무선 네트워크 환경 하에서의 음성인식에 관한 고찰

임 수 호, 신 광호, 함 성준, 김 주 곤, 정 호 열, 정 현 열
영남대학교 전자정보공학부

A Speech Recognition in a Wireless Network Environment

Soo-Ho Lim, Seong-Jun Hahm, Guang-Hu Shen, Joo-Gon Kim,
Ho-Youl Jung, Hyun-Yeol Chung
Dept. of Information and Communication Eng., Yeungnam University
E-mail: shlim@yumail.ac.kr

요 약

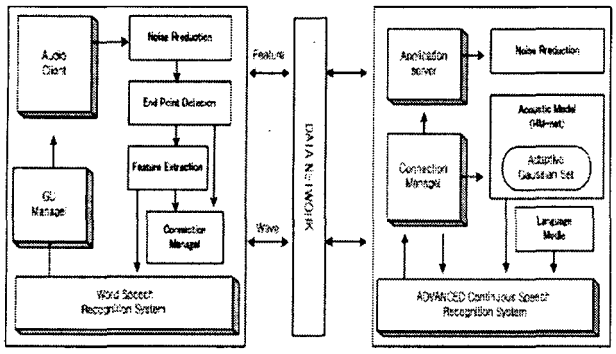
최근 PDA(Personal Digital Assistants)와 같은 휴대형 단말기들은 다양한 멀티미디어 기술과 무선 인터넷 기술의 영향으로 정보단말기로서 각광을 받고 있다. 그러나 현재의 단말기는 프로세서와 메모리의 한계로 인하여 원활한 음성인식 시스템을 구축하기에는 한계가 있다. 이를 보완하는 방법으로 본 논문에서는 Client/Server로 분리된 음성 인식 시스템을 구축하였다. 구축한 시스템은 무선 네트워크 환경을 이용하여 PDA(Personal Digital Assistants)에서 음성 파일 또는 특징 파라미터를 Serve 측으로 전송하여 Server측에서 음성 인식을 수행한 후 그 결과를 모바일 단말기로 되돌려 주는 시스템이다. 구성된 시스템을 평가하기 위해서는 국어공학센터의 음성 DB(KLE 452DB)를 이용하여 음향 모델을 생성한 후 다양한 환경(연구실, 복도, 주차장, 도서관 로비)에서 발성한 후 이를 교내 무선 인터넷망(Nespot)을 통하여 송신하여 실시간 인식하였다. 실험 결과, 각각 84.04% 72.28% 69.47% 67.61%의 평균 인식률을 얻을 수 있었다.

1. 서 론

오늘날 이동형 소형 단말기들은 다양한 멀티미디어 기술과 무선 인터넷의 영향으로 포스트 PC의 대표적인 정보통신 수단으로 각광을 받고 있다. 그러나 현재의 단말기의 사용상 편이를 위한 man-machine 인터페이스로는 사용 프로세서와 메모리 사용량의 한계로 인해 대부분 온라인 제어형 명령어를 위한 고립단어 음성인식 등에 주로 이용되고 있으나 휴대형 단말기에서의 사용자 편이성을 증대시키기 위해 사용가능한 입력 장치를 확대하기 위한 연구도 진행되고 있다. 그 예로서 무선랜 환경에서 클라이언트 서버 음성 인식 시스템을 구현한 MIPAD 시스템, Voice XML을 이용한 시스템 등이 있다. 휴대용 모빌 환경에서 사용자의 편이성 제고를 위한 인터페이스를 확대하기 위한 입력 인식하는 시스템 설계가 필요하다. 현재 화자 독립 고립단어의 인식을 수행하는 상용단말기들이 출현하고 있으나, 소형 단말기의 한정된 처리장치의 속도, 메모리 용량으로 인해 인식성능과 입력 가능한 어휘 수에 제한을 받고 있다. 또한 화자 독립 연속 음성인식 시스템을 클라이언트 환경에서 구현하는 데에는 아직까지 단말기 시스템의 성능에 한계를 나타내고 있다. 따라서 본고에서는 무선망환경하에서의 연속음성인식 시스템구현을 위한 기초연구로서 무선망을 이용하여 음성파일을 송신한 후 서버에서 인식하여 되돌려 받는 방식의 시스템 구현하고 그 결과를 보고한다[1][2].

2. 무선망을 이용한 인식 시스템

이동형 모바일 단말기에서 고립단어 음성인식과 연속 음성 인식을 이용한 멀티모달 시스템 구성도를 그림.1에 나타내었다.



CLIENT(PDA& nart-phone) SERVER(PC)

그림 1. 시스템 구성도

클라이언트/서버 환경으로 분리된 시스템은 무선망을 이용하여 정보를 전달하게 된다. 클라이언트 환경에서는 작은 고립단어 음성인식을 수행하는 시스템을 사용한다. 클라이언트의 인식속도 및 메모리 사용량 최소화를 위해 상태 및 혼합 수 분할알고리즘을 이용한 축약형 모델을 사용한다.

클라이언트가 무선망을 이용할 수 있는 경우 연속 음성을 그대로 서버로 전송하거나, 연속음성에 대한 파라미터를 추출한 후 이를 서버로 전송하여 서버에 탑재된 인식 엔진으로 연속음성 인식을 수행하게 된다[1][2][3].

이때, 클라이언트 단말기에서는 잡음환경에 강인한 음성인식을 위해 스펙트럼 평균 차감법과 켑스트럼 평균 차감법을 이용한다. 또, 자연스러운 발성을 통한 음성입력을 위해 실시간 무제약 음성구간 검출 알고리즘을 이용한다. 무선망 환경의 음성 인식시스템으로 클라이언트는 Pocket PC 2002 환경의 컴팩 iPaq에서 구현한다. 그림.2에서 시스템 구성을 나타낸 것이다.

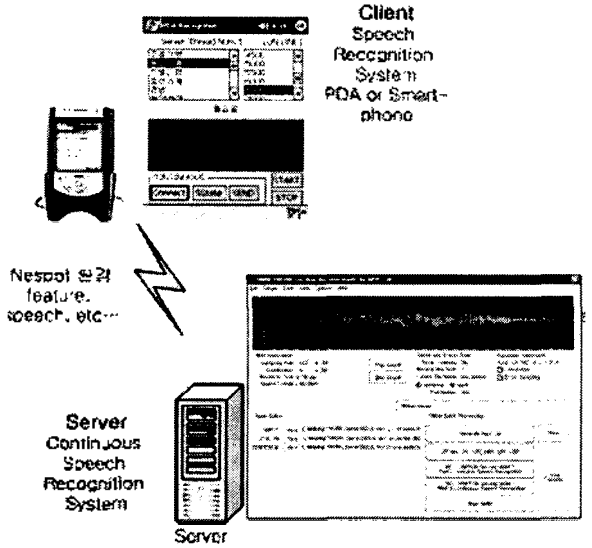


그림 2. 무선망 환경의 음성 인식 시스템

이때 사용된 무선망은 공중망 무선랜의 표준인 IEEE 802.1X에 따른 망과, 향후 IMT-2000 등의 차세대 이동전화 표준에 의한 망으로 하고, 무선망이 제공해야할 전송속도는 실시간으로 파라미터 혹은 음성을 전송하기 위한 2.5~8kbps 이상이다.

서버측 연속 음성 엔진은 음소결정트리를 이용하여 생성된 대어휘 상태 공유 Tri-phone 은닉마르코프 네트워크(HM-Net : Hidden Markov Network) 모델을 이용하고, bi-gram 언어모델을 이용한다. 또, 서버는 여러 단말기를 통하여 동시에 전송된 요구를 동시에 처리하기 위해 쓰레드를 이용한 다중 소켓 처리 방법을 이용하여 인식을 수행 할 수 있도록 하였다[5].

표.1 과 표.2에 본 논문에서 제안된 인식 시스템의 데이터분석 조건을 나타내었다.

표 1. Client System

	Word Speech Recognition
전처리	8kHz Sampling,16bits 16ms Hamming Window 5ms shift
특징파라미터	MFCC 39차
학습 DB	주식 상장사명 100단어 20명
Model	Variabel Parameter CHMM, 48 monophone

표 2. Server System

	Word Speech Recognition	Continuous Speech Recognition
전처리	8kHz Sampling,16bits 25ms Hamming Window 10ms shift	Use Wore
특징 파라미터	MFCC 39차	Use Wore
학습 DB	KLE 452 단어 35명	Trade data 문장 90명
Model	DT-HMnet, Tyied state Triphone (2000 Diagonal Gaussian)	DT-HMnet Tyied state Triphone (3000 Diagonal Gaussin)

3. 실시간 성능 평가

본 논문에서 구현된 시스템의 성능 평가하기 위해서 국어공학센터에서 제공하는 452단어(KLE452)를 이용하였다. 이 중에서 남성 화자 35명이 발성한 단어로부터 추출한 음소를 모델을 만든 후 KLE452 단어를 4set으로 나누어 인식 실험을 수행하였고, 이를 교내 무선 인터넷망(Nespot)을 이용하여 기준 환경인 연구실 환경에서 교내 남자학생 5명을 대상으로 인식 실험을 수행하였다. 또한 다양한 잡음 환경을 모사하기 위하여 복도, 주차장, 도서관에서도 실시간 인식 실험을 수행하였다.

3.1 기준 환경에서의 실시간 음성인식 실험

객관적인 평가를 위해 연구실환경을 기준 환경으로 설정하고 본 논문에서 제안한 시스템을 무선 네트워크 환경(Nespot)을 이용하여 실시간 인식실험을 실시하였다.

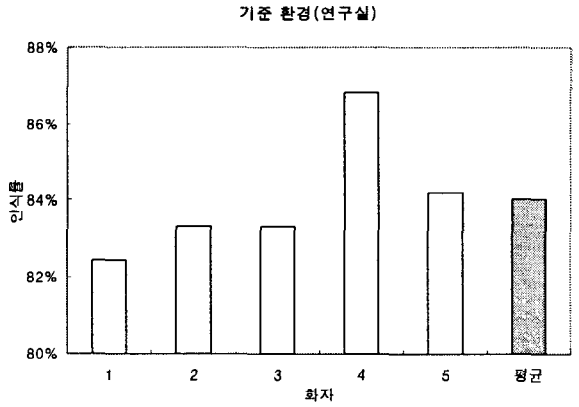


그림 3. 기준 환경의 인식률

그림.3에 인식결과를 보인다. 화자 5명의 평균인식률은 84.04%이었다.

3.2 잡음환경에서의 실시간 음성인식 실험

연구실 환경을 기준으로 다양한 잡음 환경에서의 인식 성능 평가를 위해 교내의 복도, 주차장, 도서관에서 실시간으로 음성 인식 실험을 수행하였다.

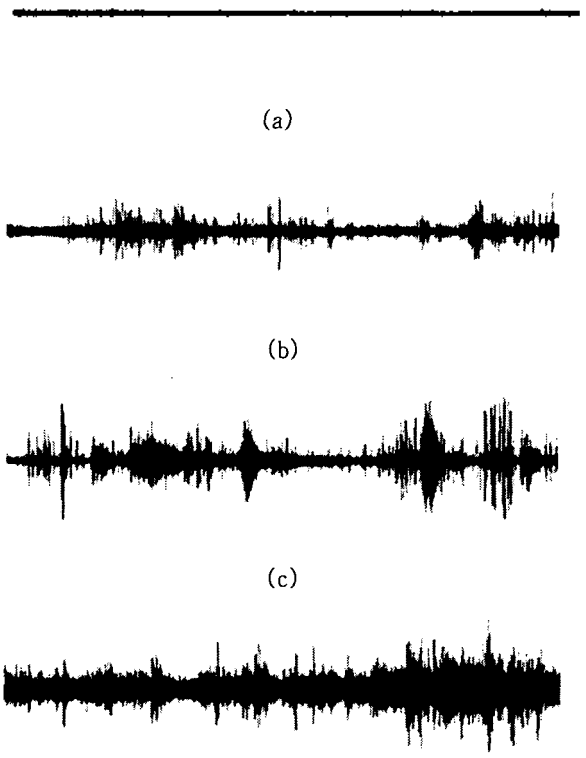
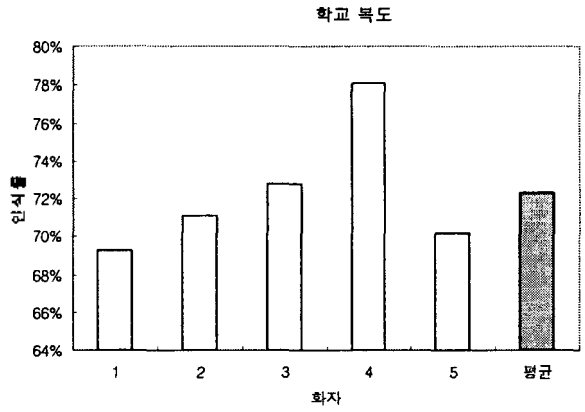
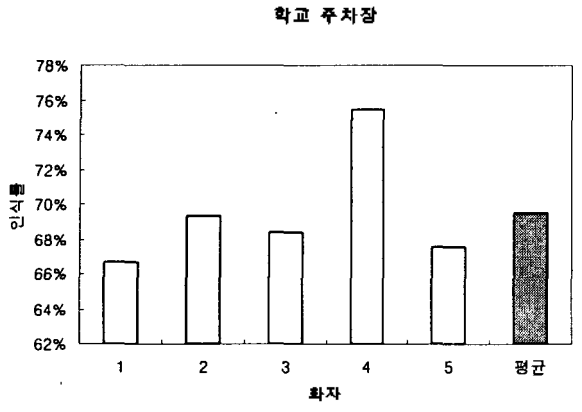


그림 4. (a)연구실, (b)교내복도 (c)주차장 (d)도서관 로비 잡음비교의 예

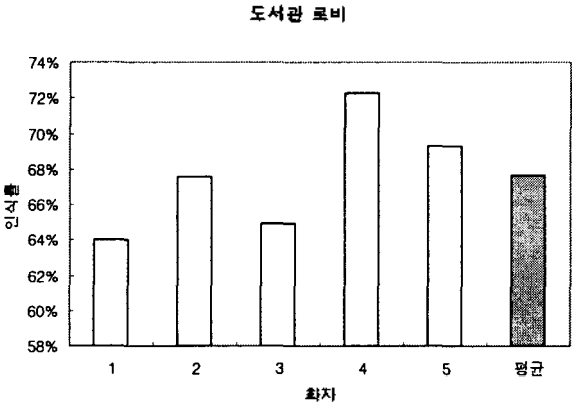
그림.4에 교내 복도, 주차장, 도서관 로비의 환경의 잡음파형을 나타내었다. 연구실 환경을 기준으로 할 때 잡음 레벨은 각각 9dB, 12dB, 15dB이다.



(a)



(b)



(c)

그림 5. 잡음 환경에서 실시간 인식률
((a)교내복도 (b)주차장, (c)도서관 로비)

그림.5 에 본 논문에서 제안한 시스템을 각기 다른 잡음(복도, 주차장, 도서관 로비)환경에서 실시간으로 인식을 수행 한 결과를 나타내었다. 평균인식률이 각각 72.28% 69.47% 67.61% 로 나타남을 알 수 있다.

4. 결론

본 논문에서는 무선 네트워크 환경을 이용하여 PDA(Personal Digital Assistants)와 같은 이동형 모바일 단말기에서 음성 파일 또는 특징 파라미터를 Server 측으로 전송하여 Server측에서 음성 인식을 수행한 후 그 결과를 모바일 단말기로 알려주는 시스템을 구현하고, 그 성능을 평가하였다.

제안된 시스템을 평가하기 위해서 국어 공학센터의 음성 DB(KLE 452DB)를 이용하여 음향 모델 만든 후 이를 교내 무선 인터넷망(Nespot)을 이용하여 다양한 환경(연구실, 복도, 주차장, 도서관 로비)에서 실시간 인식결과를 도출하였다. 실험 결과, 평균 인식률이 각각 84.04% 72.28% 69.47% 67.61%을 얻을 수 있음을 확인할 수 있었다.

향후 실시간 환경에서의 강건한 시스템을 구축하기 위해 잡음 환경에서의 끝점 검출 및 잡음 처리를 함으로서 시스템의 성능 향상을 위한 연구를 계속 하고자 한다.

참 고 문 헌

[1] 석수영, 김춘영, 조재원, 정호열, 정현열, "무선 네트워크 환경을 위한 PDA용 음성, 문자 공용 인식 시스템," 한국음향학회 추계학술대회 논문집 제22권 제2(s)호, pp. 19-20, 2003.

[2] Li Deng,. "Distributed speech Processing in MiPad's Multimodal User Interface," IEEE Trans. Speech and Audio., Vol 10, No 8, PP 605-619, 2002.

[3] Suk, S.Y., Jung, H.Y.. and Chung, H.Y. "Automatic Generation of Context-Independent Variabel Parameter Models using Successive State and Mixture Spitting," EuroSpeech Proc., 2003.

[4] WeiQi, Zhang., Liang, He., Yen-Lu, Chow., RongZhen, Yang., YePing, Su., "The Study on Distributed Speech Recognition System." Acoustics, Speech, and Signal Processing, 2000.ICASSP'00. proceedings.2000 IEEE International Conference on, Vol 3, pp 1431-1434, 5-9 June 2000

[5] Takaki, H., "A Study on HM-Nets using Decision Tree-based Successive Splitting," ICSP Proc., 383-387, 1997