

음성 인식을 이용한 자동 교환 시스템 구현

함정표, 김현아, 박익현

LG 정보통신 정보시스템 연구소, 미디어 기기실

An Implementation of the Automatic Switching System using Speech Recognition

Jeong-Pyo Ham, Hyun-A Kim, Ik-Hyun Park

Media Communication Hardware Dept.

Digital Network System Lab. Digital Network Company

LG Electronics Inc.

Email : { jpham, hakim, v108369 }@lgic.co.kr

요약

본 논문에서는 음성 인식을 이용하여 전화를 교환해주는 자동 교환 시스템을 구현하고, 성능을 평가하였다. 구현된 시스템에는 필수적인 음성인식 이외에도 DSP 진단 기능, 인식 대상 어휘의 추가 및 변경기능, 음성 수집 기능 등이 구현 되었다. SCHMM (Semi-Continuous Hidden Markov Model)을 이용한 전화망에서의 화자 독립 고립 단어 가변 어휘 인식을 대상으로 하였으며, 실시간 구현을 위하여 Texas Instrument사의 TMS320C32를 사용하였다[6]. 인식 어휘는 부서명 및 인명이고 1300 여 단어일 때, 인식 성능은 91.5%이다.

1. 서론

음성 인식 기술이 점차 발달하고 전화를 이용한 다양한 서비스가 개발되면서 전화망을 통하여 음성 인식 서비스를 제공하고자 하는 요구가 증대하고 있는 실정이다. 더욱이 휴대폰의 보급으로 전화 사용이 증가하여 중권 정보 서비스, 자동 교환 등의 부가 서비스를 보다 손쉬운 사용자 인터페이스를 통하여 제공하는 것이 절실하다.

본 논문에서는 현재 교환원이 교환해 주거나 내선 번호를 통하여 교환되던 기능을 음성 인식을 통하여 자동으로 교환되는 자동 교환 시스템을 구현하였다.

교환 시스템은 여러 사용자에게 서비스하기 위하여 다수의 독립적인 음성 인식 모듈을 실장하여 구성된다. 따라서 별도의 제어용 모듈에서 보내는 메시지를 통하

여 음성 인식 모듈들의 동작을 제어한다. 또한, 인식 모듈은 메시지를 통하여 끝점 및 인식 결과 보고 등을 제어 모듈과 통신하게 된다. 메시지와 별도로 인식 모듈의 프로세서는 디지털 트렁크를 통하여 전송되고 있는 음성을 직렬 포트를 통하여 받는다.

인식 모듈은 다수의 사용자를 대상으로 동작하므로 화자 독립 인식 시스템으로 구현되었으며, 가변적인 인식 어휘를 가지므로 음소 단위로 인식한다. 각각의 음소는 SCHMM으로 구현되었다[5]. 인식 어휘는 음소들을 이용하여 사전으로 구성되며, 실시간 인식을 위하여 목구조형으로 연결된다. 음성 인식을 실시간 구현하기 위해서는 상태 병합 및 빔 탐색 알고리즘을 적용하여 인식 시간을 줄였다.

인식 모듈의 DSP는 TI사의 TMS320C32를 사용하였으며 각 포트당 4.5 Mbyte의 메모리를 사용하였다.

서론에 이어 2장에서는 음성 인식 자동 교환 시스템의 구성 및 기능을 소개하고, 3장에서는 음성 인식 모듈의 구성 및 기능에 대한 내용을 서술한다. 4장에서는 구현된 음성 인식 알고리즘과 실험 결과에 대하여 적고, 마지막으로 5장에서는 결론을 맺는다.

2. 음성 인식 자동 교환 시스템

2.1 음성 인식 자동 교환 시스템의 구성

전화망 부가 서비스 시스템인 VIPS-HM(Value Added Information Processing System -Hybrid Medium)은 공중 전화망(PSTN), 이동 전화망(AMPS/DCN/PCS) 또는 사설

교환(PBX)망과 연결되어 가입자에게 실시간으로 원하는 음성이나 FAX 정보를 제공한다[4]. 운용 단말은 (CMS:Central Management System) VIPS-HM의 동작 상태를 확인하고 제어할 뿐만 아니라 각종 통계 정보 및 장애를 관리한다. 또한 음성 인식 자동 교환 서비스와 관련하여 인식 대상 어휘의 리스트, 전화번호 및 기타 데이터베이스를 관리하고 추가, 변경, 삭제에 따른 데이터베이스의 변경 시에 이에 관련된 데이터 파일을 생성하여 VIPS-HM으로 전송하여 주는 역할을 담당한다.

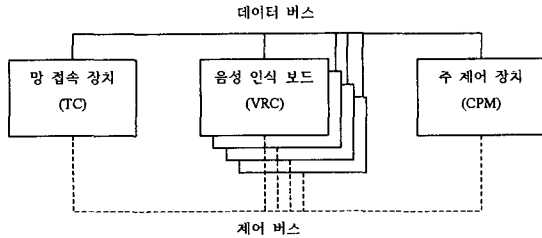


그림 1. 시스템 구성도

주 제어 모듈(CPM : Centralized Processing Module)은 VIPS-HM 시스템 전체를 제어하는 장치로서 망 접속 장치(TC : Trunk Circuit)를 통하여 접속된 통화로의 연결, 유지, 복귀 등 호 처리 기능을 제공하며 접속된 호에 대한 통계를 수집한다. 또한 시스템 전체의 동작 상태를 감시하고, 장애 발생시 운용 단말 및 원격 정보 장치를 통하여 가시 및 가청 정보를 알려주며 서비스 시나리오를 수행하는 모듈이다.

망 접속 모듈은 공중 전화 망 및 사설 교환기와의 Digital/Analog Trunk 접속 및 국간 신호처리 기능을 수행하는 모듈이다.

음성 인식 보드(VRC : Voice Recognition Circuit)는 음성 인식 서비스를 담당하는 가입자 보드로서 보드 당 4 채널 또는 8 채널의 음성인식, 음성 Play, DTMF/Tone 검출, 디지털 음량조절, ADPCM Encoder/Decoder 기능을 제공하며 VIPS-HM에는 4 모듈까지 장착된다. 음성 인식 프로그램은 음성 인식 보드의 DSP에 탑재되며 주 제어 모듈(CPM)로부터 음성 인식 시작 명령을 받아 음성 인식 결과를 보고하는 기능을 수행한다

망 접속 장치와 음성 인식 보드, 그리고 주 제어 장치는 음성 및 데이터의 경로인 데이터 버스와 상호 제어 및 상태 관련 경로인 제어 버스와 연결되어 있다.

2.2 제어 버스와 데이터 버스

VRC와 CPM 그리고 TC와의 메시지는 제어 버스를 통하여 구현된다. TC는 호 관련 메시지를 CPM과 교환하며, VRC는 CPM과 음성 인식 관련 메시지를 교환한다. 음성 인식 모듈에 있어서, CPM으로부터 VRC로 보

내는 메시지는 인식 시작, 인식 멈춤, 진단 명령 등이 있으며 VRC가 CPM으로 보내는 메시지는 시작점, 끝점, 인식 결과, 진단 결과 보고 등이 있다. 정의된 메시지는 하드웨어 타이머 인터럽트를 이용한 메시지 폴링 방식을 사용하여 교환한다.

음성 인식 자동 교환 서비스 진행을 위한 안내 음성을 CPM과 VRC 사이에서 주고 받거나, VRC의 프로그램과 기타 인식 관련 데이터를 CPM이 전송하는 기능은 데이터 버스를 통하여 구현된다.

3. 음성 인식 모듈

3.1 음성 인식 모듈의 구성

음성 인식 모듈은 크게 전처리부, 특징 벡터 추출부, 인식 핵심부, 및 메시지 처리부로 나눌 수 있다.

Highway를 통하여 들어온 원천 음성 데이터는 음성 인식 모듈 프로세서의 직렬 포트를 통하여 들어온다. 직렬 포트 인터럽트 루틴은 A/u-law 신호를 linear PCM 신호로 변환하며, 전처리 고역 통과 필터 $1-z^{-1}$ 을 통과하게 된다. 이러한 기능을 담당하는 전처리부는 인식 모듈이 동작하는 동안 배경 작업으로 수행된다.

특징 벡터 추출부는 CPM에서 전송된 메시지에 따라 전처리를 통과하여 저장된 신호를 normalized LMS 알고리즘으로 반향 제거한다. 사용자에게 전송되는 안내 멘트 등이 섞인 원천 신호에서 사용자의 음성 만을 남기고 제거하게 된다. 반향 제거된 음성은 끝점 검출 루틴을 통하여 시작점과 끝점 메시지를 발생시킨다. 끝점 검출된 부분의 음성은 특징 벡터가 추출되어 메모리에 저장된다.

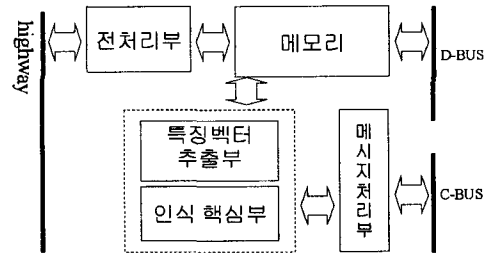


그림 2. 인식 모듈 구성도

인식 핵심부는 특징 벡터의 관찰 확률 계산 및 Viterbi 빔 탐색 과정을 통하여 HMM의 발생 확률을 구한다. 이들을 비교하여 최고의 발생 확률을 가지는 단어를 인식된 단어로 선택한다.

메시지 처리부는 특징 벡터부와 인식 핵심부의 메시지를 CPM에 전달하고 받는다. 메시지는 DPRAM(Dual-Port RAM)을 통하여 주고 받으며 타이머 인터럽트 루틴에서 메시지 큐를 검색하고 적합한 루틴을 호출한다.

3.2 인식 모듈 기능

인식 모듈은 인식하는 기능외에도 상용화되어 유지, 운용되기 위하여 여러 가지 기타 기능들을 필요로 한다.

3.2.1 자체 진단 기능

음성 인식 자동 교환 시스템 구현에서 시스템의 신뢰성은 매우 중요하다. 시스템의 신뢰성 향상을 위한 조치로 VIPS-HM의 주 제어 장치는 Active/Stand-By 방식을 이용한 이중화로 구성되며, 그 외 각 부분들은 주기적인 진단을 통해 장애 발생시 장애를 복구하도록 되어 있다.

음성 인식 모듈의 경우 주기적으로 운용 단말로부터 내려오는 진단 명령을 받아 자체적인 진단을 수행하여 오류가 있거나 진단 결과를 보고할 수 없을 정도의 장애가 있을 시 장애의 복구를 수행한다. 음성 인식 모듈의 자체 진단 방법으로는 DSP 프로그램과 파라미터인 코드북, 사전 및 모델 등 인식 데이터 메모리 영역과 제어 버스의 메시지 전달 기능에 대한 오류 검사를 수행한다. 오류 검사를 위해서 인식 모듈의 메모리 영역은 에러 체크 값(checksum value)을 가지고 있다.

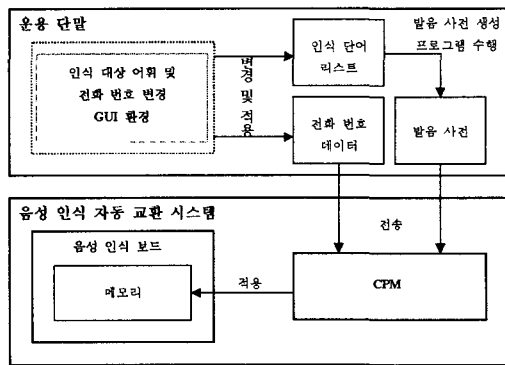


그림 3. 음성 인식 대상 어휘 변경 작업 과정

3.2.2 인식 대상 어휘 변경 기능

음성 인식 구내 자동 교환 시스템은 특정 부서나 사람의 전화 번호를 외우지 않고서도 부서명이나 인명을 발성함으로써 해당 전화 번호로 연결하므로 사용자는 편리하지만 부서나 인원의 추가, 삭제, 변경에 따른 인식 대상 어휘의 변경 발생시 인식 대상 어휘의 변경이 뒤따라야 한다는 측면에서 운용자는 인식 모듈 자체를 교체하거나 인식 모델을 재생성하는 어려움이 있다. 그러나, 이러한 작업을 운용 단말 GUI(graphic User Interface)를 통해 이루어지게 함으로써 부서나 인원의 추가, 삭제, 변경 사항만 입력하면 자동으로 인식 모델을 생성할 뿐만 아니라 운용자의 실수로 인하여 생길 수 있는 오류를 최소화 하도록 하였다. 즉, VIPS-HM의 운용단

말에서 부서나 인원의 리스트 및 전화번호 등 필요한 정보를 데이터 베이스로 구축하여 관리하고 변동 사항이 생길 때 운용자는 GUI를 통해 편집하고 적용 버튼을 누르면 자동으로 새로운 인식 모델이 생성된 후, 시스템으로 전송되어 각 인식 모듈의 해당 메모리의 위치로 전송되어 인식 대상 어휘의 변경이 실시간 적용된다. 그림 3은 인식 대상 어휘가 변경되어 전송되는 과정을 나타낸다

3.2.3 음성 수집 기능

음성 인식기 구현에서 시스템에 입력된 신호를 분석하거나 학습 데이터 베이스로 사용하기 위해서 음성 데이터의 지속적인 수집은 필수적이다. 시스템의 동작 모드에서 음성 수집 모드를 설정하면 끝점 보고된 음성 혹은 호 접속이 일어난 시점부터의 음성이 수집된다. 음성 수집은 데이터 버스를 통하여 시스템의 하드 디스크로 수집 되거나, LAN을 통하여 특정 컴퓨터에 저장된다.

4. 인식 알고리즘 및 실험 결과

4.1 인식 알고리즘

음성 인식은 음성을 확률 모델로 가정하고 입력된 음성을 학습된 모델 중에서 발생할 확률이 가장 큰 모델을 선택하는 과정으로 정의할 수 있다. 현재 사용되고 있는 음성 모델링 방법으로는 HMM(Hidden Markov Model)이 가장 효과적으로 알려져 있다. 사용된 HMM은 관찰 확률을 정규 분포의 합으로 계산하는 5 mixture Semi-Continuous HMM을 사용하였으며, 128 크기의 코드북을 사용하였다.

인식 단위는 음소(phoneme)이며, 인식 성능을 높이기 위하여 문맥에 따른 모델을 사용하는 방법 중에서 전후 모델에 따라 다른 HMM을 사용하는 tri-phone 모델링을 하였다[1]. tri-phone을 사용하면서 증가하는 모델 개수의 증가에 따른 영향을 줄이기 위하여 확률적으로 근사한 상태(state)를 병합시켰다. 정의된 mono-phone은 모두 51개이며, 각 음소를 3 state SCHMM으로 모델링하였다. 병합되어 모두 상태의 숫자를 1000개 내외로 하였다[2][3].

독립 단어 인식이라는 점을 고려하여 인식 성능은 저하되지 않으며 속도를 높이기 위하여 목구조 탐색을 구현하였다. 목구조 탐색(tree search)은 사전을 정렬한 후 공통되는 부분의 상태의 확률값을 공유하여 탐색 범위를 줄이는 방법이다. 또한 인식 지연 시간을 줄이기 위하여 전방향 탐색 과정중에 확률이 문턱치보다 작은 경로에 대하여는 탐색을 생략하는 빔탐색(beam search)를 사용하였다.

음성은 8 kHz 표본화 된 신호를 고역 통과 필터를 통

과한 전처리된 신호를 사용하였으며, 특징 벡터는 12 차 LPC 캡스트럼을 사용하였다. 동적 정보를 위해서 전후 방 각 10ms 와 20ms 의 차분 캡스트럼과 이차 차분 캡스트럼을 사용하였으며, 에너지와 차분 에너지 그리고 이차 차분 에너지를 사용하여 모두 51 차 특징 벡터를 사용하였다. 이를 이용하여 2 가지 코드북이 각각 128 차의 벡터를 가지게 된다. 또한 전화망에서 채널 잡음을 제거하기 위해서 CMS(cepstrum mean subtraction)을 사용하였다.

4.2 실험 및 결과

학습 데이터 베이스는 사용될 전화 환경에서 직접 시스템을 통하여 수집되어 최대한 인식 환경과 시험 환경이 동일하도록 하였다. 전화망을 통하여 남녀 각각 250 명씩 500 명의 수집된 데이터 베이스를 통하여 학습되었으며, 시험을 위하여 남녀 각 8 명의 사용자가 실제 사용과 같은 상황에서 전화를 통하여 시험하였다.

자동 교환 시스템 용 인식 대상 어휘는 부서명과 인명 및 직책명으로 이루어져 있으며, 부서명은 다시 사용자의 편의성을 고려하여 정식 명칭 외에도 통상적으로 사용되는 약식 명칭과 기능에 따른 민원 명칭으로 구성된다. 표 1 은 인식 대상 어휘들의 분류표이다.

표 1. 인식 대상 어휘들의 분류.

구분	예
인명	손영재, 조세환, 금광현, ...
직책명	시장, 부시장, 세무 과장, ...
부서명(정식명칭)	기획 감사실, 문화 공보실, ...
부서명(약식명칭)	기감실, 공보실, ...
부서명(민원명칭)	등초본 발급, 출생 신고, ...

인식 결과는 표 2 와 같고, 음성 전처리부는 실시간 구현이 되며 탐색 과정의 최대 지연 시간은 1.5 초 이내이다.

표 2. On-line 인식 실험 결과

	시도 횟수	성공횟수	인식률(%)
화자 1	438	365	83.3
화자 2	219	202	92.2
화자 3	438	421	96.1
화자 4	441	429	97.3
화자 5	215	190	88.4
화자 6	551	523	94.9
화자 7	551	487	88.4
화자 8	551	496	90.0
계	3404	3113	91.5

5. 결론

본 논문에서는 음성 인식 자동 교환 시스템을 구현하였다. 음성 인식 서비스를 위하여 필요한 기능들을 살

펴보고 실시간 처리하기 위하여 DSP 로 구현하였으며, 인식 성능을 평가하였다. 인식 전처리 과정은 모두 실시간 처리되어 저장되고 탐색 구간은 최대 지연 시간이 1.5 초 이내로 인식이 가능하였다. 또한 사용자의 편의를 위하여 안내 음성이 송출되고 있는 상황에서 인식이 가능하게 반향 제거 기능을 추가하였다.

운영자 측면에서는 인식 어휘의 변경과 시스템의 신뢰성 유지를 위하여 고려된 기능들이 구현되었다. 또한, 전화망에서 인식 성능을 고려하면서도 계산량을 고려하여 캡스트럼 차감법으로 채널 잡음을 제거하였다. 실제 상황에서 온라인을 통하여 실험한 결과 인식 성능은 91.5%를 얻었다.

향후 본 논문에서 구현된 시스템은 인식 어휘에 따라 교환 시스템 뿐만 아니라 증권 시황 정보 제공 서비스나 개인 사서함 관리 서비스등에 다양하게 구현될 수 있다.

참고문헌

- [1] Bahl L.R., Bakis R., Cohen P.S., Cole A.G., Jelinek F, Lewis B.L., Mercer R.L., "Further Results on the Recognition of a Continuously Read Natural Corpus", IEEE International Conference on Acoustics, Speech, and Signal Processing, April. 1980
- [2] Jae-Seung Choi, Jong-Seok Lee, Hee-Youm Lee, "Smoothing and Tying for Korean Flexible Vocabulary Isolated Word Recognition", ICSLP'98
- [3] M.Y. Hwang, X. Huang, "Shared-Distribution Hidden Markov Models for Speech Recognition", IEEE Trans. Speech and Audio Processing, vol. 1, no. 4, 1993, pp.414-420.
- [4] "VIPS-HM 설치 및 사용 설명서", LG 정보통신(주), 1999.1
- [5] Rabiner L. R "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proc. IEEE, vol. 77, no. 2, pp. 257-286
- [6] TMS32C3x User's Guide, Texas Instruments