

CHMM을 이용한 전화번호 인식에 관한 연구

이성권*, 김순협*

* 광운대학교 컴퓨터공학과

A Study of Telephone Digit Recognition Using CHMM

Seong-Kwon Lee*, Soon-Hyob Kim*

* Kwangwoon University

joyer@explore.kwangwoon.ac.kr

요 약

본 연구는 음소 단위의 CHMM(Continuous Hidden Markov Model)을 이용한 연속어 숫자음 인식에 관한 내용이다. ETRI 445 데이터를 사용하여 초기의 모델은 ML(Maximum Likelihood) 추정법을 이용하여 작성하였고 적응화를 위해 최대 사후 확률 추정법을 사용하였다. 또한 한국어 숫자음 음성의 음향학적 특성을 고려하여 발성 사전에 작성하였고 유절단위로 되어있는 한국어 숫자음의 모든 경우를 고려하여 복수개의 단어를 사전에 등록하였다. 또한 적응화 학습에 있어서 숫자음의 앞 뒤 모든 경우를 고려하여 작성한 21 종류의 7자리 전화 번호 숫자음 DB로 사용하였고 이의 효율성을 입증하기 위하여 ETRI에서 작성한 35종류의 4연속 숫자음 복복을 대상으로 인식실험을 수행하였다. 그결과 5인의 화자에 대하여 4연속 숫자음에 대하여 96%의 인식률을 보이고 있으며 7연속 숫자음에 대하여도 약 91%의 결과를 보여주고 있다. 또한 후처리를 두어 연음 현상으로 인한 오인식의 경우에 대해서도 약 2%의 인식율의 증가를 보여주었다.

I. 서론

음성인식을 통한 인간과 기계와의 의사소통이라는 주제는 지난 수십년 동안에 음성연구 관련 공학자와 과학자들에게 있어서 커다란 꿈이자 목표이었다. 이러한 꿈과 목표가 요즈음에 들어 하나씩 구현되어가고 있다. 컴퓨터의 태스크 톨 상에서 음성을 이용하여 응용 프로그램을 구동시키고, 대화형 음성인식 시스템을 이용하여 열차나 비행기의 티켓을 예매하는 등 음성 인식이 우리의 일상생활에 일부로서 활용되기 시작하고 있다.

또한 현대 사회가 국제적인 정보화 사회로 넘어 감에 따라 인간과 컴퓨터 간의 대화 단계를 넘어 언어가 서로 다른 통화자 간의 통신 내용을 자동적으로 쌍방의 언어로 변환 시켜주는 자동 통역 시스템의 개발에 대한 연구가 활발히 진행 중에 있다. 이러한 핵심 기술로서 음성인식 및 음성 이해, 자동 번역 및 음성 합성의 기술등을 들 수 있는데 그 중에서 숫자음에 대한 인식은 우리 생활의 각 분야에 두루 사용될 수 있기에 많은 데이터베이스의 구축과 더불어 연구가 활발히 진행 중에 있다. 신원조회에 쓰이는 주민등록 번호라든가 각종 신

CHMM을 이용한 전화번호 인식에 관한 연구

용카드의 비밀번호 등등 우리 주위에 숫자를 이용하는 경우가 많이 있다. 그 중에서도 우리 일상 생활에 아주 밀접한 전화를 사용할 때 음성을 사용할 수 있다면 많은 장점을 얻을 수 있을 것이다.

II. 음성 데이터베이스

국외의 경우 음성 데이터베이스가 이미 공동으로 이용 가능 하도록 구축되어 있으며 이를 CD-ROM으로 제작하여 유료로 배포하고 있는 실정이다. 미국의 경우 불특정화자 연속 숫자음 (Corpus(TIDIGITS))에는 326명의 77개 숫자음의 발성을 데이터 베이스화 하여 유료로 배포하고 있다. 여기에는 22개의 단독 숫자음 11개외에 2연, 3연, 4연, 5연, 7연 숫자음이 들어 있고 약 1000달러의 가격으로 판매하고 있으며 각종 학술 발표에 객관적으로 사용되고 있다. 프랑스의 경우에도 GRECP Project에 의해 이미 1980년대 후반에 숫자음에 대한 데이터베이스를 구축하였고 일본의 경우도 전자기술종합연구소, 동북대학, ATR 등을 중심으로 데이터베이스를 구축해 놓고 있는 실정이다. 그러나 국내의 경우 지금까지 인식, 합성에 사용되는 음성 데이터는 각자가 부분적으로 제작하여 사용하였고 외부에 공개하지 않았다. 따라서 데이터량 및 이용형태가 제한되고, 동시에 각 연구자가 발표한 인식 시스템의 성능 혹은 분석방식의 평가도 객관적인 평가가 이루어지고 있지 않은 실정이다. 본 연구에서는 초기 HMM 작성을 위한 음성 데이터를 한국 전자 통신 연구소에서 작성한 PBW(Phoneme Balanced Word) 445 단어 음성 데이터베이스 중 14인의 2회 발성중에서 1회분 총 6,230 단어를 수작업에 의해 이루어진 유사 음소단위 레이블링 정보를 이용하여 구성하였고 직음화와 인식에 있어서는 무방향 방음부스에서 작성된 ETRI 445DB와는 달리 일반적인 연구실 환경에서 남성화자 5인에 의해 4회씩 발성된 데이터와 적응화를 위한 남성 화자 8명의 각 2회씩 발성된 데이터를 사용하여 실험하였다. 마이크로는 berydynamic 사의 다이내믹 마이크를 사용하였고 비교하여 SONY사의 콘덴서 마이크를 최종 인식 실험에 사

용하였다. 본 연구실에서 앞 뒤 숫자음 음절을 고려한 21종류의 연속 숫자음을 화자 적응용으로 실험하였고 이의 적합성을 입증하기 위해 ETRI에서 만든 4연속 숫자음에 실험하였다.

표 1. 21종류의 적응화용 연속 숫자음 표

	전화번호		전화번호
1	512-0257	12	270-9483
2	630-1349	13	396-0011
3	745-6780	14	408-6281
4	826-9318	15	689-6542
5	904-0371	16	209-1921
6	910-2388	17	147-3324
7	843-4616	18	986-5066
8	729-5522	19	569-1775
9	607-7641	20	795-9785
10	358-8736	21	448-1234
11	153-0599		

표 2. 35 종류의 인식용 연속 숫자음 표

ETRI 4 연속 숫자음(35단어)									
1	0287	8	5732	15	9601	22	4156	29	1199
2	1398	9	6843	16	0712	23	5267	30	6633
3	2409	10	7954	17	1823	24	6378	31	8877
4	3510	11	8065	18	2934	25	7489	32	2244
5	4621	12	9176	19	3045	26	8590	33	5500
6	6972	13	5861	20	3649	27	0316	34	7083
7	8194	14	9205	21	1427	28	2538	35	4750

III. 음성의 특징 파라미터 추출

실험에 사용된 데이터는 A/D 변환된 후 Pre-emphasis 필터를 통과한후 16ms 길이의 해밍윈도우를 거쳐 구간 분석된다. 이때 각 구간은 5ms마다 중첩된다. 여기서 자기 상관 계수를 구하고 20차 LPC 계수를 구한 후 14차 LPC 퀘스트럼 계수를 구하고 나서 10차의 회귀계수를 구하여 특징 파라미터로 한다. 여기에 음소 지속 시간정보를 추가로 이용하였다. 다음 표 3에 음성의 분

석 조건을 나타내었다.

표 3. 음성 데이터 분석

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

IV. 한국어 숫자음 발음 사전 구성

한국어 연속 숫자음인식, 특별히 전화번호 인식에 사용되는 숫자음은 특별히 숫자음간 혼동이 많은 단음절로 구성되어 있다. 다섯 개의 모음과 다섯 종류의 자음만으로 구성되어 있으며 연음이나 동시 조음 효과에 의하여 '일'과 '이', '일'과 '칠', '삼'과 '사', '오'와 '구' 그리고 '구'와 '꿈'이 빈번한 오인식을 일으킨다. 기본 10개의 음절 외에도 동시 조음 효과에 의하여 다음 표 2에서와 같은 음절이 상호 음절의 연결에 의하여 발생된다. 또한 이러한 점을 고려하여 숫자음 발음 사전에 이러한 음절을 포함하여 구성하였다.

표 4. 동시 조음효과에 의한 음절 발생

대상 숫자	조음결과
0(공), 9	유성음화('ㄱ') 경음화(공, 구)
0(영)	령 녕 경, 녕
1, 2	릴, 리 밀, 미
3, 4	경음화(쌈, 싸)
7, 8	리(l -> r)
6	유성음화('ㄱ') 륙(rjug) 륙(ljug) 늑(njug) 움(jung) 룡(rjung) 룡(ljung) 늣(njung)

V. 전화번호 인식 시스템 구성

아래 그림 1은 전체 시스템의 구성도를 나타낸다. 초기의 HMM의 모델작성은 ML 추정법을 사용하였으며, 적응화에는 MAP 추정법을 사용한다. 인식의 기본 단위로는 묵음을 포함한 48개의 유사 음소단위를 이용한다.

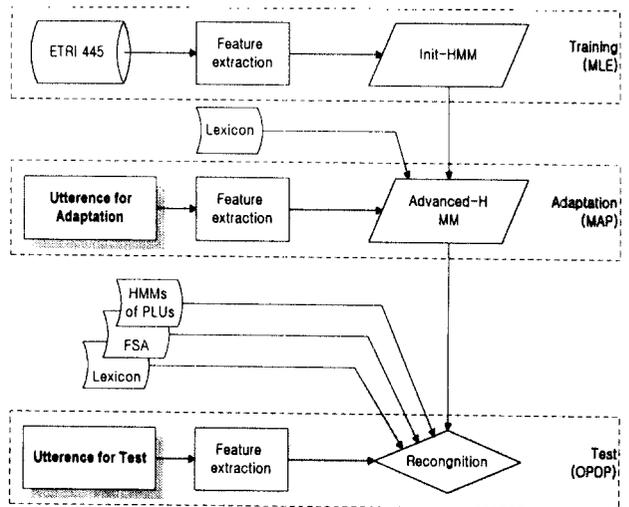


그림 1. 전체 인식 시스템

표 5. 유사음소 단위의 로마체 심볼

모음	aa /아/	axr /어/	ao /오/	uh /우/
	u /으/	ih /이/	ae /에/	eh /예/
	ja /야/	jv /여/	jo /요/	ju /유/
	wa /와/	wv /워/	we /외/	wee /웨, 왜/
	wi /위/	je /예/		wii /외/
자음	b~ /ㅂ/	d~ /ㄷ/	g~ /ㄱ/	z~ /ㅈ/
	bb /ㅃ/	dd /ㄸ/	gg /ㄲ/	zz /ㅉ/
	p /ㅍ/	t /ㅌ/	k /ㅋ/	ch /ㅊ/
	s /ㅅ/	ss /ㅆ/	hh~ /ㅎ/	r /ㄹ/
	n /ㄴ/	m /ㅁ/	ng /ㅇ/	
첫음절	b /ㅂ/	d /ㄷ/	g /ㄱ/	z /ㅈ/
	hh /ㅎ/			
종성	bp /ㅃ/	dp /ㄸ/	gp /ㄲ/	l /ㄹ/
묵음	SIL			

본 연구에서 이용한 HMM 모델은 4상태 3출력 1 혼합

CHMM을 이용한 전화번호 인식에 관한 연구

분포의 연속출력확률 이산 시간제어 HMM이다.

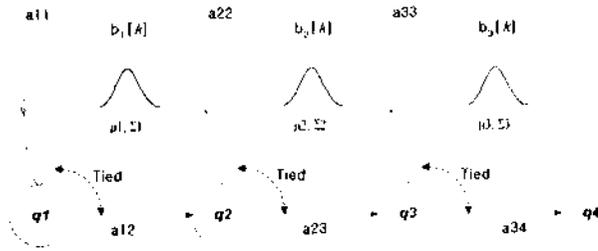


그림 2. 4상태 3출력 1 혼합분포 CHMM

VI. 실험 및 검토

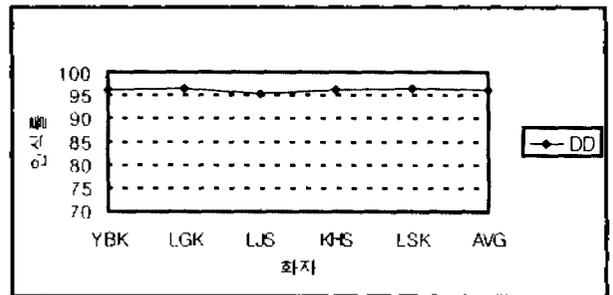
본 연구에서는 앞에서 언급했던 우리말의 숫자음 발성에 따른 음향학적 고찰을 통하여 숫자음 조합에서 발생할 수 있는 음운 현상을 고려하여 발음 사전에 등록하였다. 숫자음의 모든 경우를 고려하여 작성한 21종류의 7자리 연속 숫자음 DB로 적용화를 하였고 이의 효율성을 입증하기 위하여 ETRI에서 작성한 35종류의 4연속 숫자음 목록을 대상으로 인식 실험을 하였다. 그리하여 연구실 환경에서 실시간으로 ETRI 4연속 숫자음에 대하여 약 96%의 인식율을 얻음으로 발음 사전의 구축에 대한 효율성을 증명하였다. 또한 우리말의 '에'를 포함한 7연속 숫자음에 대하여도 실험한 결과 약 91%의 인식율을 보이고 있다.

VII. 결론

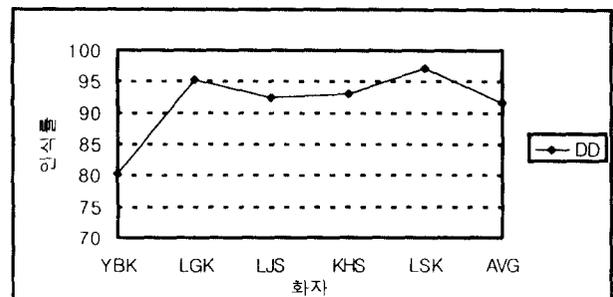
7연속 숫자음에 대해서는 YBK특정 화자에 대해서만 오차가 크게 났을 뿐 대부분 4연속 숫자음과 비교하여 높은 인식율을 보이고 있다. 그리고 숫자음에 대해서는 몇몇 특정 음소에 대해서만 학습이 이뤄지므로 다른 특정 응용분야에 비해 학습이 잘 되는 것을 볼 수 있다. 앞으로 더욱 많은 화자에 의한 학습이 이루어지고 지역 번호에로의 확장이나 더 많은 연속 숫자음으로의 확장성을 고려한다면 응용 분야에 다양하게 적용 가능할 것으로 보인다.

[참고문헌]

- [1] John R. Deller, Jr. John G. Proakis and John H. L. Hansen, "Discrete Time Processing of Speech Signals", Macmillan Publishing Company, 1993
- [2] SAEED V.VASEGHI "Advanced Signal Processing and Digital Noise Reduction", pp.111-139 Wiley Teubner
- [3] J.K. Baker, "The DRAGON System - An Overview", IEEE Trans. Acoust. Speech, Signal Processing, ASSP-23(1), pp. 24-29, February 1975
- [4] Kai-Fu Lee, Raj Reddy, "Automatic Speech Recognition", KLUWER ACADEMIC PUBLISHERS
- [5] MIN ZHOU "A Study on Stochastic Models for Spoken Language" Ph.D Thesis Toyohashi Univ. 1996
- [6] M. K. Ravishankar, "Efficient algorithms for speech recognition", Ph.D. thesis, Computer Science Department, Carnegie Mellon University, May 1996
- [7] 오 세진, "문백자유문법을 이용한 한국어 연속음성 인식에 관한 연구", 영남대학교 대학원 석사학위 논문, 1997, 12



(a) 4 연속 숫자음 인식



(b) 7 연속 숫자음 인식

그림 3. 연속 숫자음 인식