

# 숫자음 분석과 인식에 관한 연구<sup>†</sup>

## (A Study on Spoken Digits Analysis and Recognition)

김 득 수\*, 황 철 준\*\*  
(Deok-Soo Kim, Chul-Joon Hwang)

**요 약** 본 논문에서는 한국어의 음성학적 규칙을 적용한 연속 숫자음 인식에 관하여 연구한다. 연속 숫자음의 인식률은 일반적으로 음성인식 시스템 중에서 낮은 인식률을 나타낸다. 따라서 숫자음에 대하여 강건한 모델을 작성하기 위하여 음성 특징 파라미터와 음성학적 규칙을 적용하고 실험을 통하여 그 유효성을 확인하고자 한다. 이를 위하여 음성자료는 국어공학센터(KLE)에서 채록한 4연속 숫자음을 사용하며 인식의 기본단위로서는 음성학적 규칙을 적용한 19개의 연속분포 HMM을 유사음소 단위(PLUs)로 사용한다. 또한, 인식실험에 있어서는 일반적인 멜 캡스트럼과 회귀계수를 이용한 경우와 음성학적 규칙과 특징을 확장하여 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, 멜 캡스트럼만을 사용한 경우 55.4%, 멜 캡스트럼과 회귀계수를 사용한 경우에는 64.6%, 특징 파라미터를 확장한 경우 74.3%, 음성학적 특징까지 고려한 경우 75.4%로 기존의 경우보다 높은 인식률을 보였다. 따라서, 음성 특징 파라미터를 확장하고 음성학적 규칙까지 함께 적용한 경우 비교적 높은 인식률을 보여 제안된 방법이 연속 숫자음 인식에 유효함을 확인하였다.

**Abstract** This paper describes Connected Digit Recognition with Considering Acoustic Feature in Korean. The recognition rate of connected digit is usually lower than word recognition. Therefore, speech feature parameter and acoustic feature are employed to make robust model for digit, and we could confirm the effect of Considering Acoustic Feature throughout the experience of recognition. We used KLE 4 connected digit as database and 19 continuous distributed HMM as PLUs(Phoneme Like Units) using phonetical rules. For recognition experience, we have tested two cases. The first case, we used usual method like using Mel-Cepstrum and Regressive Coefficient for constructing phoneme model. The second case, we used expanded feature parameter and acoustic feature for constructing phoneme model. In both case, we employed OPDP(One Pass Dynamic Programming) and FSA(Finite State Automata) for recognition tests. When applying FSN for recognition, we applied various acoustic features. As the result, we could get 55.4% recognition rate for Mel-Cepstrum, and 67.4% for Mel-Cepstrum and Regressive Coefficient. Also, we could get 74.3% recognition rate for expanded feature parameter, and 75.4% for applying acoustic feature. Since, the case of applying acoustic feature got better result than former method, we could make certain that suggested method is effective for connected digit recognition in korean.

### 1. 서 론

최근 개인용 컴퓨터의 보급의 가속화와 컴퓨터를 이용한 신호처리기술 및 정보처리기술의 급속한 발전으로 인하여 인간의 가정 편리한 정보 교환 수단인 음성을 이용하고자 하는 Man-machine Interface 기술의 중요성이 강조되고 있으며, 현재 국외의 경우에 있어서는 단어 및 대어휘 연속음성인식 시스템에 대한 연구가 활발하게 수행되어 일부

한정된 태스크를 대상으로 한 상용화 시스템이 개발되고 있다. 또한 국내의 경우에 있어서는 단어와 연속음성을 대상으로 한 인식시스템에 있어서는 비교적 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 음성인식 시스템의 실용화를 위해서는 발성화자의 개인성, 환경 잡음, 인식의 기본단위 등에 대한 연구와 더불어 숫자음의

\* 대구공업대학 컴퓨터정보과

\*\* 대구과학대학 정보전자통신계열

† 본 연구는 1997년도 대구공업대학 교내연구비 지원에 의한 것임

음성학적 특징을 고려한 정확한 분석을 통한 모델 작성과 인식에 대한 연구가 요구되고 있지만 현재 대부분의 숫자음을 대상으로 한 인식 시스템의 경우 이에 대한 충분한 검토 없이 구현되고 있다.[1,2,5]

따라서 본 연구에서는 연속 숫자음의 실용화를 목표로 4연속 숫자음을 대상으로 음성 특징파라미터의 확장 및 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-켄트럼(Mel-Frequency Coefficient; MFC)과 동적 특징으로서 회귀계수(Regressive Coefficient; RGC)를 추출하여 연속 숫자음 인식실험을 위한 특징파라미터로 사용한다.

이때, 인식의 기본단위로서는 음성학적 특징을 고려한 경우의 18개의 연속분포 HMM 유사음소단위(PLUs)와 그렇지 않은 경우의 48개의 연속분포 HMM 유사음소단위(PLUs)를 사용한다.

또한, 인식 실험에 있어서는 기존의 방법으로 음소모델을 작성한 경우와 연음현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 음소모델을 작성한 경우, 그리고 각 화자의 음성학적 발성 특징에 대해서 유한 상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP[6,9,10]법에 의한 구문제어법으로 인식실험을 수행하여 그 결과를 비교 검토하고 제안한 방법의 유효성을 확인하고자 한다.

본 논문의 구성은 다음과 같다. II장에서는 연속 숫자음의 음성학적 특징에 대하여 설명하고, III장에서는 음성 자료 및 분석 방법, IV장에서는 음성학적 특징을 고려한 모델 작성 및 인식 방법에 대하여 기술한 후, V장에서는 인식 실험 및 결과 고찰을 통하여 제안한 방법의 유효성을 확인한다. 그리고, 마지막으로 VI장에서는 결론을 맺는다.

## 2. 연속 숫자음의 음성학적 규칙

음성인식에 있어서 인식하고자 하는 언어의 특징을 이해하는 것은 중요하다. 이것은 일본어, 영어, 한국어 등의 각 언어가 자기의 특징을 가지고 있기 때문이며, 또한 그러한 각 언어의 특징은 발음시에도 그대로 나타난다. 예를 들어, 영어나 일어를 인식하는 인식기가 인식률이 좋다고 해서 이것을 한국어에 적용한다고 해도 한국어 나름대로의 특징을 이해하지 않고 적용시키지 않는다면 그 결과는 좋지 않을 것이다. 여기서는 한국어의 음성학적 특징을 먼저 살펴보기로 한다.

### 2.1 국어의 음소 변동 규칙[3]

국어에서 나타나는 음성학적 규칙으로는 국어의 음소 변동 규칙으로서 종성 규칙, 자음 동화, 구개음화, 전설 모음화, 경음화, L-첨가, 사이 시옷의 첨가 등이 있다.

- (1) 종성 규칙 : 종성(받침)이 발음되는 양상을 규칙화한 것을 종성(받침) 규칙이라 하는데, 종성 규칙은 중화 규칙, 겹받침 단순화, 격음화, 연음 규칙으로 나뉜다. 이 규칙들은 음소 연결의 제약과 밀접한 관계를 가진다.
- (2) 자음 동화 : 자음과 자음이 이어나올 때 두 자음이 서로 비슷해지거나 같아지는 현상을 자음동화라 한다. 국어의 자음 동화는 유음화, 장애음의 비음화, 유음의 비음화, 변자음화로 나뉜다.
- (3) 구개음화 : 국어의 형태소 끝 자음 /ㄷ, ㅌ/은 /ㅣ/로 시작하는 의존 형태소(조사, 어미, 접미사) 앞에서 구개음화되어 /ㅊ, ㅌ/으로 바뀐다.
- (4) 전설 모음화 : 후설 모음 /ㅏ, ㅣ, ㅜ, ㅠ, -/가 뒤 음절의 전설 고모음 /ㅣ/의 영향으로 같은 높이의 전설 모음 /ㅘ, ㅙ, ㅚ, ㅛ, ㅣ/로 바뀌는 현상을 전설 모음화('ㅣ' 모음 역행 동화, umlaut)라 한다.

이 외에도 경음화 등이 있으며, 이들은 연속 숫자음에서 나타나는 현상이므로 다음절에서 다루기로 한다.

### 2.2 연속 숫자음에 나타나는 음운 현상

연속 숫자음 인식에 나타나는 음운 현상은 연속 숫자음 자체가 한정되어 있기 때문에, 앞 절에서 살펴본 모든 음운 현상이 나타나는 것은 아니다. 그러므로 이 절에서는 연속 숫자음에서 나타날 수 있는 음운 현상에 대해서만 살펴보기로 한다.

- (1) 연음 규칙 : 연음 규칙이란 형태소 끝 자음이 모음으로 시작하는 의존 형태소(어미, 조사, 접미사)가 이어 나올 때 중화 규칙의 적용을 받지 않고 다음 음절의 초성으로 발음되는 것을 말한다. 그러나 형태소의 끝 자음이 /ㅍ, ㅌ, ㅋ, ㆁ, ㆁ/인 낱말들은 같은 환경에서 사람에 따라, 방언에 따라 서로 다른 자음으로 실현된다.
- (2) 자음 동화 : 국어의 자음동화는 유음화, 장애음의 비음화, 유음의 비음화등으로 나뉜다.
  - 유음화 : 국어에서 /ㄴ/과 /ㄹ/은 연이어 나올 수 없기 때문에 /ㄴ/이 /ㄹ/의 앞이나 뒤에서 /ㄹ/로 바뀐

다. 이 현상은 비음/ㄴ/이 유음/ㄹ/로 바뀌므로 유음화라 한다.

- 유음의 비음화 : 국어에서 /ㄹ/은 /ㄴ/과 /ㄹ/ 이외의 자음 뒤에서 /ㄴ/으로 바뀌는데, 이 현상을 유음의 비음화라 한다.
- 장애음의 비음화 : 장애음 /ㅂ, ㄷ, ㄱ/은 비음 앞에 나타날 수 없기 때문에 비음 앞에서 /ㅃ, ㄸ, ㅇ/으로 바뀐다. 이 현상을 장애음의 비음화라고 한다.

(3) 경음화 : 국어에서의 경음화 현상은 그 종류가 다양하며, 지역에 따라, 그리고 세대간에 따른 차이도 있다. 같은 서울 토박이라도 세대에 따라 같은 낱말을 다르게 발음하는 경우가 있다. 기성 세대, 젊은 세대를 막론하고 서울 토박이들이 연음으로 발음하는 소리를 다른 방언 사용자들(특히 경상도 방언 사용자들)이 경음으로 발음하는 경우가 있다. 혹은, 그 반대의 경우도 있다.

### 2.3 DATABASE에 나타나는 음운 현상

본 연구에서 사용되는 연속 숫자음의 음성학적 특징을 살펴보면, 연음현상과 경음화 현상이 많이 나타나며, 그 외의 다른 음성학적 현상도 나타난다. 실제 data로 사용되는 음성 데이터를 살펴보면 연음 현상의 예로서 연속 숫자음에서 '12'가 /일리/로, '75'가 /치로/로 발음된다. 경음화 현상의 예는 '63'이 /육쌘/으로 발음되는 예이다. 또한, 이 이외의 다른 음성학적 현상으로는, 예를 들어, '84'가 /팔짜/, '96'이 /구육/으로 발음되는 현상이 나타난다.

본 실험에 사용된 4연속 숫자음에서 나타나는 대표적인 음성학적 특징으로 연음 현상과 경음화 현상을 표 1에 나타내고, 그 외의 음성학적 특징을 표 2에 나타낸다.

<표 1> 연속 숫자음의 연음 현상과 경음화 현상 및 기타 음운 현상

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

<표 2> 연속 숫자음의 연음 현상과 경음화 현상의 음운 현상

선행 숫자	후행 숫자
	6
4	사육 -> 사륙
5	오육 -> 오륙
9	구육 -> 구륙

## 3. 음성 자료 및 분석 방법

### 3.1 음성 자료

연속 숫자음의 인식 실험을 위한 음성자료는 국어공학센터(KLE)에서 구축한 한국인 남·여 72인이 4회 발성한 4연속 숫자음 중에서 남성 20인이 발성한 4연속 숫자음을 모델 학습에 사용하고, 학습에 참여하지 않은 5인의 남성화자가 발성한 4연속 숫자음을 평가용 자료로 사용한다.

HMM 작성을 위한 음성 데이터는 국어공학연구소에서 채록한 4연속 숫자음 데이터베이스를 사용하여 유사 음소 단위의 레이블 정보를 이용하여 작성한다. 표 3은 실험에 사용된 음성 데이터베이스를 나타낸다.

인식의 기본 단위로는 한국어의 음향학적 특징과 대어휘 연속 음성인식으로서의 확장성을 고려하여 48개의 유사음소 단위를 이용한다.

<표 3> KLE 4연속 숫자음 데이터베이스

1	공이팔칠	0287	19	칠사팔구	7489
2	오칠삼이	5732	20	이이사사	2244
3	구육공일	9601	21	사육이일	4621
4	사일오육	4156	22	구일칠육	9176
5	일일구구	1199	23	삼공사오	3045
6	일삼구팔	1398	24	팔오구공	8590
7	육팔사삼	6843	25	오오공공	5500
8	공칠일이	0712	26	육구칠이	6972
9	오이육칠	5267	27	오팔육일	5861
10	육육삼삼	6633	28	삼육사구	3649
11	이사공구	2409	29	공삼일육	0316
12	칠구오사	7954	30	칠공팔삼	7083
13	일팔이삼	1823	31	팔일구사	8194
14	육삼칠팔	6378	32	구이공오	9205
15	팔팔칠칠	8877	33	일사이칠	1427
16	삼오일공	3510	34	이오삼팔	2538
17	팔공육오	8065	35	사칠오공	4750
18	이구삼사	2934			

### 3.2 음성 분석

A/D 변환된 모든 음성자료는 Pre-emphasis 필터를 통과한 후 16ms 길이의 해밍 윈도우를 거쳐 구간으로 분석된다. 이 때 각 구간은 5ms씩 shift된다. 이로부터 자기 상관(Autocorrelation) 방법을 사용하여 20차의 LPC 계수를 구하고 14차의 LPC 켈스트럼 계수를 구한 후, 마지막으로 이 14차의 LPC 켈스트럼으로부터 10차의 멜 켈스트럼 계수를 추출한다. 그리고 이로부터 전후 40ms의 동적 특징을 고려한 10차의 회귀계수를 추출하여 음성의 특징 파라미터로 사용한다[8]. 표 4에 음성 분석 조건을 나타낸다.

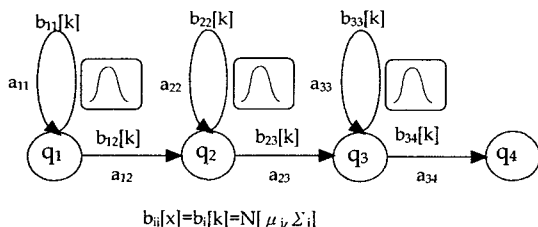
<표 4> 음성자료의 분석조건

Sampling Frequency	16 kHz
Resolution	16 bits
Hamming Window	16 msec ( 256 points )
Frame Rate	5 msec ( 80 points )
Analysis	14 order LPC
Feature Parameters	10 order MFCC + 10 order RGC

## 4. 음성학적 규칙을 적용한 인식 방법

### 4.1 음소 모델

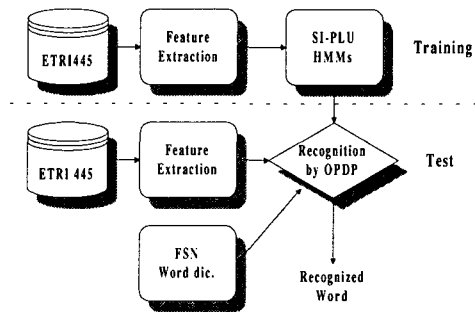
HMM(Hidden Markov Model)은 출력확률의 분포에 따라 크게 이산분포 HMM과 연속분포 HMM으로 분류한다. DHMM에서는 추출된 음성 특징 파라미터들의 출력확률분포가 벡터양자화에 의해 코드북내의 코드워드로 매핑되므로 벡터 양자화에 따르는 양자화 오차가 발생한다. 그러나, CHMM에서는 출력확률분포를 Gauss 분포나 Cauchy 분포로 직접 모델링 함으로써 양자화 오차를 막을 수 있다 [4,7,8]. 따라서 본 연구에서는 CHMM을 이용하여 초기 음소모델을 작성하여 인식에 이용한다. 이때 CHMM 음소모델의 구조는 4상태 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 숫자음 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.



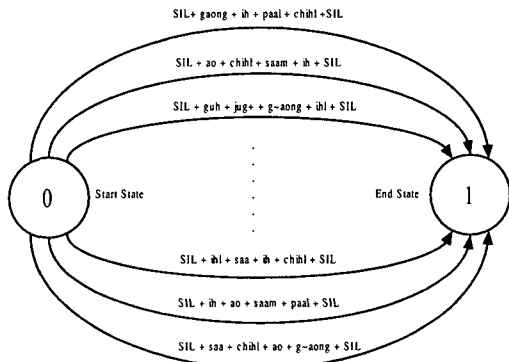
<그림 2> 4연속 숫자음 인식 시스템의 전체 구성도

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

### 4.3 연속 숫자음 인식을 위한 유한상태 오토마타

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 구문제어를 통해 효율적으로 입력음성을 정합시키는 방법이 널리 사용되고 있다.

그림 3에 4연속 숫자음에 대한 유한 상태 오토마타의 예를 나타낸다.

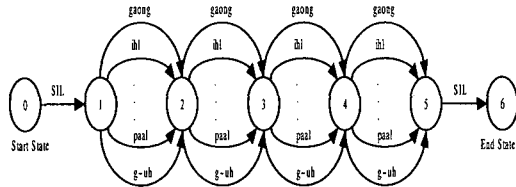


<그림 3> 한정된 연속 숫자음 인식을 위한 FSN의 구성 예

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

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

본 연구에서는 가능한 모든 4연속 숫자음 인식을 고려하여 그림 4와 같이 유한 상태 오토마타를 구성하여 인식 실험을 수행한다.



<그림 4> 가능한 모든 연속 숫자음 인식을 위한 FSN의 구성 예

#### 4.4 음성학적 특징의 분석과 모델의 구성

인식 실험을 위해서는 II절에서 음성학적 특징으로서 검토한 연속 숫자음에서의 연음 현상과 경음화 현상을 고려하여 음소모델을 구성하였다.

이러한 모델을 구성하는데 있어서, 데이터베이스에 일괄적으로 음운현상을 적용하는 것은 무리가 있다. 이것은 발성화자에 따라 음운 현상이 나타나는 화자가 있고 나타나지 않는 화자가 있기 때문이다. 이것은 화자의 구성이 서술말(표준말)을 쓰는 화자들로 구성된 것이 아니라 각 지방의 방언을 고려하여 각 지방마다 동등한 비율로 화자를 구성하였기 때문이다. 그러므로 이 경우에 대해서도 고려

를 해주어야 한다.

먼저, 이러한 것을 검토하기 위하여 데이터베이스에 일괄적으로 음운현상을 적용시켜 모델을 작성하여 보았다.

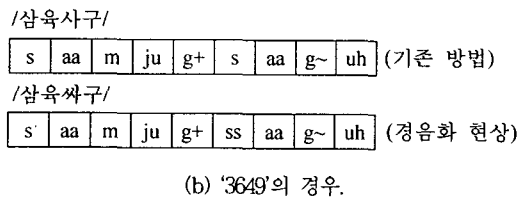
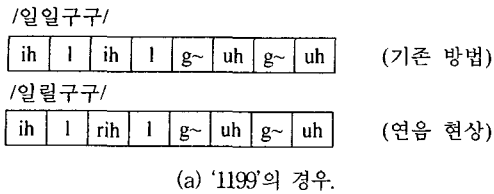
다음으로, 데이터베이스의 모든 화자의 발음을 직접 wav 파형을 관찰하며 시각적으로 관찰하고, 청각적으로 들어서 여기에서 나타나는 음운현상을 적용시켜 모델을 작성하여 보았다. 이때 이러한 방법으로 모델을 구성하는데는 객관성이 결여되기 때문에 최대한의 객관성을 유지하기 위하여 여러명의 관찰자가 지정한 음운 현상을 모델의 구성에 적용하였으며, 이러한 관찰을 반복적으로 시도함으로써 객관성을 유지하도록 하였다. 이러한 관찰의 결과를 표 5에서 보인다. 표 5는 데이터베이스에서 음운 현상이 나타나는 데이터만을 표시하였다.

이러한 모델 구성의 예를 그림 5에서 보인다. 그림에서 기존의 방법은 연속숫자음에서 나타나는 음운 현상을 전혀 고려하지 않은 일반적으로 현재 많이 사용하는 표기식 발음 사전을 이용하여 모델을 구성하기 때문에 연속 숫자음에서 나타나는 연음 현상이나 경음화 현상을 충분히 표현하지 못한다. 따라서 본 연구에서는 연속 숫자음에서 나타나는 음성학적 특징을 고려하기 위하여 그림에서와 같이 음소모델을 작성하여 인식실험을 수행하였다.

<표 5> KLE 4 연속 숫자음 데이터 베이스에서 나타나는 음운현상

data num	Database	음운현상 결과	data num	Database	음운현상 결과
2	오칠삼이	오칠사미	17	팔공육오	팔공육고
3	구육공일	구륙공일	21	사육이일	사륙이일
4	사일오육	사일오륙	22	구일칠육	구일칠륙
5	일일구구	일릴구구	24	팔오구공	파로구공
6	일삼구팔	일쌈구팔	26	육규칠이	육구치리
8	공칠일이	공칠일리	27	오팔육일	오팔륙일
10	육육삼삼	육육삼삼	29	공삼일육	공삼일륙
13	일팔이삼	일팔리삼	31	팔일구사	파릴구사
16	삼오일공	사모일공 삼모일공	35	사칠오공	사치로공

여기서, 인식 실험을 수행함에 있어서, 위의 표에 나타나는 음성학적 현상을 모두 적용하지는 않았다. 이것은 앞에서 언급한 화자마다 다르게 발생할 수 있기 때문이다. 그래서, 모델을 작성할 때는 화자 발성의 관찰에서 가장 많이 나타나는 음운현상만을 적용하였으며, 이것 또한 객관성을 유지하기 위한 것이다. 참고로, 이러한 관찰의 결과, 데이터 베이스에 나타나는 음운현상은 연음 현상 및 경음화 현상 등이 있었으며, 이것을 모델의 구성에 적용하였다.



<그림 5> 음성학적 특징을 고려한 음소모델

인식실험에서 있어서도 위에서 언급한 연음현상이나 경음화 현상, 이용하기 위해 음성데이터를 직접 듣고 나서 음성학적 특징이 확실하게 나타나는 4연속 숫자음에 대해서 위의 그림과 같이 구성하였다.

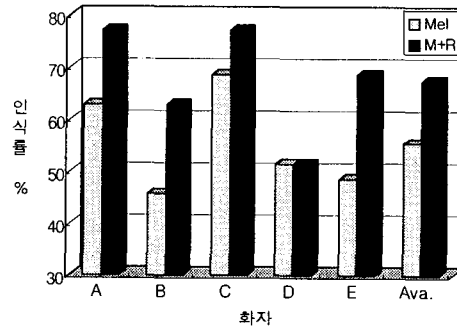
### 5. 인식 실험 및 고찰

HMM 학습과 인식 과정은 크게 학습 단계, 적응화 단계, 인식 단계로 나눌 수 있다. 학습 단계에서는 최우 추정법(ML)을 사용하여 4상태 1혼합 left-to-right CHMM 모델로 초기 HMM을 작성하며, 적응화 단계에서는 최대 사후 확률 추정법(MAP)을 사용하여 화자 종속 HMM을 작성하여 인식단계에서 OPDP 알고리즘을 이용하여 인식을 수행한 결과를 출력한다.

연속 숫자음 인식의 실용화를 위한 특징 파라미터의 확장과 음성학적 특징 적용의 유효성을 확인하기 위해, 인식 실험을 위한 음성자료로는 국어공학센터(KLE)에서 구축한 한국인 남·여 72인의 4회 발성한 4연속 숫자음 중에서 남성 20인이 발성한 4연속 숫자음으로 표준 패턴을 작성하고, 학습에 참여하지 않은 남성 5인의 화자가 발성한 4연속 숫자음을 평가용 자료로 사용하여 ultra-10s 워크스테이션에서 인식 실험을 수행하였다.

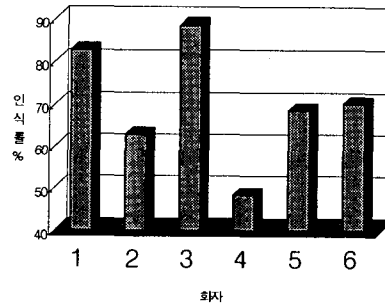
먼저, 기존에 일반적으로 사용되고 있는 멜 캡스트럼과 회귀계수를 사용한 경우와 음성 특징 파라미터를 확장한 경우, 그리고 숫자음의 음성학적 특징인 연음 현상 및 경음화 현상을 고려한 경우에 대하여 4연속 숫자음 인식을 위한 FSN을 구성하여 인식실험을 수행하였다.

멜 캡스트럼과 회귀계수를 사용한 경우의 결과를 그림 6에 나타내었다.



<그림 6> 기존의 방법을 사용한 4연속 숫자음 인식을 그림에서와 같이 5인의 남성화자 평균 멜 캡스트럼의 경우 55.4%, 멜 캡스트럼과 회귀계수의 경우 67.4%의 인식을 얻었다.

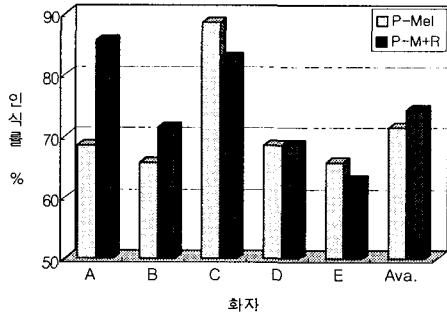
기존의 방법에 연속 숫자음에서 나타나는 음성학적 특징을 적용한 경우의 인식 수행 결과를 그림 7에 나타내었다.



<그림 7> 음성학적 특징을 적용한 4연속 숫자음 인식을

위 그림에서 기존의 방법에 연음 현상 및 경음화 현상과 같은 음성학적 특징을 적용한 경우 70.28%로 향상된 인식을 얻어 숫자음 인식에 있어 음성학적 특징이 유효함을 확인할 수 있었다. 그림에서 화자 4의 경우 발성이 다른 화자에 비하여 발성변화가 심하여 낮은 결과를 보였다.

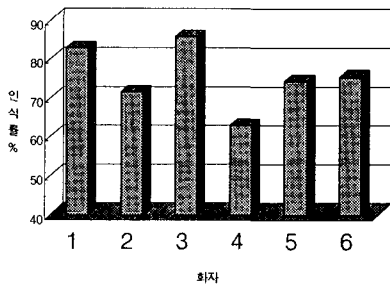
다음으로 기존의 방법에 에너지에 대수를 취한 특징값을 사용하여 음성의 특징 파라미터를 확장한 경우의 인식 실험을 수행하였다. 그 결과를 그림 8에 나타내었다.



<그림 8> 특징 파라미터를 확장한 4연속 숫자음 인식률

인식 실험 결과 5인 화자 평균 대수 에너지와 멜 캡스트럼의 경우 71.4%, 회귀계수와 함께 사용한 경우 74.3%의 인식률을 보여, 에너지값에 대수를 취한 음성 특징 파라미터 확장의 유효성을 확인하였다.

그림 9에 음성 특징 파라미터를 확장하고, 연음 현상 및 경음화 현상 등과 같은 음성학적 특징까지 적용한 경우에 대하여 인식 실험을 수행한 결과를 나타내었다.



<그림 9> 특징 파라미터 확장과 음성학적 특징을 적용한 4연속 숫자음 인식률

음성 특징 파라미터를 확장하고 음성학적 특징을 적용한 경우 위 그림에서와 같이 5인의 남성화자 평균 75.42%의 인식률을 보였다.

이상의 결과로부터 음성의 특징 파라미터를 확장하고 음성학적 규칙을 적용한 경우 향상된 인식률을 얻어 한국어 연속 숫자음 인식의 실용화를 위해 에너지에 대수를 취한 특징값을 사용한 음성 특징 파라미터의 확장과 연음 현상 및 경음화 현상과 같은 음성학적 규칙의 적용이 유효함을 확인하였다.

## 6. 결 론

본 논문에서는 인식시스템에서 일반적으로 인식률이 낮은 연속 숫자음의 실용화를 위한 인식률 향상에 관한 연구로서 4연속 숫자음을 대상으로 음성 특징 파라미터를 확장하고 연음 현상과 경음화 현상 등과 같은 음성학적 특징을 고려하여 숫자음에 강건한 모델을 작성하는 방법을 제안하고 인식실험을 통하여 그 유효성을 확인하였다.

인식 실험 결과, 기존의 방법에서 멜 캡스트럼을 사용한 경우 55.4%, 멜 캡스트럼과 회귀계수를 사용한 경우 67.4%의 인식률을 얻었다. 여기에 음성학적 특징을 적용한 경우 70.28%로 다소 향상된 인식률을 얻었다.

그리고 음성 특징 파라미터를 확장한 경우 멜 캡스트럼을 사용한 경우 71.4%, 회귀계수까지 함께 사용한 경우 74.3%의 인식률을 얻어, 기존의 방법보다 크게 향상되어 에너지에 대수를 취한 음성의 특징값의 유효성을 확인할 수 있었다.

또한 음성 특징 파라미터를 확장하고 음성학적 규칙까지 함께 적용한 경우 5인 남성화자 평균 75.42%로 비교적 높은 인식률을 보여 제안된 방법이 연속 숫자음 인식의 실용화에 유효함을 확인하였다.

향후 이 실험 결과를 카드결제 시스템이나 전화번호 시스템 등 연속 숫자음이 필요한 다양한 분야에 적용할 수 있을 것으로 기대된다.

## 참 고 문 헌

- [1] 정현열, "음성인식 연구의 국내외 현황과 전망," 제15회 음성통신 및 신호처리 워크샵 논문집, pp. 23-30, 1998. 8.
- [2] 김순협, "음성인식의 현황과 최근 연구 동향," 2000년도 한국음향학회 학술발표대회 논문집, Vol. 19, No. 2(s), 2000. 11.
- [3] 이호영, "국어 음성학," 태학사 1996.
- [4] 우인봉, 이강성, 김순협, "HMM의 교정학습과 후처리를 이용한 연결 숫자음 인식에 관한 연구," 제11회 음성통신 및 신호처리 워크샵 논문집, pp.161-165, 1994.
- [5] M. K. Ravishankar, "Efficient Algorithms for Speech Recognition," Ph. D Thesis, Carnegie

Mellon University, 1996.

2000년 3월 ~ 현재 대구과학대학 정보전자통신계열  
전임강사

[6] 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.

관심분야 : 음성분석 및 인식, 디지털 신호처리

[7] X. D. Huang, Y. Ariki, M. A. Jack, "Hidden Markov Models for Speech Recognition", Edinburgh Univ., 1990.

[8] Rabiner, Juang, "Fundamentals of Speech Recognition," Prentice-Hall International, Inc, 1993.

[9] 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.

[10] 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.



**김 득 수 (Deok-Soo Kim)**  
1977년 2월 한양대학교 전자공학과 (공학사)  
1987년 2월 영남대학교 대학원 전자공학과 (전자계산기전공 공학석사)  
1993년 8월 영남대학교 대학원 전자공학과 (전자계산기전공 박사수료)

1983년 3월 ~ 현재 대구공업대학 컴퓨터정보과 부교수  
관심분야 : 디지털 신호처리, 마이크로프로세서 응용



**황 철 준 (Chul-Joon Hwang)**  
1996년 2월 영남대학교 전자공학과 (공학사)  
1998년 2월 영남대학교 대학원 전자공학과 (정보통신전공 공학석사)  
1998년 3월 ~ 현재 영남대학교 대학원 전자공학과 (정보통신전공 박사수료)