 통해 이산데이터로 변환되고 Preemphasis 필터를 통과한 후 16ms(256 points) 길이의 해밍 윈도우를 사용하여 5ms(80points)씩 쉬프트 시키면서 분석된다. 이로부터 14차 LPC 캡스트럼 계수를 구하고, 10차의 LPC 멜-캡스트럼을 구하여 정적 특징파라미터로 사용한다. 또한 이로부터 10차의 회귀계수를 추출하여 동적 특징파라미터로 사용한다. 인식실험은 10차의 멜-캡스트럼과 10차의 회귀계수를 이용한다.

표 1. 음성자료의 분석조건.

Speech Data	KLE 4연속 숫자음
Sampling frequency	16khz
Filtering	LPF, 7khz
Resolution	16bits
Hamming window	16ms (256points)
Frame rate	5ms (80points)
Analysis	14order LPC analysis
Static Feature parameters	10order Mel-Cep. coeff.
Dynamic Feature parameters	10order Regressive coeff.

III. 연속 숫자음의 음성학적 특징

본 연구에서 사용되는 연속 숫자음에 있어서 일반적으로 나타나는 음성학적 특징으로서 연음 현상 및 경음화 현상이 있다. 연음 현상은 음절의 끝 자음이 모음으로 시작하는 음절이 이어 나오면 다음 음절의 초성으로 발음되는 현상을 말한다. 예를 들면, 연속 숫자음에서 '12'는 /일리/로, '75'는 /지로/로 발음된다. 또한, 경음화 현상은 파열음 /ㅂ, ㄷ, ㄱ, ㅅ, ㅈ/이 파열음, 마찰음, 파찰음 뒤에 오게 되면 경음 /ㅃ, ㄸ, ㄲ, ㅆ, ㅊ/으로 발음되는 현상을 말한다. 예로서는 '63'이 /육쌈/으로, '84'가

/팔썸/로 발음되는 것이 많이 나타난다[4].

본 실험에 사용된 4연속 숫자음에서 나타나는 대표적인 음성학적 특징으로 연음 현상 및 경음화 현상의 예로 표 2에 나타낸다.

표 2. 연속 숫자음의 연음 현상과 경음화 현상의 예.

선행 숫자	후행 숫자							
	0	1	2	3	4	5	6	9
1		일일 일릴	일이 일리	일삼 일쌈	일사 일싸	일오 이로	일육 일륙	
3		삼일 사릴	삼이 사미			삼오 사오	삼육 삼륙	
6	육공 육궁	육일 육질	육이 육기	육삼 육쌈	육사 육싸	육오 육고	육육 육륙	육구 육구
7		칠일 치릴	칠이 치리	칠삼 칠쌈	칠사 칠싸	칠오 치로	칠육 칠륙	
8		팔일 파릴	팔이 파리	팔삼 팔쌈	팔사 팔싸	팔오 파로	팔육 팔륙	

IV. 인식 방법

4.1 음소 모델

HMM(Hidden Markov Model)은 출력확률의 분포에 따라 크게 이산분포 HMM과 연속분포 HMM으로 분류한다. DHMM에서는 추출된 음성 특징 파라미터들의 출력확률분포가 벡터양자화에 의해 코드북내의 코드워드로 매핑되므로 벡터 양자화에 따르는 양자화 오차가 발생한다. 그러나, CHMM에서는 출력확률분포를 Gauss 분포나 Cauchy 분포로 직접 모델링 함으로써 양자화 오차를 막을 수 있다[4,5,6]. 따라서 본 연구에서는 CHMM을 이용하여 초기 음소모델을 작성하여 인식에 이용한다. 이때 CHMM 음소모델의 구조는 4상태 1혼합을 사용한다. 그림 1에 본 연구에서 사용한 연속분포 HMM 모델의 구성을 나타내었다.

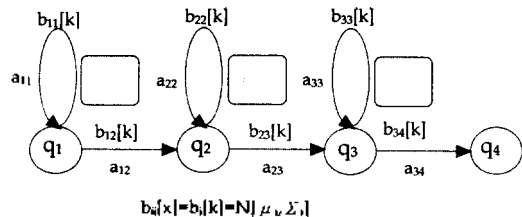


그림 1. 연속분포 HMM의 구성(4상태 1혼합)

4.2 인식 시스템

인식시스템은 표준패턴을 작성하기 위한 학습 단계와 표준패턴과 입력패턴과의 유사도를 측정하여 최적의 상태열을 찾는 인식 단계로 구성되며 그림 2에 4연속 숫자

음 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.

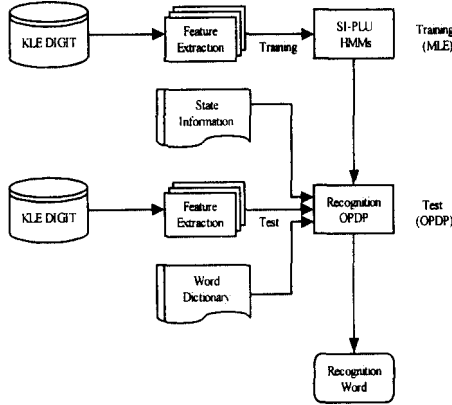


그림 2. 연속 숫자음 인식 시스템의 흐름도.

이때, 학습 단계에서 CHMM을 이용하여 음소 표준패턴을 작성하고, 인식단계에서 미리 작성한 단어사전과 유한상태 오토마타(FSA)에 의한 구분제어를 통하여 OPDP법으로 인식을 수행한다.

4.3 연속 숫자음 인식을 위한 FSN의 구성

표준패턴과 입력패턴 사이의 유사도를 측정하기 위한 일반적인 방법으로는 예측되어진 전체 표준패턴과 입력패턴을 정합시키는 방법이다. 그러나 이 방법은 인식하고자 하는 카테고리가 증가하고 인식 알고리즘이 복잡해짐에 따라 많은 시간과 문법적인 제약에 영향을 받는다. 따라서 유한상태 오토마타에 의한 구분제어를 통해 효율적으로 입력음성을 정합시키는 방법이 널리 사용되고 있다. 그림 3에 4연속 숫자음에 대한 유한 상태 오토마타의 예를 나타낸다.

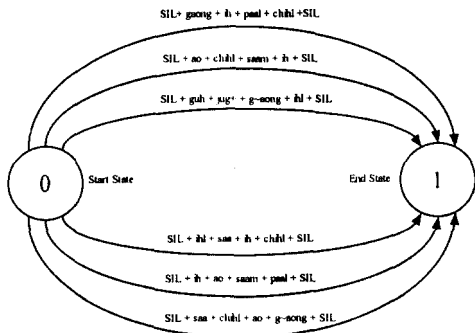


그림 3. 한정된 연속 숫자음인식을 위한 FSN의 구성 예.

그림 3의 경우에 있어서는 한정된 4연속 숫자음에 대해서는 효율적이지만 가능한 모든 4연속 숫자음을 대상으로한 인식을 고려할 경우 확장성 등에 문제점이 있다.

따라서 연속 숫자음을 대상으로한 실용화 시스템을 구성하기 위해서는 그림 4와 같이 유한 상태 오토마타를 작성하는 것이 매우 유리하다.

그러므로 본 연구에서는 가능한 모든 4연 숫자음 인식을 고려하여 그림 4와 같은 방법으로 유한 상태 오토마타를 구성하여 인식 실험에 사용한다.

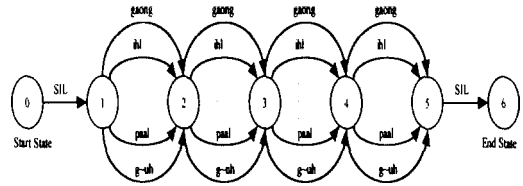


그림 4. 가능한 모든 연속 숫자음인식을 위한 FSN의 구성 예.

음성학적 특징을 이용하기 위하여 III절에서 검토한 연속 숫자음에서의 연음 현상과 경음화 현상을 고려하여 모델을 그림 5와 같이 작성한다. 기존의 방법의 경우는 현재 일반적으로 많이 사용하는 표기식 발음 사전을 이용하기 때문에 연속 숫자음에서 나타나는 연음 현상이나 경음화 현상을 충분히 고려하지 못한다. 그러나 본 연구에서는 연속 숫자음에서 나타나는 음성학적 특징을 고려하여 모델을 작성하여 인식실험을 수행한다.

/일일구구/

ih | l | ih | l | g~ | uh | g~ | uh (기존 방법)

/일릴구구/

ih | l | rih | l | g~ | uh | g~ | uh (연음 현상)

(a) '1199'의 경우.

/삼육사구/

s | aa | m | ju | g+ | s | aa | g~ | uh (기존 방법)

/삼육싸구/

s | aa | m | ju | g+ | ss | aa | g~ | uh (경음화 현상)

(b) '3649'의 경우.

그림 5. 음성학적 특징을 고려한 모델.

이때 인식실험에 있어서 연음 현상 및 경음화 현상 등을 고려하기 위해서 음성타이틀을 화자별로 직접 들으면서 음성학적 특징이 확실하게 나타나는 4연속 숫자음을 추출하여 사용하였다.

V. 인식 실험 및 고찰

인식실험에 있어서는 II절에서 설명한 음성자료를 사용하여 한국인 남·여 72인의 4회 발성한 4연속 숫자음 중에서 남성 5인이 발성한 4연속 숫자음으로 음성학적 특징을 고려한 모델을 작성하여 표준 패턴을 작성하고, 학습에 참여하지 않은 남성 5인의 화자가 발성한 4연속 숫자음을 평가용 자료로 사용하여 인식 실험을 수행하였다.

먼저, 기존의 방법으로 모델을 작성하고 일반적으로 많이 사용하고 있는 표기식으로 4연속 숫자음 인식을 위한 FSN을 구성하여 인식실험을 수행하였다. 그 결과물 그림 6에 나타내었다.

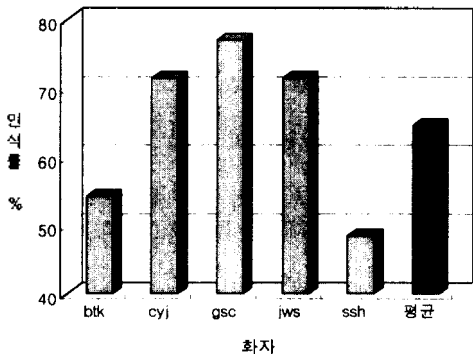


그림 6. 4연속 숫자음 인식률(기존의 방법).

이상의 인식 실험결과 5인의 남성화자 평균 64.6%의 인식률을 보였다.

또한 본 연구에서 제안한 음향학적 특징으로서 연음현상 및 경음화 현상을 고려하여 모델을 구성하고 FSN을 적용하여 인식실험을 수행하였다. 그림 7에 인식결과를 나타내었다.

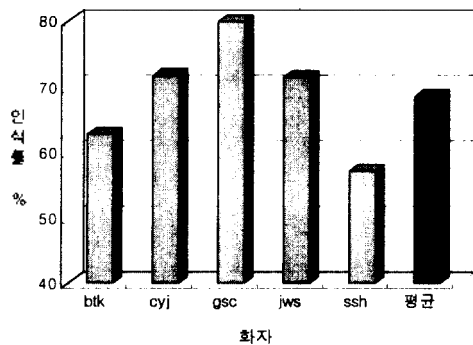


그림 7. 4연속 숫자음 인식률(제안한 방법).

인식 실험결과, 연음현상 및 경음화 현상을 고려한 경우 5인의 남성화자 평균 68.6%의 인식률을 보였다.

이상의 결과로부터 음성학적 규칙을 적용한 경우와 있어서는 음성학적 규칙을 적용하지 않은 경우보다 4.0%의 향상을 보여 한국어 연속 숫자음에 있어서 음성학적 특징의 적용이 유효함을 확인하였다.

VI. 결론

본 논문에서는 숫자음 인식시스템의 인식률 향상을 위한 연구로서 4연속 숫자음을 대상으로 연음 현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 숫자음에 강건한 모델을 작성하는 방법을 제안하고 인식실험을 통하여 그 유효성을 확인하였다.

인식 실험 결과, 음성학적 규칙을 적용한 경우 68.6%로서 음성학적 규칙을 적용하지 않은 기존 방법의 경우 64.6%에 비해 4.0%의 향상된 인식률을 보여 제안한 방법의 유효성을 확인하였다.

전체적으로 볼 때 보다 높은 인식률 향상을 위해서는 연속 숫자음에 나타나는 음성학적 특징을 잘 표현할 수 있는 보다 많은 음성데이터를 채록하여 모델을 작성하여 인식에 이용하면 보다 향상된 인식성능이 기대된다.

향후 이들 결과를 단어나 연속문장으로 확장하여 음성 인식 성능 향상에 대한 연구를 진행하고자 한다.

* 본 연구에서 사용한 단어데이터베이스는 국어공학센터에서 구축한 4연속 숫자음 음성데이터베이스를 사용하였습니다.

참고 문헌

1. J.H. Lee, B.K. Kim and H.Y. Chung, "Environmental Adaptation Using Maximum A Posteriori Estimation for Korean Word Recognition," Proceeding of IEEE Invited Workshop on Pattern Recognition for Multimedia Techniques, 1996.
2. 越川忠, "連続音聲認識システムにおけるHMMの話者適應化に関する研究," 修士學位論文, 1993.
3. 中川聖一, 甲斐充彦, "文脈自由文法制御によるOnePass型HMM音聲認識法," 信學論誌 D-II, Vol. J76-D-II, No.7, pp. 1337-1345, 1993.
4. 우인봉, 이강성, 김순협, "HMM의 교정학습과 후처리를 이용한 연결 숫자음인식에 관한 연구," 제11회 음성통신 및 신호처리 워크샵 논문집, pp.161-165, 1994.
5. 中川聖一, "確率モデルによる音聲認識," 電子情報通信學會編, 1989.
6. X. D. Huang, Y. Ariki and M. A. Jack, "Hidden Markov Models for Speech Recognition," Edinburgh Univ., 1990.