

대용량 음성인식 전화정보시스템 개발 및 시험운용

°장 경애, 김 재인

한국통신 멀티미디어연구소 음성언어연구소

The Development of a Speech Recognition System with Large Channel over the PSTN and it's Field Trial

Kyung-Ae Jang, Jae-In Kim

Spoken Language Research Division, Multimedia Technology Research Lab., Korea Telecom

(kajang,jikim)@smm.kotel.co.kr

요약

본 논문은 대용량, 대어휘, 화자독립 음성인식시스템의 개발과 1998년 3월 16일부터 일반인들을 대상으로 시험운용하고 있는 음성인식 증권서비스에 대한 것이다. 이 시스템은 상용서비스를 위하여 한대당 120명이 동시에 사용할 수 있는 대용량 시스템으로 HMM(Hidden Markov Model)기술에 기반을 둔 고립 단어 인식시스템이다. 이 시스템은 음소를 기본 인식단위로 사용하여 인식단어의 추가 및 변경이 자유로우며[1], 추가 또는 변경된 회사명칭을 운용자의 개입없이 자동적으로 시스템에 반영될 수 있다. 본 서비스의 개발과정에서 인식대상단어를 결정하는데 발생된 문제점과 인식단어의 변경방법 및 적용후의 효과 등을 살펴 보았다.

1. 서론

음성인식 전화정보서비스는 전화기를 통한 기존의 음성정보서비스(700번 서비스)에 음성인식기능을 부가한 서비스로서, KT에서는 95년 11월에 이미 PC를 기반으로 한 5채널 규모의 소용량 음성인식시스템으로 대우중

권에서 시험운용을 하였고[2], 이를 기반으로 상용서비스를 위한 수용용량이 큰 시스템의 개발이 요구되어, 96년부터 97년말까지 128채널 규모의 음성인식 시스템을 개발하였고, 현재 이를 이용하여 시험운용 중이다.

PC기반의 프로토타입 시스템을 2년간 대우증권에서 시험운용 하면서 발생된 가장 큰 문제점 중의 하나는 새로운 상장회사명칭을 상장되기 전날에 시스템에 적용하는 것이었다. 새로 상장된 종목에 대한 문약은 상장되는 날에 가장 빈번하므로 주식시장의 변화 즉, 인식대상단어의 변화가 음성인식시스템에 적시에 반영되는 것이 매우 중요하다. 현재 시험운용중인 상기 시스템에서는 회사명 변경 한 후, 인식단어리스트를 작성하고, 발음사전을 생성하여 시스템에 반영하는 모든 과정을 운용자의 개입없이 자동으로 수행하도록 하였다.

또 서비스를 제공하면서 문제되었던 것은 음성인식 증권정보서비스에서 인식대상단어를 정하기가 어려운 점, 인식시스템을 전혀 사용해 보지 않은 사람을 고려한 시나리오 구성, 음량이 작은 사용자를 위한 점등이 있었다.

본 논문에서는 상용서비스를 위한 대용량 음성인식 전화정보시스템의 개발과 음성인식 증권정보서비스의 개발시 고려되어야 할 사항과 서비스과정에서 나타난 문제점등과 운용결과들을 살펴 보았다.

2. 음성인식 시스템의 대용량화

기존의 전화정보 시스템은 채널당 가격이 100만원이 하로 형성되어 있고, 신뢰도도 안정되어 있기 때문에 음성인식 전화정보 시스템을 처음부터 새로 개발하는 것 보다 기존의 전화정보 시스템에 음성인식기능을 추가 하는 것이 개발비와 개발기간을 줄일 수 있고 또한 시스템의 전반적인 가격 상승을 최소화 할 수 있다. 개발된 시스템은 음성인식 S/W를 위한 H/W를 만들어 전화정보 시스템에 추가하였다. PC를 기반으로 한 프로타입 시스템에서는 PC 한대 당 한채널을 수용하였고, 채널당 두 개의 DSP(Digital Signal Processor)를 사용하였다. 두 개의 DSP중 첫 번째 것은 입력된 음성을 분석하는 일을 담당하고, 두 번째 것은 음성을 인식하는 일을 담당하고 있다. 첫 번째 DSP는 사용자의 음성입력을 감지 해야 하기 때문에 서비스 제공시에는 항상 연결되어 있어야 하지만, 음성인식을 담당하는 두 번째 DSP는 사용자 음성이 감지 된 뒤에만 연결되어도 되므로 다른 채널과 공유하여 사용될 수 있다. 두 번째 DSP를 다른 채널과 공유하면 시스템내에 추가되는 H/W의 줄일 수 있고 따라서 시스템 가격 상승을 낮출 수 있다. 개발된 시스템에서는 약3개의 음성분석 DSP를 1개의 음성인식 DSP가 담당하도록 설계되어 있다. 시험운용 결과에 의하면 사용자가 음성을 입력하는 시간과 인식결과를 듣는 시간이 평균1:11이었지만 우선 위와같은 비율로 제작하였다.[4]

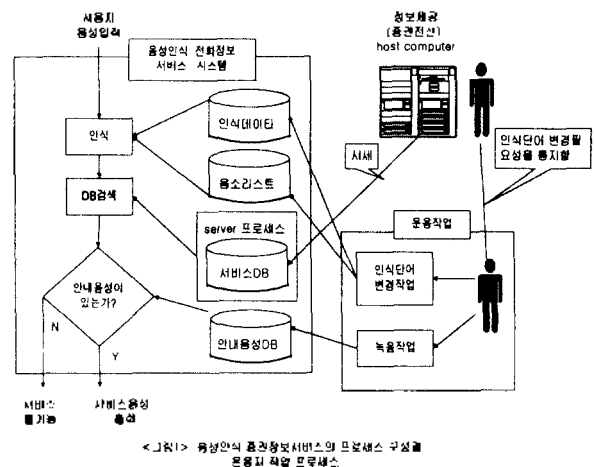
3. 인식단어의 관리

3.1 인식대상단어의 변경

음성인식 전화정보시스템은 원하는 정보를 찾기 위하여 서비스코드를 기억하여 버튼으로 입력해야하는 대신 음성입력을 사용하므로 사용자측면에서는 매우 편리한 반면, 운용자 측면에서는 서비스단어 즉, 인식대상단어가 변경되면 <그림1>에서와 같이 인식관련 데이터를 변경해야 하는 번거로움이 있다.

<그림1>의 음성인식 증권정보서비스 시스템은 사용자음성이 들어오면 인식단어를 참고하여 인식작업을 수

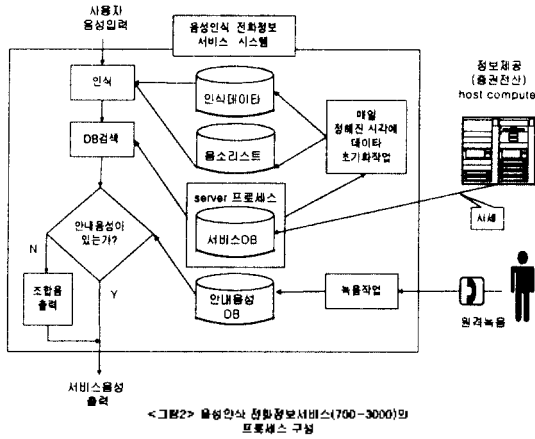
행하고 그 결과에 따라 서비스DB를 검색하여, 그 결과를 음성으로 사용자에게 출력하여 준다. 서비스DB는 증권전산에서 실시간으로 정보를 제공하는 host computer에 의하여 실시간으로 update된다. 새로운 회사명을 인식하기 위해서는 인식단어 데이터와 음소리스트가 변경되어야 하며, 새로운 회사명에 대한 녹음작업을 수행하여 안내음성DB가 변경된 후에, 그 회사에 대한 서비스가 가능해진다. 이러한 일련의 변경작업은 변경 필요성을 운용자가 바로 알 수 없으며, 증권회사나 다른 경로로 연락받아야 되는데, 정확한 날짜에 정확한 회사명을 전달받아야 하고, 상장당일 상장회사에 대한 문의가 가장 많다는 점을 고려할 때, 상장일 전까지 정확하게 인식단어를 변경하는 작업은 운용자에게 큰 부담이 된다.



<그림1> 음성인식 증권정보서비스의 프로세스 구성과 운용자 작업 프로세스

이러한 부담을 해결하기 위하여 인식단어도 일반데이터와 같은 방식으로 증권전산으로부터 받아서 운용자의 개입없이 자동으로 처리하도록 하였다. <그림2>에서는 음성인식 증권정보시스템에서 인식단어리스트가 전달되면 음성인식 전화정보시스템에서 자동으로 음소리스트를 생성하고 DSP프로그램을 변경시켜 서비스를 재시동하는 작업을 수행하는 과정을 나타낸다. 정보제공 host computer는 매일 새벽 상장 또는 변경내용을 포함하여 그날 서비스할 종목리스트를 정하여 매일 아침 한번 server 시스템으로 전송한다. 이때 시스템은 매일 일정한 시간이 되면 시세데이터와 함께 단어리스트를 자동적으로 sever로 요구하여 수신받고, 인식데이터를 변경하여 그날 서비스 할 인식단어들로 시스템을 초기

화 시킨다.



<그림2> 음성인식 전용정보서비스(700-3000)의 프로세스 구성

하지만 어떤이유에서든 host에서 보내온 data가 불완전한 경우 서비스를 제대로 해 줄 수 없기 때문에 전날의 상장회사수에 대한 정보를 이용하여 전송받은 데이터의 타당성 여부를 판단하여 사용한다.

3.2 인식대상단어의 선정

음성인식 증권정보서비스에서는 사람들이 부르는 회사이름에 대한 선정이 경우에 따라 어려운 경우가 있다.[3] 첫째, 사람들이 부르는 이름과 증권시장에 상장된 회사명칭이 다른 경우이다. 실제로 대우증권에서의 시험서비스의 경우에도 회사이름이 4자가 넘는 경우 4자정도로 축약해서 사용하는 경우가 많았다.[3] 물론, 이것은 음성인식서비스가 초기단계여서 그 동안 회사이름을 사용하지 않고 코드입력으로 습관화된 사용자들은 회사의 정식명칭보다는 단축명령을 부르는 것에 익숙해져 있으며 정식명칭을 부를 필요성을 느끼지 못하고 있다. 본 서비스 시스템 설계 당시 이러한 문제를 해결하기 위하여 <표1>과 같이 회사의 정식명칭(또는 단축명칭)이 생소하고 어색한 경우에는 단축명칭(또는 정식명칭)으로 등록하도록 하였다.

<표 1> 정식명칭과 단축명칭의 예

정식명칭	단축명칭
동화약품공업	동화약품
동아건설산업	동아건설
영창악기제조	영창악기
동양에레베이터	동양에레

둘째, 한 회사이름을 여러 다른 이름으로 부름으로써 한 이름으로 정할 수 없는 경우로써, 서비스 운용도중에도 몇차례 수정이 있었다. 근본적인 해결을 위하여 동일회사에 대한 여러 이름을 수용하는 방법이 연구되고 있다.

<표 2> 동일회사의 다른이름의 예

엘지정보통신/엘지정보
현대자동차/현대차
한국전력/한전/한국전력공사

셋째, 회사이름은 고유한 것이므로 특이하고 부르기에 부적합하더라도 틀린 이름이라고 할 수가 없으며 서비스에서 수용하여야 하는 경우가 있다. <표3>에서와 같이 같은 업종을 가진 회사들간에 일관성없는 이름 및 외래어의 부자연스런 발음법을 그대로 회사명으로 정한 경우 등이 있다. 음성인식의 관점에서 볼 때 이러한 이름들은 인식을 낮 사용자 만족도에 큰 영향을 끼친다. 앞으로 음성인식이 활성화되어 인식이 잘되는 이름이 각 회사들의 이익과 연관된다면 자연스럽게 회사들의 명명방법에 변화가 있을 것이라 생각된다.

동양에레베이터/현대엘리베이터
진로인더스트리즈/화승인더스트리/ 삼우인다스트리얼
현대차서비스/기아차서비스

<표 3> 회사들간의 일관성없는 명명의 예

4. 음성인식 증권정보 서비스의 시험운용

4.1 이용호수(채널당 사용호수)

시간대별 통화분포는 그림3에 나타내었다. 그림2를 보면 호의 대부분이 집중시간대에 몰려 있는 것을 볼 수 있다. 여기서 집중시간대란 증권시장에서 증권을 거래할 수 있는 시간 즉 평일에는 오전장(9시반 - 11시반) 과 오후장(오후1시- 3시)을 토요일에는 오전 장을

대용량 음성인식 전화정보시스템의 시험운동

말한다.

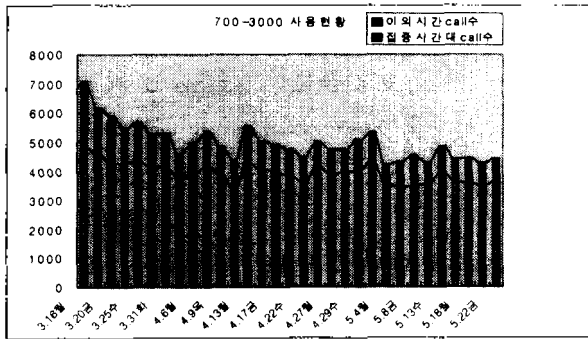


그림 3. 시간대별 사용호수

이 때문에 거래를 할 수 있는 시간을 중심으로 통화가 집중되는 경향이 있다.

4.2 채널당 평균사용호수

채널당 평균 사용호수에 대한 변화를 그림 4에 나타내었다. 분석결과 약 156호를 상회하고 있으며, 이는 대우증권에서 시험서비스를 한 경우의 85호보다 배이상 증가하였다.

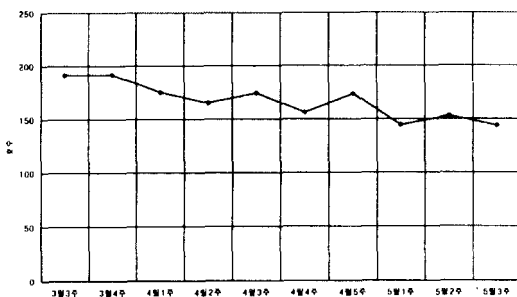


그림 4. 채널당 평균호수

4.3 DTMF와 음성인식사용비율

개발된 시스템은 음성입력과 DTMF입력을 사용할 수 있다. 처음에는 DTMF입력에 대한 멘트없이 허용하다가 후에 DTMF입력이 가능하다는 멘트를 넣어서 사용자들이 선택하여 사용하도록 하였다. 두가지 방식에 대한 사용 비율을 그림 5에 표시하였다.

그림에서와 같이 사용비율은 82%정도를 상회하고 있으며 이는 DTMF를 사용하는 것 보다 음성인식기능을

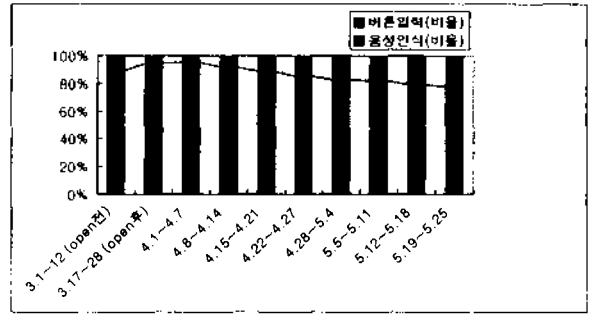


그림 5. DTMF와 음성인식의 사용비율

사용하는 것이 편하다는 것을 나타내고 있다. 실제로 DTMF를 자주 이용하시는 분도 자신이 알고 있는 회사 이외에는 DTMF입력이 아닌 음성인식기능을 이용하여 서비스를 사용하고 있었다.

4.4 완료율

개발된 시험시스템은 교환기에 연결되어 있기 때문에 완료율을 알 수 있다. 이에 대한 결과를 그림6에 나타내었다. 그림6에서는 앞에서 설명한 호 집중시간대의 완료율에 대한 것을 나타내었는데, 그림 3에서와 같이 대부분의 사람들이 주식거래를 할 수 있는 시간대에 조회를 하고 있는 것을 알 수 있다.

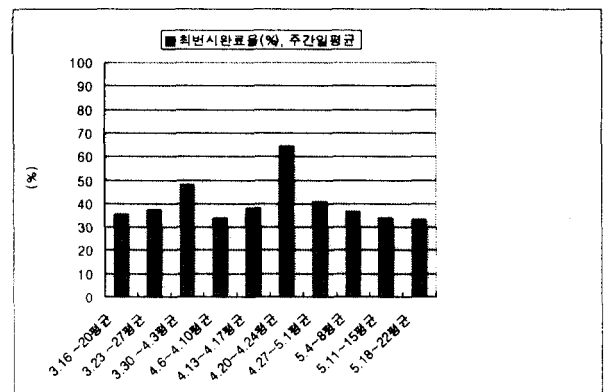


그림 6. 최번시 완료율

그리고 시스템의 용량을 30채널정도로만 운용하고 있기 때문에 아직 완료율이 그리 높지 못함을 볼 수 있다.

5. 개선되어야 할 점

서비스 개시한 시점에 사용량과 그로부터 얼마간 시

간이 지난뒤의 사용량에 차이점이 발생하는데 그이유중에 하나가 처음 사용자에 대한 배려가 없어서 서비스의 사용을 포기한 경우가 있다고 판단되어, 사용법안내를 사용자가 원하던 원하지 않던 무조건 방송되도록 수정하였다. 또한 사용자의 특성에 시스템을 적용시키려는 시도로 초심 사용자의 음량이 너무 작아 끝점검출기가 이를 검출하지 못하는 점을 해결하기 위하여, 무입력으로 입력시간이 초과되었을 경우 끝점검출기의 임계치를 자동적으로 재조정하는 방법을 시험중에 있다.

6. 결 론

음성인식 전화정보서비스는 일반 전화정보서비스(기존의 700서비스)와 달리 사용자에게 들려줄 데이터 뿐만 아니라 음성인식 기능을 위한 인식관련 데이터의 관리가 필요하다. 인식관련 데이터의 관리는 사용자들이 사용할 인식대상단어를 선정하는 방법과 변경작업을 포함한다.

본 논문에서는 이와 같은 인식대상단어의 선정과 효과적인 변경방법을 다루었다. 서비스의 사용자 및 시스템의 고객들의 인식개념의 부재로 인하여 인식단어 선정이 어려웠으며 우선 운용초기에 단어를 변경하기도 했으나, 근본적인 해결책으로서 동일종목명의 다른 여러 이름을 수용하는 방법을 연구중이며, 명명방법의 점차적인 변화 등을 고려할 수 있겠다.

그리고, 서비스 종류에 따라 정보의 변경 및 인식대상단어의 변경 빈도가 높은 경우가 있으며, 인식데이터 변경작업이 적절한 시간에 정확하게 수행되기 위해서는 변경작업의 자동화는 반드시 필요하다. 음성인식 중권정보서비스에서는 운용자가 특별히 음성인식에 관련된 일상적인 작업을 하는 일은 없도록 하였으며, 출력할 안내음성이 없을 경우를 대비하여 조합음을 출력하도록 하였다.

6월 말까지의 운용결과, 총 30채널을 기준으로 채널당 일일 평균 156호를 사용하였고, 한 호당 평균 127초를 이용하였으며, 하루중 최번시 통화완료율은 40%수준으로서 앞으로 서비스채널을 확장할 계획이다. 시스템은 DTMF와 음성인식을 동시에 수용하고 있는데, 음성인식기능의 사용비율은 평균 82%로 나타났다.

운용하는 동안 실제 일반 사용자들의 문의전화를 받아 음성인식 응용서비스의 사용자 반응을 조사하고 있을 뿐 아니라 서비스 시나리오 및 인식기능을 운용자들과 사용자들의 성향에 맞추어 보완하고 있다. 앞으로도 계속 운용하면서 rejection이나 keyword spotting과 같은 새로운 기능에 인식기를 적용하여 상용서비스를 할 수 있는 음성인식시스템으로 개선해 나갈 예정이다.

참고문헌

- [1] 전예임, 박상규, 구명완, "단어독립 음성인식 시스템을 위한 운용시스템 개발", SCAS-12권 1호, pp. 65-68, 1995
- [2] M. W. Koo, et al, "A Stock information system over The telephone network", Proc. of Int. Conf. on Signal Processing Applications & Technology, pp. 2039~2043, Boston, Oct. 1995
- [3] 김재인, 구명완, "음성인식 중권정보시스템의 운용 결과", 한국통신학회 '96하계종합학술대회 논문집, 1996.7.4 ~ 7.5,
- [4] Seong-June Park, Jae-In Kim, Chu-Shick Jhon, "The architecture of a speech recognition system for cost reduction", Proceedings of ICSP'97, August 26-28, 1997, Seoul, Korea