1999년도 한국음향학회 학술발표대회 논문집 제18권 제1(s)호

이동전화를 위한 단어 인식기의 성능평가

김민정*, 황철준*, 정호열*, 정현열*
* 영남대학교 정보통신공학과

Evaluation of Word Recognition System For Mobile Telephone

Min-Jung Kim*, Cheol-Jun Hwang*, Ho-Youl Chung*, Hyun-Yeol Chung*
*** Department of Information and Communication Eng., Yeungnam University***
{kmj, hcj, hoyoul, chy}@speech.yeungnam.ac.kr

요 약

본 논문에서는 음성에의해 구동되는 이동전화를 구현하기 위한 기초 실험으로서, 이동전화상에서 많이 사용되는 단어 데이터를 직접 채록하여 단어 인식 실험을 수행하여 인식기의 성능을 평가하였다.

인식 실험에 사용된 단어 데이터베이스는 서울 화자 360명(남성화자 180명, 여성화자 180명), 경상도 화자 240명(남성화자 120명, 여성화자 120명)으로 구성된 600명의 발성을 이용하여 구성하였다. 발성 단어는 이동전화에 주로 사용되는 중요 기능과 제어 단어, 그리고 숫자음을 포함한 55개 단어로 구성되었으며, 각 화자가 3회씩 발성하였다. 데이터의 채집환경은 잡음이 다소 있는 사무실환경이며, 샘플링율은 8kHz였다.

인식의 기본단위는 48개의 유사음소단위(Phoneme Like Unit ; PLU)를 사용하였으며, 정적 특징으로 멜캡스트럼과 동적특징으로 회귀계수를 특징 파라미터로 사용하였다. 인식실험에서는 OPDP(One Pass Dynamic Programming)알고리즘을 사용하였다.

인식실험을 위한 모델은 각 지역에 따라 학습을 수행한 모델과, 지역에 상관없이 학습한 모델을 만들었으며, 기존의 16kHz의 초기 모델에 8kHz로 채집된 데이터를 적응화시키는 방법을 이용하여 학습을 수행하였다.

인식실험에 있어서는 각 지역별 모델과 지역에 관계없이 학습한 모델에 대하여, 각 지역별로, 그리고 지역에 관계없이 평가용 데이터로 인식실험을 수행하였다. 인식실험 결과, 90%이상의 비교적 높은 인식률을 얻어 인식 시스템 성능의 유효성을 확인할 수 있었다.

I. 서 론

음성인식에 대한 연구는 외국에서는 수 십년 전부터 연구가 수행되었으며, 우리 나라의 경우도 약 20여년전부터 연구가 수행되어 왔다.

현재, 외국의 경우에는 숫자음이나 연속음성인식을 이용한 시스템이 상용되고 있으며, 우리 나라의 경우에는 일부 제한된 영역의 단어 인식 시스템은 상용화되고 있다. 한편, 우리 나라의 이동전화 보급은 세계적 수준이며, 향후 더욱 증가할 것으로 예상된다. 이렇게 이동전화의 사용이 늘어남에 따라 사용자들은 좀더 쉽고 편리하게 보다더 안전하게 이동전화를 사용하고자 하는 욕구가 증가하고 있다. 이와같은 요구에 부응하여 최근 음성인식 기능을 이용한 다이얼링 방법이 채용되고 있다.

그러나, 현재까지 이동전화에 채용된 음성인식 다이얼링 시스템은 인식 단어의 수가 극히 제한되어 있고 인식률도 낮아서 사용자의 요구를 충족시키지 못해 인기를 잃어가고 있는 실정이다.

따라서, 본 연구에서는 사용자 요구에 부응하는 음성인식기술을 이용한 이동전화를 위한 기초 연구로서, 현재까지 이동전화에 이용되고 있는 단어 및 향후 이동전화상에서 이용될 것으로 기대되는 단어들을 데이터베이스화하여 단어 인식 실험을 수행하고자 한다.

인식 실험에 사용되는 단어데이터는 이동전화의 작동에 필요한 중요기능과 제어단어, 그리고 숫자음 등이며, 55개의 단어로 구성되어 있고, 8 kHz로 샘플링되었다. 음성의 특징 파라미터는 정적 특징으로 멜켑스트럼과 동적특징으로 회귀 계수를 추출하여 사용한다.

인식의 기본단위로서는 48개의 연속분포 HMM 유사음소단위(PLU)를 사용한다. 인식실험은 유한상태 오토

마타(FSA)에 의한 구문제어를 통한 OPDP[1,2,3]법으로 인식실험을 수행하였다.

본 논문의 구성은 다음과 같다. II장에서는 본 연구에 사용된 음성자료 및 분석방법, III장에서는 모델작성과 인식방법에 대해 설명하고, IV장에서는 인식실험 및 고찰, 마지막으로 V장에서 결론을 맺는다.

II. 음성자료및 분석방법

2.1 음성자료

본 연구를 위한 음성자료로는 서울 남성화자 180명, 서울 여성화자 180명, 대구 남성화자 120명, 대구 여성화자 120명, 총 600명이 발성하였다. 학습용 자료와 평가용 자료의 구성을 표 1에 나타낸다.

표 1. 학습용 자료와 평가용 자료.

학습을 위한 자료	평가를 위한 자료
서울 남성화자 100명	서울 남성화자 80명
서울 여성화자 100명	서울 여성화자 80명
대구 남성화자 100명	대구 남성화자 20명
대구 여성화자 100명	대구 여성화자 20명
서울 남성화자 50명 + 대구 남성화자 50명	학습에 참여하지 않은 화자
서울 여성화자 50명 + 대구여성화자 50명	

이동전화를 사용할 때 많이 쓰이는 중요기능과 제어단어, 그리고 숫자음을 포함한 55개 단어로 구성되었으며 표 2에 나타내었다.

각 화자가 3회씩 발성하였고, 데이터의 채집환경은 잡음이 섞인 사무실환경이며, 8kHz로 샘플링되었다.

2.2 분석방법

음성자료의 분석은 표 3과 같이 음성 데이터를 샘플링 주파수 8kHz, 양자화 정도 16Bits A/D 변환기를 통해 이산 데이터로 변화되고 Pre-emphasis 필터를 통과한 후, 16ms(128 points)길이 헤밍 윈도우를 사용하여 5ms(40 points)씩 쉬프트 시키면서 분석된다. 이로부터 14차 LPC 켑스트럼 계수를 구하고, 정적 특징으로 10차의 멜켑스트럼을 구하여 특징파라미터로 사용한다. 또한 이로부터 10차의 회귀계수를 추출하여 동적 특징파라미터로 사용한다. 인식실험에서는 10차의 멜-켑스트럼과 10차의 회귀계수를 이용한다.

표 2. 이동전화에서 사용되는 단어.

1	추가	20	되감기	39	별표
2	리스트	21	빨리감기	40	전자우편
3	전호번호부	22	첸슬	41	읽기
4	다음	23	저장	42	답신
5	취소	24	이전	43	전체답신
6	삭제	25	공	44	앞으로
7	발신	26	영	45	전송
8	지움	27	일	46	통화
9	종료	28	하나	47	검색
10	아니오	29	이	48	메뉴
11	예	30	둘	49	정지
12	네	31	삼	50	녹음
13	재발신	32	사	51	인사말변경
14	응답	33	오	52	전화정보
15	번호	34	육	53	통화기록
16	음성메모	35	칠	54	비밀기능
17	재생	36	팔	55	경보기능
18	스톱	37	구		
19	일시정지	38	우물정자		

표 3. 음성자료의 분석조건.

Speech Data	
Sampling frequency	8khz
Resolution	16bits
Hamming window	16ms (128points)
Frame rate	5ms (40points)
Analysis	14order LPC analysis
Static Feature parameters	10order Mel-Cep. coeff.
Dynamic Feature parameters	10order Regressive coeff.

III. 모델작성과 인식방법

3.1 모델작성

일반적인 음성인식에서는 모델의 작성을 위하여 음성데이터를 Labeling한 후, 인식 파라미터로서 멜-켑스트럼과 회귀계수를 추출한 다음 학습단계를 거쳐 모델을 만든다. 그러나, 이러한 방법은 실제 응용제품에 적용하기에는 사용상의 번거로움이나 응용제품의 메모리의 증가로 인한 cost의 상승 등으로 인하여 적용하기에는 무리가 있다. 따라서, 본 연구에서는 모델의 작성을 위하여 특징추출과정, Labeling과정, 그리고 학습과정을 거치지 않고, 16kHz의 초기 HMM모델에 8kHz의 음성데이터를 직접 적응화하는 방법을 이용하였다. 이렇게 함으로서, 이동전화의 실제 사용시에 적응화과정이 필요한 경우 부가적인 학습단계나 기타 특징추출단계를 거치지 않고

도 바로 모델을 만들 수 있는 이점이 있다.

3.2 음소 모델

HMM(Hidden Markov Model)은 출력확률의 분포에 따라 크게 이산분포 HMM과 연속분포 HMM으로 분류한다. DHMM에서는 추출된 음성특징 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 작성한 것을 적응화과정에 이용한다. 이때 CHMM 음소모델의 구조는 4상태 1혼합을 사용한다. 그림 1에 본 연구에서 사용한 연속분포 HMM 모델의 구성을 나타내었다.

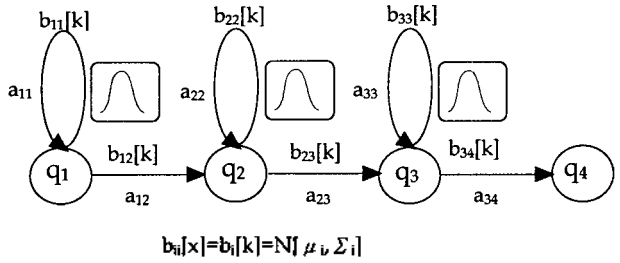

그림 1. 연속분포 HMM의 구성.(4상태 1혼합)

3.3 인식 시스템

인식시스템은 표준패턴을 작성하기 위한 적응화 단계와 표준패턴과 입력패턴과의 유사도를 측정하여 최적의 상태열을 찾는 인식 단계로 구성되며 그림 2에 이동전화 단어 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.

이때, 인식단계에서 미리 작성한 단어사전과 유한상태 오토마타(FSA)에 의한 구문제어를 통하여 OPDP법으로 인식을 수행한다.

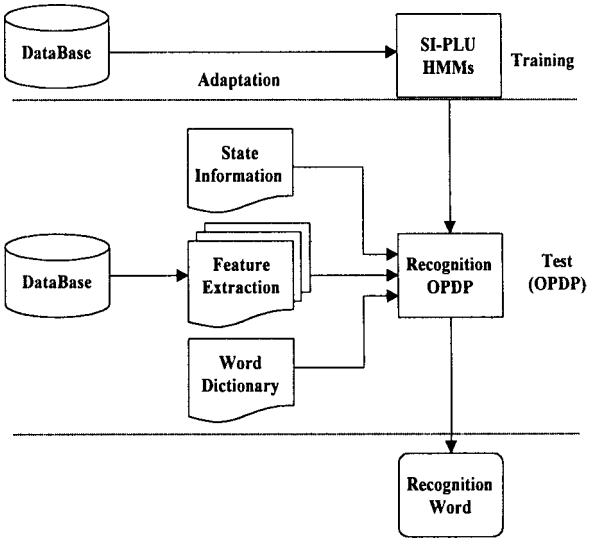

그림 2. 연속 숫자음 인식 시스템의 흐름도.

IV. 인식실험 및 고찰

인식실험에 있어서는 II절에서 설명한 음성자료를 사용하여 서울 남성화자 100명, 서울 여성화자 100명, 대구 남성화자 100명, 대구 여성화자 100명, 서울-대구 남성화자 100명, 서울-대구 여성화자 100명이 3회씩 발성한 단어들로 적응화 방법에 의해 학습을 수행하여 각각 표준패턴을 작성하였다. 이때, 지역이 다른 화자들로 구성된 경우, 비율은 1:1로 하였다. 적응화에 참여하지 않은 나머지 화자들이 발성한 단어를 평가용 자료로 사용하여 인식실험을 수행하였으며, 인식실험에서 특징파라미터로는 멜켑스트럼과 회귀계수를 사용하였고, 인식실험은 멜켑스트럼만 사용한 경우와 멜켑스트럼과 회귀계수를 함께 사용한 경우에 대하여 각 모델에 따라 인식실험을 수하였다. 표 4, 5, 6에 인식 실험결과를 나타내었다.

표 4. 학습용과 평가용이 같은 지역 화자의 인식률.

학습	평가	Mel	Mel + Rgc
서울남성화자 100명	서울남성화자 80명	88.41	93.8
서울여성화자 100명	서울여성화자 80명	88.52	92.88
대구남성화자 100명	대구남성화자 80명	85.99	91.44
대구여성화자 100명	대구여성화자 80명	86.66	92.9

표 5. 서울/대구화자로 구성된 모델의 지역별 인식률.

학습	평가	Mel	Mel + Rgc
서울대구남성 화자100명	서울남성화자 130명	88.48	93.77
〃	대구남성화자 70명	85.49	92.78
서울대구여성 화자100명	서울여성화자 130명	87.64	92.27
〃	대구여성화자 70명	85.34	90.9

표 6. 서울/대구화자로 구성된 모델의 서울화자 인식률.

학습	평가	Mel	Mel + Rgc
서울대구남성 화자100명	서울남성화자 80명	88.19	93.63
서울대구여성 화자100명	서울여성화자 80명	85.64	92.34

이상의 인식실험 결과, 멜켑스트럼만 사용한 경우보다 멜켑스트럼과 회귀계수를 함께 사용했을 때 더 높은 인식률을 보였으며, 서울 남성화자 100명으로 학습하고 서울 남성화자 80명으로 인식했을 경우가 93.8%로 가장 높게 나타났다.

V. 결 론

본 논문에서는 음성으로 구동하는 이동전화를 구현하기 위한 기초 실험으로서, 이동전화상에서 많이 사용되는 단어 데이터를 직접 채록하여 단어 인식 실험을 수행하여 인식기의 성능을 평가하였다.

모델작성시 적응화를 수행하는 방법으로 학습을 실시하여 초기 HMM모델을 작성하여 인식실험에 사용하였으며, 인식 실험에서는 멜켑스트럼과 회귀계수를 특징파라미터로 사용하였으며, 인식화자와 학습화자를 바꾸어 가며 인식실험을 수하였다.

인식실험결과, 학습화자와 인식화자가 같은 지역인 경우가 인식율이 높았으며, 가장 높은 인식률은 서울 남성화자 100명으로 학습한 모델에 서울 남성화자 80명을 인식한 경우가 93.8%로 가장 높게 나타났다.

인식률이 다소 낮은 원인으로서는 화자의 발성의 실수와 주변의 소음이 특히 심한 경우였다.

음성인식 기능을 가진 이동전화를 구현하기 위해서는 보다 고정도의 인식성능이 필요하며, 향후 다양한 방법으로 인식률 향상을 위한 연구를 수행하고자 한다.

참 고 문 헌

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, "連續音聲認識システムにおけるHMMの話者適應化に關する硏究," 修士學位論文, 1993.
3. 中川聖一, 甲裴充彦, "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.