제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

연속음성인식기술을 이용한 음성인식 증권정보 시스템의 성능 향상에 대한 연구

구 명완
한국통신 멀티미디어연구소 음성언어연구실

A Study on the Performance Improvement of a Stock Information Retrieval System using Continuous Speech Recognition Technology

Myoung-Wan Koo
Spoken Language Research Team, Multimedia Technology Research Lab., Korea Telecom

mwkoo@smm.kotel.co.kr

요 약

본 논문에서는 한국통신이 개발하여 현재 700-3000번으로 서비스되고 있는 음성인식 증권정보시스템을 소개하고, 음성인식 성능을 향상시키기 위한 한국통신의 연구현황을 기술하고자 한다. 현재 운용중에 있는 서비스 시스템은 120명이 동시에 사용할 수 있는 시스템이며 S/W와 H/W를 분리시켜 S/W의 버전을 갱신하더라도 H/W의 변경이 최소화 되도록 설계되었다. 현재 고려하고 있는 성능 향상방법은 연속음성인식 기술을 이용하여 고립단어 인식을 시도하는 것과 거절기능 구현 및 tied-state에 의한 문맥종속 음소를 구하는 것이다. 또한 연속HMM 모델 방식으로의 변경도 연구중에 있다.

I. 서론

음성인식 기술의 실용화 방향은 통신사업자 중심과 컴퓨터 사업자 중심으로 크게 나누어 진다[1]. 통신 사업자 중심의 실용화 방향은 전화 서비스의 품질 향상 및 부가서비스에 음성인식 기술을 이용하는 것이다. 1981년 일본 NTT는 ANSER(Automatic Answer Network System for Electrical Request)라는 전화정보 검색 시스템을 개발하여 은행업무에 대한 정보 서비스를 제공해 왔고, AT&T는 VRCP(voice recognition call processing) 서비스를 개발하여 수신자 부담(collect call), 전화 카드(calling card), 지명통화(person to person) 등과 같이 과거에 교환원에 의해 처리되는 장거리 전화에 대한 서비스를 음성인식을 이용하여 처리하고 있다[1]. 그리고 캐나다의 Northern Telecom에서는 1992년부터 음성인식을 이용한 증권정보 안내에 대한 서비스를 제공하기 시작하였고[2], 미국의 NYNEX에서는 음성으로 전화걸기 원하는 곳을 말하면 전화를 자동으로 걸어 주는 voice dialing서비스를 제공하고 있다[3]. 유럽에서는 MIVA(multilingual interative voice activated telephone services)라는 과제를 통하여 유럽 내의 국가로 출장을 갈 경우 현지 전화를 사용하는 방법 및 긴급 전화번호 등을 알려주는 시스템을 개발하고 있으며, 여기에서는 6개국어의 음성인식 기술을 사용하는 연구를 수행하고 있다[4]. 또한 철도정보를 전화로 알려주기 위하여 음성인식 기술을 이용하기도 한다[5].

컴퓨터 사업자의 실용화 연구는 컴퓨터의 명령을 음성으로 제어하거나 워드프로세서에 음성인식 기능을 제공하는 소프트웨어를 개발하는데 중점을 두고 있다. 마

이크로 소프트 회사는 휘스퍼(Whisper)라는 음성인식 및 합성 소프트웨어를 개발하였으며[6] IBM에서는 "ViaVoice"라는 음성인식 타이프라이터를 개발하였다[7].

국내에서도 최근에는 전화망을 통한 음성인식 시스템 개발에 성공한 사례가 발표되고 있다[8]. L&H화사는 한국어 음성인식 S/W를 개발하여 국내의 기업체에 제공하고 있으며 음성인식 전화정보 시스템 시장에도 진출하고 있다. 몇몇 대학이나 연구소의 실험실에서도 음성인식시스템이 개발되었으나, 아직 상용화를 할 수 있는 시스템은 완성하지 못하고 있는 실정이다.

한국통신에서는 91년부터의 음성언어(spoken language)에 대한 연구 결과물의 하나로, 지난 94년에 음성인식 증권정보시스템을 개발하였고[9][10], 95년도에는 내부 시험을 거쳤으며, 95년 11월 중순부터 실제로 증권회사에 설치하여 1차 시험운용을 하였으며, 1996년부터는 동시에 120명까지 음성인식이 가능한 시스템을 기업체와 개발을 시작하였다. 그리고 1998년 3월 16일부터는 700-3000번으로 시험운용을 본격적으로 진행하고 있다.

본 논문에서는 음성인식 전화정보시스템에 대한 소개와 성능향상을 위한 연구내용을 소개하고자 한다. 서론에 이어 2장에서는 시스템에 대한 소개를 하고, 3장에서는 성능향상에 연구내용을 소개하고자 한다. 그리고 4장에서는 결론을 맺는다.

2. 음성인식 증권정보 시스템

현재 한국통신이 운용중에 있는 시스템은 상용시스템으로 설계되어 있으며, 연구용 시험시스템[10]에 비하여 다음과 같은 특징을 가지고 있다.

o 대용량 음성인식 채널: 120명이 동시에 음성인식 기능을 사용할 수 있는 시스템으로 구성되어 있다.

o DSP의 효율적인 사용: 1 채널의 음성인식을 위하여 2개의 DSP를 사용하는 대신에 특징 추출용 DSP와 탐색용 DSP를 적당한 비율로 재 설계하여 1 채널의 음성인식에 1.xx의 DSP를 사용할 수 있도록 하였다.

o 전화버튼과 음성인식의 동시사용: 전화정보 시스템에서 사용되고 있는 입력방식인 전자식 전화기 버튼을 누르더라도 작동이 가능하도록 설계하였으며, 음성과 동시에 전화기 버튼이 입력이 되면 전화기 버튼에 의해 우선적으로 동작하도록 순위로 두었다.

o 인식 대상단어 자동 갱신: 매일 상장회사 갯수가 조정이 되는 주식시장의 특성을 감안하여 매일 일정한 시간에 음성인식 대상단어가 자동적으로 시스템으로 로딩(loading)되어 인식 대상단어가 갱신되도록 하였다. 그림 1에는 상용 음성인식 증권정보시스템의 개요도가 나타나 있다.

2.1 특징 추출

전화망을 통해 들어온 음성은 8kHz로 표본화되고, (1 - 0.95 z^{-1})의 전달함수를 갖는 필터를 사용하여 pre-emphasis된다. 이 음성은 20 msec의 길이의 프레임(frame) 단위로 분할된다. 이 프레임은 10 msec씩 중첩된다. 자기상관계수(autocorrelation) 방법을 사용하여 14차 LPC 분석을 수행하고, 이 LPC 계수를 이용하여 켑스트럴(cepstral) 계수를 구한다. 이 계수는 아래 수식의 창(window) $W_c(m)$을 사용하여 weighting된다.

$$W_c(m) = 1 + \frac{Q}{2}\sin\left(\frac{\pi m}{Q}\right),\ 1 \le m \le Q \qquad (1)$$

사용되는 음성특징으로는 이렇게 구한 weighted LPC cepstral 계수 외에 이것의 델타, 이차 델타, 로그 파워의 일차값 등이 있다[11]. 이 계수들은 각 종류별로 벡터 양자화된다. 이때 4개의 벡터 코드북(codebook)을 사용하는데, 로그 파워 부분은 64개의 코드 워드(codeword)를 가지며, 나머지 3개는 각각 256개의 코드 워드를 갖는다. VQ 알고리즘으로는 Linde-Buzo-Gray (LBG) 방식을 사용하였다.

2.2 음소 모델

HMM에 기반을 둔 음성인식 시스템을 위해서는 기본단위가 필요하다. 본 시스템은 유사음소(phoneme-like phone)를 기본단위로 사용하였다. 먼저 61개의 문맥독립(context-independent) 음소 모델을 선정하였다.

사용한 모델은 7개의 노드(node)와 12개의 변이(transition)로 구성된다. 이 변이는 3개의 그룹으로 묶

어서 같은 그룹에 있는 변이는 같은 출력 확률을 갖도록 하였다. 그리고 문맥독립 음소 모델을 문맥종속(context-dependent) 음소모델로 확장하였는데, 각 DSP에서 사용가능한 메모리의 크기를 고려하여 300개의 문맥종속 음소모델을 선정하였다. 통계적으로 신뢰성있는 모델을 생성하기 위하여 인식단위 축소법칙(unit reduction rule)을 사용하였다[12].

3. 음성인식 성능향상 방법

3.1 연속음성인식 기술 사용

현재 상용되고 있는 음성인식 기술은 고립단어 인식기술인데, 그림 2에서와 같이 인식 대상 단어 양쪽에 묵음 구간을 두어서 발화자의 잘못된 입력 혹은 주변 소음등에 쉽게 대처하도록 되어 있다. 그러나 연속음성 인식기술을 적용하면 그림 3과 같이 되어 고립단어 인식기술에 비해 묵음 구간이 looping이 있으므로 발화자의 입력방법이 변하더라도 쉽게 적응될 수 있다.

3.2 거절기능 구현

현재 사용되고 있는 거절기능은 filler 모델을 모델링[13]하여 사용하고 있는데 성능향상 및 단어독립 음성인식시스템에 적합한 발화검증 알고리즘을 이용한 거절기능으로 교체할 예정이다[14].

3.3 Tied-state에 의한 문맥종속 음소 사용

문맥종속 음소는 기본 음소로부터 triphone을 구하고 필요한 문맥종속 음소 개수에 따라 unit reduction 규칙[12]을 사용하였다. 최근에는 state의 확률분포를 공유할 수 있어 메모리 양 및 인식시간을 줄일 수 있으며, 인식률을 저하시키지 않는 tied-state에 의한 문맥종속 음소를 사용하는 것을 연구하고 있다. 그림 4에는 3종류의 triphone이 tied-state에 의한 문맥종속 음소로 어떻게 변화되는지를 나타내고 있다. 기존의 9개 확률밀도를 5개의 확률밀도로 변화시켜서 메모리 감소를 야기 시켰다. 그림 5에서는 tied-state를 만드는 데 사용되는 규칙이 적용되는 예를 나타내었다.

3.4 연속HMM 모델 방식

이산 HMM 모델방식은 인식시간이 빠르다는 장점이 있으나 모델능력이 연속 HMM 모델방식에 비해 떨어진다. 연속 HMM 모델방식은 실시간 처리가 어렵다는 단점이 있으나 최근에는 N-mixture 개념과 covariance 행렬을 diagonal로 정의함으로써 인식시간을 줄일 수 있게 되었다. 한국통신도 연속 HMM 모델방식을 연구하고 있다[15].

3.5 기본음소 개수 설정

현재는 59개의 기본음소를 사용하고 있으나, 음소인식률을 최고로 할 수 있는 기본음소 개수는 자동으로 구할 수 있는 알고리즘을 제안하여 실험중에 있다[16]

4. 결론

본 논문에서도 한국통신이 개발하여 현재 700-3000번으로 서비스해 주고 있는 음성인식 증권정보시스템의 상용화 버전에 대한 소개를 하였으며, 성능향상을 위해 작업 중에 있는 연구를 기술하였다. 성능향상을 위해서 연속음성인식 기술 적용, 거절기능 구현, tied-state에 의한 문맥종속음소 사용 및 연속HMM 모델방식을 연구하고 있으며 기본 음소 개수를 정하는 알고리즘도 구현하고 있다

참고문헌

[1] D. B. Roe et al., "AT&T speech recogntion in the telephone network," Speech Technology, pp. 16-21, Feb/March, 1991.

[2] M. Lenning, et al., "Flexible vocabulary recognition of Speech", *Proceedings ICSLP92*, Vol. 1, pp. 93-96 1992.

[3] G. J. Vysotsky, "VoiceDialing - The first speech recognition based service delivered to custom's home from the telephone network.", *Speech Communication*, Vol. 17, pp.235-247, 1995.

[4] C. Sorin, et al., "Current and experimental application of speech technology for telecom services in Europe," Proc. of IVTTA'96, pp.1-6, Sep. 1996.

[5] R. Billi, et al., "Field trial evaluation of two different information inquiry system," Proc. of IVTTA'96, pp. 129-134, Sep. 1996

[6] Microsoft URL: http://research.microsoft.com/stg

[7] L. R. Bahl, et al., "A fast approximate acoustic match for large vocabulary speech recognition," IEEE Trans. Speeh and Audio Proc., vol. 1, pp. 59-67, 1993.

[8] 구 명완, "음성인식기술의 현황과 실용화 전망," '98음향학회 하계학술대회 발표예정

[9] 도삼주, 김우성, 장두성, 구명완, "음성인식기술을 이용한 증권정보 안내시스템의 실험적 실용시험", 11회 음성통신 및 신호처리 워크샵 논문집 1호, pp. 241- 244, 1994.

[10] M. W. Koo et al., "KT-STOCK : A speaker-independent, large-vocabulary speech recognition system over the telephone", in *Proc. 1994 IEEE Int. Conf. Spoken Language Processing*, Sep. 1994.

[11] C.H. Lee et al., "Acoustic modeling for large vocabulary speech recognition," Computer Speech and Language, No. 4, pp.127-165, 1990

[12] C.H. Lee et al., "Acoustic modeling for subword units for speech recognition," in Proc. 1990 ICASSP, pp.721-724, Apr. 1990.

[13] 구명완, "신경망을 이용한 음성인식 거절기능 구현", 제13회 음성통신 및 신호처리 워크샵 논문집 pp.207-211, 1996

[14] 김우성, 구명완, "발화검증에 의한 음성인식 거절기능 연구", '98음향학회 하계 학술대회 pp67_70, 1998

[15] 구명완, "HMM 훈련 알고리늠에 따른 음소인식률 비교 연구" 음성통신 및 신호처리 워크샵 논문집 발표 예정, 1998

[16] 김호경, 구명완, "기본음소 설정을 위한 음소 인식률 이용방안 연구" 음성통신 및 신호처리 워크샵 논문집 발표예정, 1998.

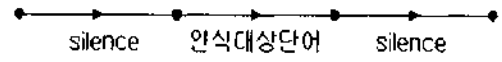

그림 2. 고립단어 인식

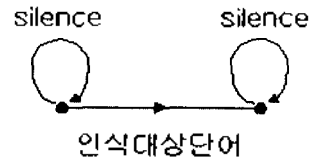

그림 3. 연속음성인식 기술을 이용한 고립단어

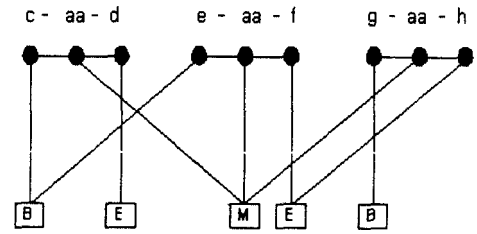

그림 4. Tied-state에 의한 문맥 종속 음소

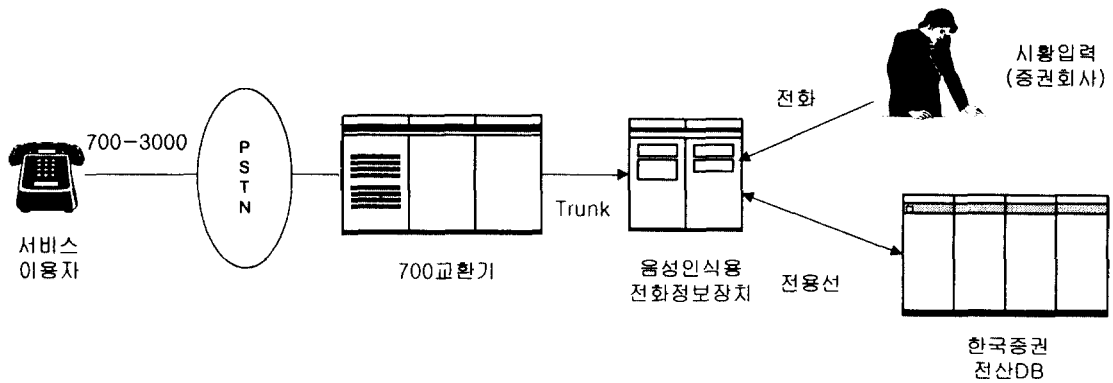

그림 1. 상용 음성인식 증권정보 시스템 개요도(700-3000번)

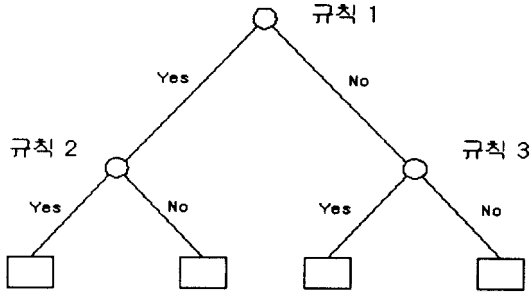

그림 5. 음성 규칙에 의한 결정 트리