

CHMM을 이용한 발매기 명령어의 음성인식에 관한 연구

A Study on the Speech Recognition for Commands of Ticketing Machine using CHMM

김범승[†] · 김순협^{*}

Beom-Seung Kim · Soon-Hyob Kim

Abstract This paper implemented a Speech Recognition System in order to recognize Commands of Ticketing Machine (314 station-names) at real-time using Continuous Hidden Markov Model. Used 39 MFCC at feature vectors and For the improvement of recognition rate composed 895 tied-state triphone models. System performance evaluation result of the multi-speaker-dependent recognition rate and the multi-speaker-independent recognition rate is 99.24% and 98.02% respectively. In the noisy environment the recognition rate is 93.91%.

Keywords : CHMM, Ticketing Machine, Speech Recognition of Station-Names

요 지 논문에서는 연속 HMM(Continuous Hidden Markov Model)을 이용하여 실시간으로 발매기 명령어(314개 역명)를 인식 할 수 있도록 음성인식 시스템을 구현하였다. 특징 벡터로 39 MFCC를 사용하였으며, 인식률 향상을 위하여 895개의 tied-state 트라이폰 음소 모델을 구성하였다. 시스템 성능 평가 결과 다중 화자 종속 인식률은 99.24%, 다중 화자 독립 인식률은 98.02%의 인식률을 나타내었으며, 실제 노이즈가 있는 환경에서 다중 화자 독립 실험의 경우 93.91%의 인식률을 나타내었다.

주 요 어 : 연속 HMM, 발매기, 역명 음성인식

1. 서 론

21세기 정보의 홍수 속에서 컴퓨터의 정보처리 속도는 빠르게 증가 하였다. 이러한 과정에서 정보기기의 성능은 더욱더 향상 되었고 유비쿼터스 환경에 적합하도록 설계되었다. 이러한 환경에서 정보기기는 소형화 되었고 이동성이 강조되었다. 이로 인하여 인간과 컴퓨터가 서로 대화할 수 있는 새로운 창구의 필요성이 절실히 대두되고 있다[1].

인간의 기본적인 의사소통 도구 중 가장 친화적이면서 편리한 것이 바로 음성이다. 음성은 전화선을 이용한 원격지 정보 입력 및 검색 등이 가능하며 정보 발생 및 처리에 따른 장소와 시간적인 제약을 극복할 수 있다는 장점이 있다.

또한 컴퓨터의 소형화에 따른 키보드를 대신할 입력 매체로 사용될 수 있다. 때문에 음성은 인간과 기계간의 인터페이스 수단으로써 여러 분야에서 응용될 수 있으며 현재 실용화되고 있는 국내외 응용 분야는 음성인식을 이용한 사무자동화 분야, 공장자동화 분야, 전화망을 이용한 분야, 의료 분야 및 멀티미디어 시스템의 조작용을 위한 인터페이스 등 다양한 분야에 응용되고 있거나 실용화 되었다.

본 논문에서는 지하철역 내에 설치되어 있는 발매기를 이용하는 고객들에게 음성을 통한 보다 편리한 인터페이스 제공을 위하여 CHMM을 이용하여 화자독립으로 발매기 명령어(314개 역명)를 인식할 수 있도록 음성인식 시스템을 구현하였으며 특징 벡터는 39 MFCC를 사용하였다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 CHMM의 기본개념에 대해서 설명하고 3장에서는 구현된 음성 인식 시스템에 대해서 소개하였고 4장에서는 실험 및 분석을 하였고 5장에서는 본 논문의 결론을 내렸다.

[†] 책임저자 : 정희원, 코레일 정보기술단 차장 · 광운대학교 박사과정
E-mail : bluedav@korail.com

TEL : (02)3149-3154 FAX : (02)3149-3195

^{*} 광운대학교 컴퓨터공학과 교수

2. CHMM(Continuous HMM)의 고찰

2.1 통계적 방법에 의한 음성인식

음성인식에는 음성·음향학적 지식을 이용하는 방법, 통계적 방법, 인공지능을 이용한 방법, 신경회로망을 이용한 방법 등이 널리 연구되었다. 통계적 음성인식에서는 아래 그림 1에서와 같이 학습단계에서는 추출된 특징벡터를 이용하여 기준이 되는 음향모델 또는 기준패턴을 만든 다음, 실제 인식단계에서는 Viterbi 알고리즘 또는 DTW와 같은 패턴 정합 알고리즘을 이용하여 미리 학습된 패턴 중에서 유사도가 가장 높은 것을 찾아서 인식결과로 출력한다. 이 중에 HMM은 통계적인 특성으로 인해서 음성 인식결과와의 이후처리에 수반될 언어처리나 의미처리 등의 통계적인 모델과 잘 융합하는 장점을 지니고 있어 최근에는 HMM을 대표로하는 통계적인 방법이 가장 널리 사용되고 있다.

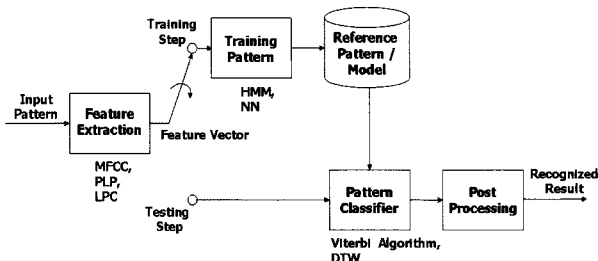


Fig. 1. Speech Recognition of statistical method

2.2 HMM(Hidden Markov Model)

HMM 알고리즘은 음성이 Markov process로 모델링 될 수 있다는 가정 하에 음성의 학습과정에서 Markov 모델의 파라미터를 구하여 기준 Markov 모델을 만들고 입력된 음성과 저장된 기준 Markov 모델을 비교하여 유사도가 가장 높은 기준 Markov 모델을 인식된 어휘로 결정한다[2]. HMM 알고리즘은 관측이 불가능한 프로세스를 관측이 가능한 다른 프로세스를 통해 추정하는 이중 확률처리 기법이며, 기준패턴을 음소, 음절 등으로 설정할 수 있고, 입력 음성으로 단어, 문장 등을 사용할 수 있기 때문에 화자독립, 연속 음성인식에 유리하다. HMM은 관측확률이 이산 분포 또는 연속분포를 갖는냐에 따라서 이산 HMM, 연속 HMM으로 나누어진다. 이산 HMM은 계산량이 적다는 이점이 있지만, 인식률이 상대적으로 낮고, 정교한 스무딩 과정이 필요하다는 약점으로 최근에는 임베디드 시스템을 위한 소형 인식기 이외에는 많이 사용되지 않고 있다. 이에 반해 연속 HMM은 계산량은 많지만 인식률이 높아 중·대형 인식기에 많이 사용되고 있다.

본 논문에서는 Gaussian Mixture 연속확률밀도에 기반한

연속 HMM을 이용한 자동발매기 명령어의 음성인식 시스템을 구현하였다.

2.3 연속 HMM

연속 HMM에서는 음성신호에서 추출한 특징벡터를 그대로 사용한다. 모델의 파라미터를 추정하기 위하여 관측된 신호들에 대한 Maximum Likelihood를 구하기 위한 방법으로 Gaussian Mixture Model을 사용한다. 연속 HMM의 경우 상태 j , 시간 t 에서 입력벡터 O_t 를 관측할 확률은 다음 식 (1)과 같이 가우시안 혼합모델(Gaussian Mixture Model, GMM)로 표현된다[3].

$$b_j(O_t) = \sum_{k=1}^M c_{jk} N(O_t, \mu_{jk}, U_{jk}), 1 \leq j \leq N \quad (1)$$

여기서 M 은 GMM을 구성하는 가우시안 mixture의 개수를, c_{jk} 는 상태 j 에서 k 번째 가우시안 mixture의 가중치(weight)를, μ_{jk} 와 U_{jk} 는 각각 상태 j 에서 k 번째 가우시안 mixture의 평균벡터(mean vector)와 공분산 행렬(covariance matrix)이고 N 은 전체 상태(state)수 이다.

$b_j(O_t)$ 를 학습 데이터로부터 구하기 위한 Baum-Welch 재추정(re-estimation) 알고리즘은 다음과 같다.

상태수를 N , 심벌의 길이를 T , 전향확률을 $\alpha_t(i)$ ($i=1, 2, \dots, N; t=1, 2, \dots, T$), 후향확률을 $\beta_t(i)$ ($i=1, 2, \dots, N; t=T, T-1, \dots, 0$)라하고, 시간 t 에서 상태가 j 일 확률 $\gamma_t(j, k)$ 를 다음 식 (2)와 같이 정의한다.

$$\gamma_t(j, k) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{j=1}^N \alpha_t(j)\beta_t(j)} \left[\frac{c_{jk} N(O_t, \mu_{jk}, U_{jk})}{\sum_{m=1}^M c_{jm} N(O_t, \mu_{jm}, U_{jm})} \right] \quad (2)$$

위의 세 가지 확률변수(α, β, γ)를 이용하여 모델 파라미터 c_{jk}, μ_{jk}, U_{jk} 를 추정하는 식은 다음과 같다.

$$\bar{c}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k)}{\sum_{t=1}^T \sum_{k=1}^M \gamma_t(j, k)} \quad (3)$$

$$\bar{\mu}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot O_t}{\sum_{t=1}^T \gamma_t(j, k)} \quad (4)$$

$$\bar{U}_{jk} = \frac{\sum_{t=1}^T \gamma_t(j, k) \cdot (O_t - \mu_{jk})(O_t - \mu_{jk})^T}{\sum_{t=1}^T \gamma_t(j, k)} \quad (5)$$

3. 발매기 명령어의 음성인식 구현

3.1 인식대상 어휘의 선정

음성인식 시스템에 적용되는 인식대상 어휘는 그 인식 시스템의 성격을 규정한다. 본 논문은 현재 지하철 구간에 사용중인 발매기 명령어의 음성인식을 구현하고자 하였다.

발매기가 고객들에게 편이를 제공하고 있고 경영자의 입장에서 어느정도 무인화를 가능하게 함으로써 상당 부분의 인력을 재배치 운용할 수 있게 되었다.

하지만 고객에게 보다 편리한 인터페이스 수단을 제공할 수 있도록 발매기 명령어(314개 역명)를 인식대상 어휘로 선정하였다. 발매기 명령어로 사용한 314개의 역명은 현재 서울 및 수도권 구간에서 지하철이 운행중인 광역 지하철역 1호선~8호선의 역명을 대상으로 선정하였다. 본 논문의 구현에 사용된 314개의 인식대상 어휘는 표 1에 정리하였다.

3.2 특징추출

특징벡터 MFCC(Mel Frequency Cepstral Coefficients)의 추출과정을 개략적으로 살펴보면 Fig. 2와 같다. 음성 신호는 anti-aliasing filter를 거친 다음, A/D변환을 거쳐서 디지털 신호로 변환된다. 이 신호는 고대역 통과 특성을 갖는 디지털 프리엠퍼시스 필터를 거친다. 프리엠퍼시스된 신호는 해밍 윈도우를 씌워서 블록 단위의 프레임으로 나누어진다. 이후부터의 처리는 모두 프레임 단위로 이루어진다. 프레임의 크기는 20ms이며 프레임 이동은 10ms를 사용하였다. 한 프레임의 음성신호는 FFT(Fast Fourier Transform)를 이용하여 주파수 영역으로 변환된다. 주파수 대역을 여러개의 필터뱅크로 나누고 각 बैं크에서의 에너지를 구한다. 밴드 에너지에 로그를 취한 후 DCT(Discrete Cosine Transform)를 하면 최종적인 MFCC가 얻어진다. MFCC 계수는 $c_1 \sim c_{12}$ 까지의 12개를 사용하며 이와는 별도로 구한 프레임 로그 에너지가 추가적으로 사용되어 13 MFCC가 된다. 13 MFCC를 갖는 음성신호를 모델링하기 위하여 하나의 프레임 보다는 여러 개의 프레임을 사용하는 것이 더 나은 성능을 얻을 수 있지만 전체 계수의 개수가 매우 많아지게 되므로 여러 개의 프레임을 최소의 파라미터로 표현할 필요가 있어 여기에 델타(미분계수), 델타-델타(가속계수) 까지 확장하여 39 MFCC를 음성신호 입력을 위한 특징벡터로 사용하였다 [4].

Table 1. Station-Names

가락시장	남부터미널	마곡	불광	시청	오금	종합운동장
가리봉	남성	마두	사가정	시흥	오류동	주안
간석	남영	마들	사당	신금호	오리	주엽
강남	남태령	마장	산본	신길	오목교	중계
강남구청	남한산성입구	마천	산성	신길온천	오이도	중곡
강동	내방	마포	삼각지	신답	옥수	중동
강동구청	노량진	마포구청	삼성	신당	온수	중앙
강변	노원	망원	삼송	신대방	올림픽공원	중화
개봉	녹번	망월사	상계	신대방삼거리	왕십리	증산
개화산	녹사평	매봉	상도	신도림	외대앞	지축
거여	녹천	먹골	상록수	신사	용답	창동
진대입구	논현	면목	상봉	신설동	용마산	창신
경마공원	단대오거리	명동	상수	신용산	용산	천왕
경복궁	답십리	명일	상왕십리	신이문	우장산	천호
경원대	당고개	명학	상월곡	신정	원당	철산
고덕	당산	모란	상일동	신정네거리	월계	청구
고려대	대곡	목동	새절	신천	월곡	청담
고속터미널	대공원	몽촌토성	서대문	신촌	을지로3가	청량리
고잔	대림	무악재	서빙고	신풍	을지로4가	초림
공단	대방	문정	서울대입구	신흥	을지로입구	충무로
공덕	대야미	미금	서울역	쌍문	응봉	충정로
과천	대청	미아	서초	아차산	응암	태평
관악	대치	미아삼거리	서현	아현	의정부	평촌
광나루	대화	반월	석계	안국	의정부북부	하계
광명	대흥	반포	석수	안산	이대	학동
광화문	도곡	발산	석촌	안암	이수	학여울
광흥창	도림천	방배	선바위	안양	이촌	한강진
교대	도봉	방이	성균관대	암사	이태원	한남
구로	도봉산	방학	성내	암구정	인덕원	한대앞
구로공단	도원	방화	성북	애오개	인천	한성대입구
구산	도화	백궁	성산	야탑	일원	한양대
구의	독립문	백석	성수	약수	잠실	합정
구일	독바위	백운	성신여대입구	양재	잠원	행당
구파발	독산	버티고개	소사	양천구청	장승배기	혜화
군자	돌곶이	범계	송내	양평	장암	홍대입구
군포	동대문	보라매	송정	여의대공원	장지	홍제
굽은다리	동대문운동장	보문	송파	여의나루	장한평	화곡
금정	동대입구	북정	수락산	여의도	정발산	화랑대
금호	동묘앞	봉천	수색	역곡	정부파견청사	화서
길동	동암	봉화산	수서	역삼	정왕	화정
길음	동인천	부개	수원	역촌	계기동	회기
김포공항	동작	부곡	수유	연신내	제물포	회룡
까치산	둔촌동	부천	수진	영등포	종각	회현
낙성대	독섬	부평	숙대입구	영등포구청	종로3가	효창공원앞
남구로	독섬유원지	북의정부	승실대입구	영등포시장	종로5가	-

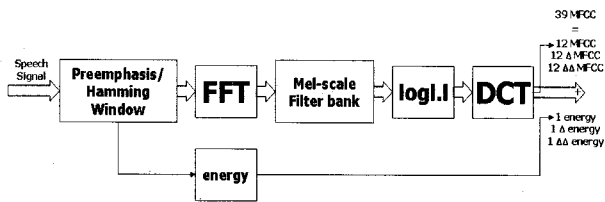


Fig. 2. Schematic diagram of feature extraction from voice

3.3 인식과정

본 논문에서는 HMM 학습을 위한 인식 단위로 음소 모델링(sub-word unit modeling) 중 앞뒤에 오는 음소를 모두 고려한 문맥 종속 트라이폰 모델을 사용하였다[5]. 트라이폰 모델은 음소 모델보다 단어 내의 음운현상을 효과적으로 반영하는 장점이 있지만, 음운조합의 다양한 표현으로 인해 그 수가 과도하게 많아지게 된다. 때문에 신뢰성 있는 모델 파라미터를 추정하기 위해서는 각 트라이폰 모델당 어느 정도의 학습 데이터 확보가 필요하다. 이러한 학습데이터 부족 현상을 해결하기 위하여 문맥이 같은 모든 트라이폰 모델은 전이확률(transition probability)과 상태 파라미터를 공유하게 하는 tied-state 트라이폰을 만들어 준다. 이때 파라미터의 특성을 공유하는 것은 각 모델간의 변별력을 떨어뜨리는 요인이 될 수 있으므로 영향을 미치지 않는 범위 내에서 결정해야 한다[6]. 본 논문에서는 895개의 tied-state 트라이폰을 구성하였다. 음소 기반의 5 state left-to-right model을 정의하였다. 모델의 각 상태에 대해 Gaussian 연속밀도함수에 기반하여 4개의 mixture와 39개의 평균(mean) · 공분산(covariance)을 추출하여 파라미터로 사용하였다. 각 음소 음향모델로부터 고립단어를 인식하는 과정은 다음 그림 3과 같다. 먼저 훈련데이터로부터 Baum-Welch 재추정(re-estimation) 알고리즘을 이용하여 음소단위의 HMM을 학습한다. 인식단계에서는 발음사전을 참조하여 음소모델을 이용하여 인식하고자 하는 단어의 모델을 조립한다. 예를 들어 “서울역”이라는 단어의 HMM은 “s”, “v”, “u”, “l”, “r”, “jv”, “gq” 모델 들을 연결함으로써 얻을 수 있다. 단어모델이 구성되면 입력된 전향 알고리즘을 이용하여 입력된 특징벡터열에 대한 관측확률을 계산하여 가장 확률이 높은 모델을 찾아 인식결과를 출력한다.

3.4 발매기 명령어의 제어과정

본 논문에서 사용한 발매기 명령어를 제어하기 위한 기술로 CHMM을 이용한 음성인식을 사용하였다. 고객들에게 좀더 편리한 서비스 제공 및 인간과 기계사이의 효율적인 인터페이스를 위하여 인간의 가장 기본적인 의

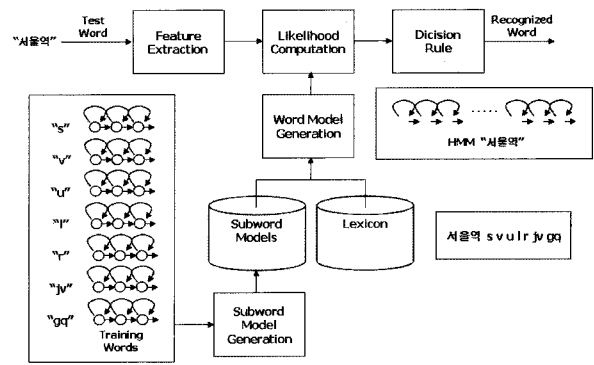


Fig. 3. Speech Recognition of phoneme base

사 소통도구인 음성을 사용함으로써 좀 더 친숙하게 다가갈 수 있도록 하기 위하여 본 기술을 적용하였다. 기존 발매기의 처리과정은 익숙하지 않은 사용자 뿐 만 아니라 익숙한 사용자라 하여도 가고자 하는 역명의 요금을 미리계산 하여야 하는 어려움이 있었다. 하지만 음성을 이용하여 발매기의 명령어를 제어함으로써 이러한 불편함을 해결 할 수 있을 것이다. 아래 Fig. 4와 같이 고객이 가고자하는 역명을 말하면, 음성을 이용한 인터페이스를 통해 원하는 역명과 요금이 자동으로 표시되게 된다. 그 후 고객은 동전을 넣고 출력된 티켓을 받아 지하철을 이용하면 된다.

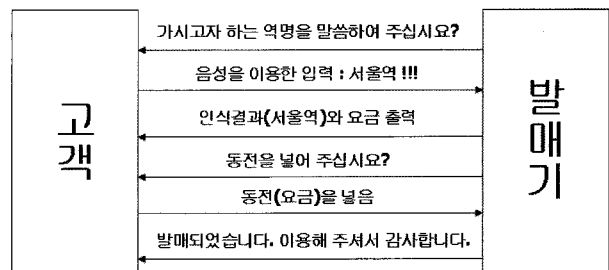


Fig. 4. Control process for commands of Ticketing Machine

발매기 명령어 음성인식 시스템은 Visual C++ 6.0으로 작성되었고 크게 세 부분으로 나누어진다. 첫 번째는 음성의 끝 구간을 검출하고 검출된 데이터를 이용하여 특징을 추출하는 입력부, 두 번째는 입력된 특징벡터를 이용하여 인식을 수행하는 음성 인식부, 세 번째는 인식된 결과를 이용하여 실제 발매기 명령어를 제어하는 제어부이다. Fig. 5는 시스템의 전체적인 개략도와 실제 구현된 화면의 UI를 보여준다.

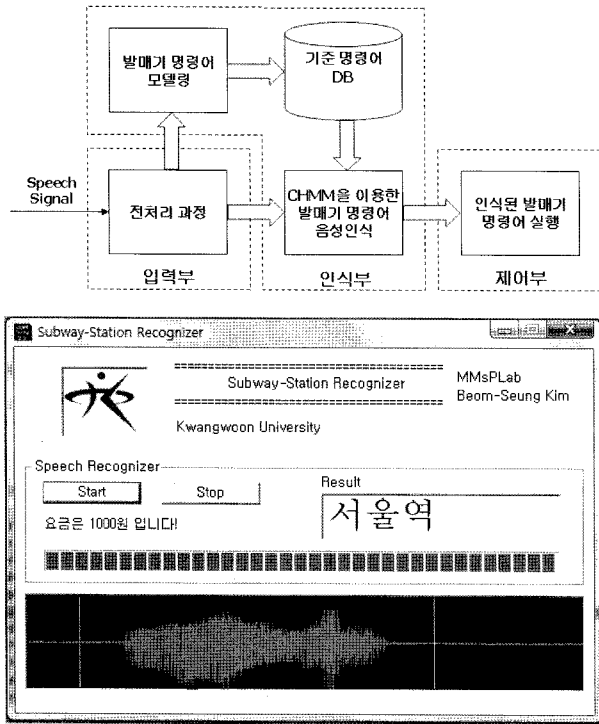


Fig. 5. Schematic diagram and User Interface Screen

4. 실험 및 분석

4.1. 실험 도구

1) 사용 PC: Pentium 200PC & Notebook LG Xnote E200, 2) Microphone: Sony ECM-221 Condenser Microphone & Xnote E200 내장 마이크, 3) Sampling Rate : 8000 Hz & 양자화 레벨 : 16 bit

4.2 실험 방법

구현된 인식 시스템의 성능을 평가하기 위하여 314개의 인식 대상 단어의 목록을 표준말을 사용하는 20대 남성·여성에게 대하여 학습에 참여한 화자 20인(남:10인, 여:10인), 학습에 참여하지 않은 화자 20인(남:10인, 여:10인)에 의해 3회 발성하였고 이를 이용하여 다중 화자 종속/독립 실험을 수행하였다. 화자는 종속 화자 그리고 독립 화자로 구별한다.

마지막으로 노이즈가 있는 환경에서 학습에 참여하지 않은 20인(남:10인, 여:10인)의 독립 화자의 인식 실험을 수행하였다.

4.3. 화자 종속/독립에 대한 인식률 비교

본 논문에서 구현된 시스템은 화자 독립 인식 시스템을 목표로 연구 개발되어졌기 때문에 본 절에서는 화자 독립

성능을 평가하기 위해서 화자 종속 및 독립 실험을 수행하였다.

본 실험에서는 제안된 모델을 사용하였고 39 MFCC 특징벡터를 사용하였다. 실험 결과는 표 2와 같이 인식률은 화자 종속의 경우 99.24%를 보였고, 화자 독립의 경우 98.02%로 나타났다.

Table 2. Recognized Result of Speaker dependent/independent

구분 \ 방법	화자종속	화자독립	Average
남성	99.54	98.12	98.83
여성	98.94	97.92	98.43
Average	99.24	98.02	98.63

4.4. 노이즈 환경에서의 인식률 비교

본 논문에서 구현된 시스템은 대중들이 많이 이용하는 역 대합실이나 지하철 역내 등 공공장소와 같은 여러 가지 노이즈가 발생할 수 있는 환경이다.

본 실험은 이러한 환경에서 시스템의 성능을 평가하기 위해서 휴대폰 벨소리, 열차에 의한 진동, 여객들의 대화소리 등 소음이 발생할 수 있는 특수한 환경을 가정하고 수행하였다. 화자 독립 인식 실험 중 전화벨 소리가 여러 번 울렸고, 주변의 사람들의 웅성대는 소리, 계단 내려오는 발소리 등 소음이 있는 환경에서 실험하였다. 아래 표 3은 노이즈 환경에서 화자 독립 실험의 인식률을 나타낸다.

Table 3. Recognized Result of noise environment

구분 \ SNR	15dB	10dB	5dB	Average
남성	97.39	94.47	90.57	94.14
여성	96.69	93.98	90.36	93.68
Average	97.04	94.22	90.47	93.91

실험 결과 평균 93.91%의 인식률을 나타내었다. 이는 노이즈가 없는 환경에서의 인식률보다 4.11% 낮은 값이다. 일반적으로 SNR(Signal to Noise Ratio)이 충분히 큰 경우(30dB 이상)는 작은 에너지를 갖는 음성신호라 할지라도 주변잡음보다는 큰 에너지 값을 가지므로 에너지 함수만으로 음성구간을 검출할 수 있지만 그렇지 않은 경우는 인식률에 큰 영향을 미친다[7,8]. 따라서 노이즈를 효율적으로 처리하는 것이 인식률을 높일 수 있는 관건이다. 또한 오인식 단어들을 보면 삼성·삼송, 신천·신촌 등 발음이 비슷한 경우로 시작되는 단어의 자음이 같거나 끝부분의 자음이 같은 경우 나타났고 처음 자음으로 시작하는 단어의 경우 음성 끝점 검출시 노이즈와 구별하지 못하여 예러가 발

생하였다.

위의 결과로 본 시스템의 경우 발매기 명령어(314개역명)의 유사단어와 주변 노이즈들이 음성인식에 영향을 미치는 것을 알 수 있었다.

5. 결론 및 향후 과제

본 논문에서는 CHMM을 이용하여 발매기 명령어 314개의 역명을 실시간으로 인식하는 화자 독립 시스템을 구현하였다. 시스템 구현을 위해 음성의 구간을 자동으로 검출하는 음성 검출 알고리즘을 구현하였으며, 특징벡터로 39 MFCC를 사용하였고 인식률 향상을 위하여 895개의 tied-state 트라이폰 음소 모델을 구성하였다. 또한 사용자에게 음성을 이용한 인터페이스를 제공함으로써 보다 쉽고 편리하게 발매기를 이용할 수 있도록 하였다.

본 논문에서 구현된 시스템은 표준말을 사용하지 않는 특정지역 사람들의 사투리 억양을 고려하지 않았다. 따라서 발음이 적절하지 않은 화자가 사용할 경우 인식률이 떨어지는 경우가 있다. 때문에 본 시스템의 성능 향상을 위해서 이러한 지역 사람들의 특징도 고려해야 할 것이다. 또한 정상인이나 장애우 모두에게 사용하기 쉽고 편리한 인터페이스를 구현하기 위해서는 더 많은 실험과 연령대 별로 사용자를 다양화해서 연구해야할 필요성이 있으며, 실제 발매기가 설치된 위치가 소음에 노출된 역 구내라는 특수한 환경을 고려하여 좀 더 잡음에 강인한 인식 시스템의 구현이 필요하다.

특히, 본 시스템을 이용하여 장애우들을 위한 시스템 구

현을 할 경우 음성을 더욱 효율적인 인터페이스로 수단으로 사용할 수 있을 것이다.

참고 문헌

1. 강점자, 강병욱, 정호영, 정훈, 이윤근, “신성장동력산업용 대어휘 음성인식 기술 동향 및 응용”, 전자통신동향분석, 제23권 제1호, pp. 65-76, 2008. 2.
2. L. R. Rabiner, “A tutorial on hidden Markov models and selected applications in speech recognition”, Proc IEEE, vol. 77, no. 2, pp. 257-286, Feb. 1989.
3. D. A. Reynolds and R. C. Rose, “Robust text-independent speaker identification using gaussian mixture speaker models”, IEEE Transactions on Speech and Audio Processing, vol. 3, no. 1, pp. 72-83, Jan. 1995.
4. D. Jurafsky and J. H. Martin, “Speech and Language Processing”, Prentice Hall(2nd), 2008.
5. L. R. Rabiner and B. H. Juang, “Fundamentals of Speech Recognition”, Prentice Hall, 1993.
6. L. Gu and K. Rose, “Sub-state tying in tied mixture hidden Markov models”, Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, pp. 1062-1065, 2000.
7. Y. Ephraim, D. Malah, B. H. Juang, “On the Application of Hidden Markov Models for Enhancing Noisy Speech”, IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 37, no. 12, pp. 1846-1856, Dec. 1989.
8. P. J. Moreno, “Speech Recognition in Noisy Environments”, Ph d. dissertation”, Carnegie Mellon Univ. Pittsburgh, Pennsylvania, 1996.

접수일(2008년 11월 8일), 수정일(2009년 1월 15일),
게재확정일(2009년 4월 13일)