1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 2(s) 호

Voice Dialing Sysytem을 위한 음성인식

이성권*, 김순협*

* 광운대학교 컴퓨터공학과

A Study on the Speech Recognition For the Voice Dialing System

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 Voice Dialing System을 위한 인속 음성인식에 관한 내용이다. 연구실 환경에서 음성으로 전화를 걸기 위하여 전국 지역명과 연속 숫자음 인식을 수행하였다. ETRI 445 데이터를 사용하여 초기의 모델은 ML(Maximum Likelihood) 추정법을 이용하여 작성하였고 적응화를 위해 최대 사후 확률 추정법을 사용하였다. 음성으로 다이얼링을 수행하기 위하여 문맥자유문법을 이용하여 제한적이나마 대화체문장으로 수행할 수 있도록 하였다. 그리하여 숫자음에 대하여 5인의 화자에 대하여 4연속 숫자음에 대하여 96%의 인식률을 보이고 있으며 7연속 숫자음에 대하여도 약 91%의 결과를 보여주고 있다. 문장으로도 음성 다이얼링을 수행하였을 경우 문장내에 단어와 숫자음에 대하여 약 80%의 인식률을 보였다.

I. 서론

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 시키주는 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 누루 사용될 수 있기에 많은 데이터베이스의 구축과 더불어 연구가 활발히 진행 중에 있다. 특별히 음성을 이용한 다이얼링을 자동으로 수행할 수 있다면 더 없이 편리함을 가져다 줄 것이다. 이 경우에는 연속 숫자음 인식이라는 어려움이 있지만 이에 대한 연구가 꾸준히 전개되어 오고 있다

II. 음성 데이타베이스

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 적응화와 인식에 있어서는 무반향 방음부스에서 작성된 ETRI 445DB 와는 달리 일반적인 연구실 환경에서 남성화자 5인에 의해 4회씩 발성된 데이터와 적응화를 위한 남성 화자 8명의 각 2회씩 발성된 데이터를 사용하여 실험하였다. 마이크는 SONY사의 콘덴서 마이크를 사용하였다. 인식 대상 단위는 전국 지역명 246개의 단어와 본 연구실에서 안 된 숫자음 음절을 고려한 21종류의 인속 숫자음을 화자 적응용으로 실험하였고 이의 적합성을 입증하기 위해 ETRI에서 만든 4연속 숫자음에 인식 실험하였다.

표 1. 35 종류의 4 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 켑스트럼 계수를 구하고 다시 10차의 회귀계수를 구하여 특징 파라미터로 한다. 여기에 음소 지속 시간정보를 추가로 이용하였다.

IV. 숫자음 발음 사전 구성

연속음성인식 중에서 특별히 숫자음 인식, 전화걸기에 사용되는 숫자음은 특별히 숫자음간 혼동이 많은 단음절로 구성되어 있다. 다섯개의 모음과다섯 종류의 자음군으로 구성되어 있으며 연음이나 동시 조음 효과에 의하여 '일'과 '이', '일'과 '칠', '삼'과 '사', '오'와 '구' 그리고 '구'와 '공'이 빈번한 오인식을 일으킨다. 기본 10개의 음절 외에도 동시 조음 효과에 의하여 다음 표 2 에서와 같은 음절이 상호 음절의 연결에 의하여 발생된다. 또한 이러한 점을 고려하여 숫자음 발성 사전에 이러한 음절을 포함하여 구성하였다.

표 2. 동시 조음효과에 의한 음절 발생

대상 숫자	조음결과
0(공), 9	유성음화 ('ㄱ') 경음화(꽁,꾸)
0(영)	렁 녕 경, 넝
1, 2	린, 리 밀, 미
3, 4	경음화(쌈,싸)
7, 8	ㄹ(l -> r)
6	유성음화('ㄱ') 륙(rjug) 륵(ljug) 뉵(njug) 융(jung) 륭(rjung) 륭(ljung) 늉(njung)

V. Voice Dialing System 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.

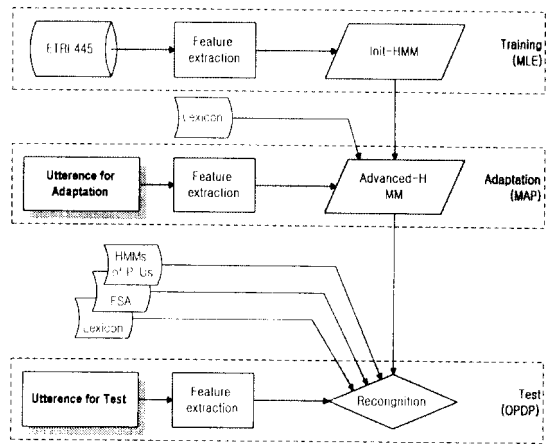


그림 1. 전체 인식 시스템

본 연구에서 이용한 HMM 모델은 4상태 3출력 1 혼합분포의 연속출력확률 이산 시간제어 HM

M이다.

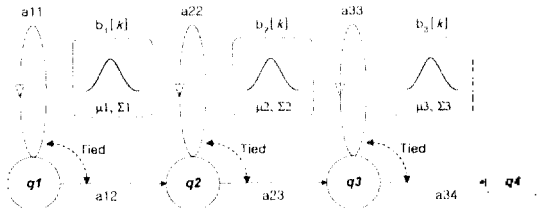
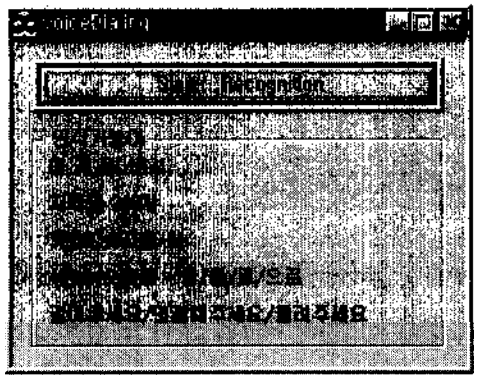
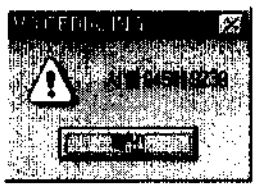


그림 2. 4상태 3출력 1 혼합분포 CHMM

또한 다음 그림3에 인식 프로그램을 수행했을 경우의 그림을 볼 수 있으며 그 결과도 메시지 박스에 출력됨을 볼 수 있다.



(a) 초기화면



(b) 인식 결과

그림 3. Voice Dialing System 시뮬레이션

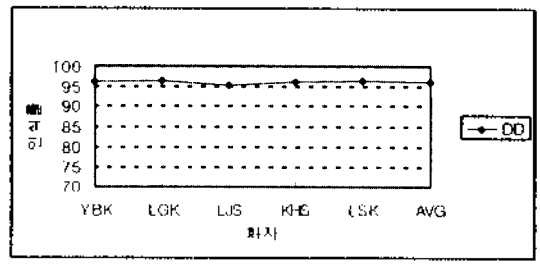
VI. 실험 및 검토

본 연구에서는 앞에서 언급했던 우리말의 숫자음 발성에 따른 음향학적 고찰을 통하여 숫자음 조합에서 발생될 수 있는 음운 현상을 고려하여 발음 사전에 등록하였고 246개의 지역명에대서도 조음현상을 고려한 복수개의 단어를 발음 사전에 등록하여 인식 실험을 수행 하였다. 숫자음의 모든 경우를 고려하여 직성한 21종류의 7자리 연속 숫자음 DB로 적응화를 하였고 이의 효율성을 입증하기 위하여 ETRI에서 작성한 35종류의 4연속 숫자음 복록을 대상으로 인식 실험을 하였다. 그리하여 연구실 환경에서 실시간으로 ETRI 4연속 숫자음에 대하여 약 96%의 인식율을 얻음으로 발성 사전의 구축에 대한 효율성을 증명하였다. 또한 우리말의 '에'를 포함한 7연속 숫자음에 대하여도 실험한 결과 약 91%의 인식률을 보이고 있다. 또한 지역명과 결합하여 간단한 문장으로 인식 실험을 수행하였을 경우엔 인식률이 낮게 나오지만 앞으로 간단한 대화체 음성으로도 음성 다이얼링을 수행 할 수 있는 길을 모색하였다는데 중점을 두었다. 다음은 인식에 사용한 간단한 문장들의 예이다.

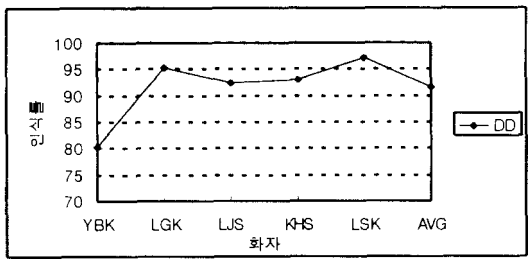
1. 비서 직성 745에 6780을 걸어주세요
2. 남양 826에 9318
3. 저 남해 904에 0371을 연결해주세요
4. 동두천 910에 2388을 눌려주세요
5. 옵 문산 843에 4616을 걸어줘
6. 봉화 729에 5522를 부탁해요

VII. 결론

7 연속 숫자음에 대해서는 특성 화자에 대해서만 오차가 크게 났을 뿐 대부분 4연속 숫자음과 비교하여 많은 차이를 보이지 않고 있다. 그리고 숫자음에 대해서는 몇몇 특정 음소에 대해서만 학습이 이뤄지므로 다른 특정 응용분야에 비해 학습이 잘 되는 것을 볼 수 있다. 단이 246개에 대해서도 많은 양의 데이터를 가지고 학습하면 더욱 좋은 결과를 얻을 수 잇을 거으로 사료된다. 문장 단위로의 인식은 좀더 다양한 문장으로서 인식이 가능하여야 겠지만 전화걸기에 사용되는 문장은 다양하지가 않으므로 추후 더욱 연구 되어야 할 것으로 본다. 앞으로 더욱 많은 화자에 의한 학습이 이루어지고 지역 번호에로의 확장을 고려한다면 응용 분야에 다양히세 적용 가능할 것으로 보인다.



(a) 4 연속 숫자음 인식



(b) 7 연속 숫자음 인식

그림 4. 연속 숫자음 인식

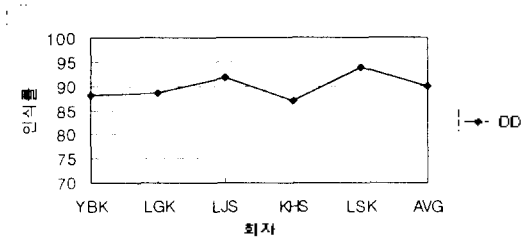


그림 5. 단어 인식

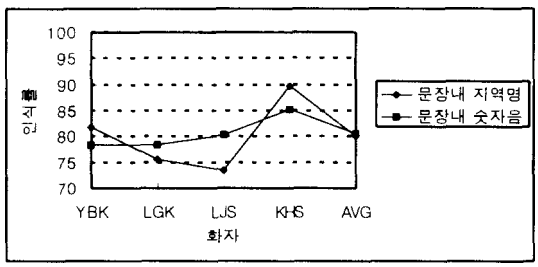


그림 6. 문장 내 단어, 숫자음 인식

[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, "문맥자유문법을 이용한 한국어 연속음성인식에 관한 연구", 영남대학교 대학원 석사학위 논문, 1997, 12

[8] Sadaoki Furui "Digital Speech Processing, Synthesis, and Recognition" , MARCEL DEKKER, INC.1991

[9] Douglas O'sShaughnessy "Speech Communication Human and Machine"Addison Wesley Publishing Company. 1987