

□ 사례 발표 □

음성인식 증권정보서비스 구축 사례

감 지 은[†] 이 문 형^{**}

◆ 목 차 ◆

- | | |
|-----------|--------------|
| 1. 서 론 | 4 서비스분석 및 전망 |
| 2. 음성인식 | 5. 결 론 |
| 3. 서비스 구성 | |

1. 서 론

증권정보를 위한 서비스의 형태는 신문과 뉴스, 컴퓨터, 전화서비스 등 여러 가지가 있다. 그러나 증권정보 서비스의 경우는 형태가 전화서비스나 인터넷 어느 경우라도 사용자가 원하는 서비스나 알고자하는 회사의 고유코드를 입력하는 방법으로 주로 이루어져왔다. 이러한 방법은 사용자가 코드를 알고 있지 않을 경우는 매우 불편하다.

음성인식(Auto Speech Recognition)의 기술은 이러한 불편함을 줄일 수 있다. 음성인식은 미래의 사회를 그리는 첨단과학영역 속에 자주 등장하는 기술로 인간사이의 중요한 통신수단인 음성을 인간과 기계사이에서도 가능하게 만들 수 있는 기술이다. 불과 아십여 년 전만 해도 이 기술은 먼 미래에나 가능할 것으로 생각되었지만 현재 부분적으로 벌써 상용화가 이루어지고 있다. 최근의 기계, 전자산업의 발달과 급격히 향상된 컴퓨터의 성능으로 기계, 컴퓨터의 보급률이 높아지면서 인간과 기계사이의 통신부분도 급속한 성장을 이루었고 그 결과로 최첨단 기술인 음성인식의 기술이 꽃을 피우게 된 것이다.

자동응답(ARS)의 형태로 증권정보를 알려주는

시스템에(700-3003) 음성인식의 기술을 적용하였다. 전화를 걸어 종목코드를 입력하는 것이 아니라 종목명을 말하면 시스템이 이를 인식하여 알려주는 것이다. 이 신기술의 사용으로 시스템에 말로써 문의하고 서비스를 받을 수 있게 되어 영화속에서나 보던 서비스를 실용화한 것이다.

2. 음성인식

음성인식의 기술은 최근 크게 부각되는 기술로서 적용할 수 있는 범위가 매우 넓다. 특히 통신 분야는 음성인식 기술이 사용된다면 그 장점이 더욱 크게 발휘될 것이다. 그 때문에 국내에서 가장 먼저 음성인식 기술이 실용화된 부분도 전화기이다. 그리고 최근에는 PCS나 Cellular phone에서 실용화되어 음성인식의 기술은 이미 생활 속에 있다.

2.1 음성인식 시스템 분류

음성인식은 사용하는 기술의 종류에 따라 몇 가지로 분류할 수 있다.

첫째로 사용자의 형태에 따라 누구나 인식 가능한 화자독립 인식 시스템과 특정화자만을 인식할 수 있는 화자종속 인식 시스템을 구성할 수

† 정회원 : 범일정보통신(주) 연구원

** 정회원 : 한국증권전산 사업부 부장

있다. 화자독립 인식이란 화자에 관계없이 누구나 말을 해도 기계가 알아들을 수 있게 하는 방법이며 화자종속은 특정화자의 말만을 인식 할 수 있는 방법을 말하는데, 특히 화자종속의 경우 자기 자신만을 인식하고 다른 사람들은 거부하는 화자 인식 시스템을 구성할 수 있어 보안용으로도 사용될 수 있다.

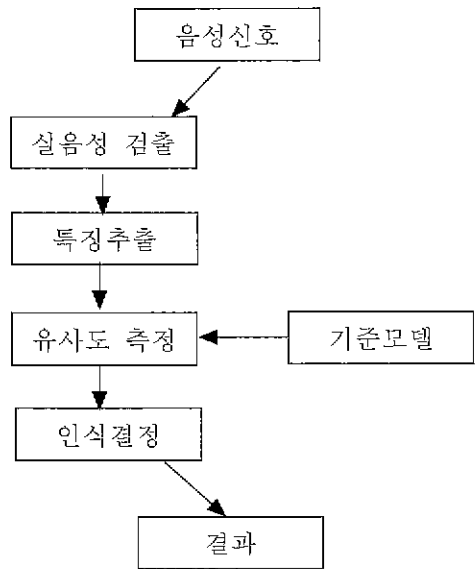
둘째로 화자의 발성에 따라 단어단위로 발성하여 인식하는 고립단어(Isolated Word) 인식, 단어들을 연결하여 발성하는 연결단어(Connected Word) 인식, 화자의 말(문장)에서 특정 단어만을 인식하는 핵심어 인식(Keyword Spotting), 그리고 평상시 말하는 자연스러운 대화체를 인식할 수 있는 연속음(Continuous Speech) 인식 등으로 분류할 수 있다. 물론 마지막으로 기술한 연속음 인식이 가장 어려우며 또한 앞으로 음성인식의 방향으로 볼 수 있다.

마지막으로 인식 가능한 어휘 수에 따라 대용량 단어인식과 소용량 단어인식이 있다. 소용량의 단어를 인식하는 경우는 인식 단위를 단어단위로 할 수 있기 때문에 고립어나 연결단어 인식 등에 사용되어 왔고 지금까지 많은 연구활동 및 제품들이 있었다. 그러나 대용량 단어는 수천 단어에서 수만 단어를 인식하여야 하기 때문에 이러한 방법으로는 곤란하다. 그래서 음소단위나 그와 비슷한 단위로 인식하는데 이러한 방법은 어떠한 형태의 인식기에도 사용할 수 있지만 그 방법이 어려워 아직까지 많은 제품들이 있지는 못하다. 그러나 많은 연구들이 이루어지고 있으므로 머지않아 모든 인식기들에 사용되리라 본다.

2.2 음성인식 시스템 구성

모든 음성인식 시스템들이 다음과 같은 구성을 가지고 있다. 우선 입력된 신호 중에서 실제 사람

이 발성한 음성신호만을 검출하여 특징을 추출하고 이것으로 기존 음성모델과의 유사도를 측정해 최종적으로 인식하게 되는 것이다. 그러나 시스템마다 전처리 단계를 가지거나 조금씩 다른 특징들을 추출함으로써 서로 다른 특성을 지니는 시스템을 구성하게 되는 것이다.



(그림 1) 음성인식 시스템의 구성

음성신호를 입력받게되면 실제 화자가 발성한 음성부분만을 검출하여야 하는데 이 실음성 검출 부분은 인식성능에 매우 큰 영향을 미친다. 왜냐하면 아무리 좋은 음성인식 알고리즘을 사용한다 하더라도 음성검출이 제대로 이루어지지 않았다면 좋은 인식률을 기대하기 어렵기 때문이다. 특히 전화선을 통한 발성에서의 같이 잡음이 많은 환경에서는 검출된 구간에 잡음이 포함되거나 음성부분이 아닌 잡음 구간만이 검출되는 경우가 많다. 따라서 음성인식에 앞서 정확한 음성검출이 요구된다.

실음성 구간을 검출한 다음에는 음성특징을 추출한다. 이 부분은 일종의 음성압축 부분으로서

불필요하게 중복되는 음성정보를 없애고 동일 음성 신호들간의 일관성을 높임과 동시에 다른 음성 신호와는 변별력을 높일 수 있는 정보를 추출한다. 방법은 인간의 발성기관을 모델링 하여 음성의 특징을 나타내는 계수들을 구하는 것이다. 이를 특징계수 혹은 특징벡터라고 한다.

이렇게 구한 특징 계수들을 가지고 실제 인식에 들어가게 되는데 이때 사용되는 음성인식 알고리즘은 여러 가지가 있으나 대표적인 것으로 VQ(Vector Quantization), DTW(Dynamic Time Warping), NN(Neural Network), HMM(Hidden Markov Model) 등이 있다.

VQ란 앞서 구한 특징벡터(계수)들을 기존모델인 코드북(Codebook)과 mapping시켜 대표 값으로 부호화 하는 방법이다. DTW는 같은 사람이 같은 발음을 해도 신호의 길이가 시간에 따라 달라지는 음성의 동적 특성을 고려해 기존모델과 비교할 때 시간 축에서의 차이를 보상하는 방법이다. NN이란 신경망으로 사람의 뇌를 모델링 한 것이다. 즉 사람이 학습하는 방법처럼 같은 음성을 반복하여 훈련시킴으로써 시스템 스스로 특징을 알아내는 계수를 구하게 한다. 그리고 HMM은 음성을 확률적으로 모델링 하여 학습시키는 방법으로 대용량 어휘 인식에 주로 사용되는 알고리즘이다.

그리고 최종적으로 인식한 결과가 올바른 것인지를 확인하는 부분이 있어야 한다. 위의 과정을 거쳐 인식한 음성이 기대하는 말이 아니었을 경우 이를 오인식 하는 것이 아니라 거절할 수 있어야 한다. 즉, 잘못된 음성 입력에 대한 검증 부분까지 포함되어야만 신뢰성이 있는 시스템이라고 할 수 있는 것이다.

2.3 음성인식 시스템의 조건

음성인식이란 간단히 말해 사람의 말을 기계에

게 전달해 원하는 동작을 유도하는 것이다. 결국 사람과 기계의 자연스러운 통신이 궁극적인 목적이라고 볼 수 있다. 따라서 불특정 화자가 언제, 어디에서 무슨 말을 하더라도 이를 인식하는 것이 최종적인 목표이다.

다시 말해, 사용환경에 영향을 받지 않는 화자 독립 무한 연속음 인식이 되어야만 한다. 또한 이러한 시스템을 사용하는 사용자의 부주의로 인한 영향도 고려해야 한다. 그러나 이러한 것은 최상의 조건으로 아직까지 이러한 제품은 없다.

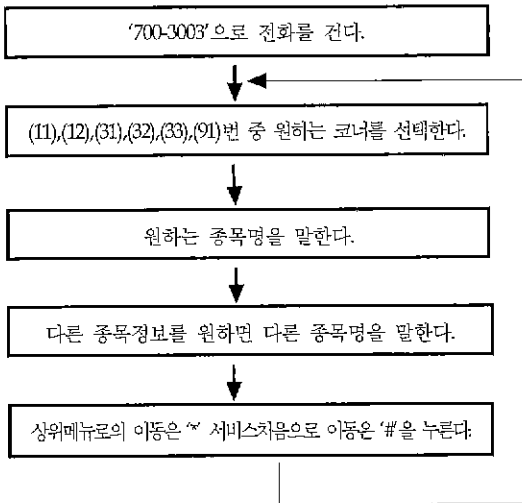
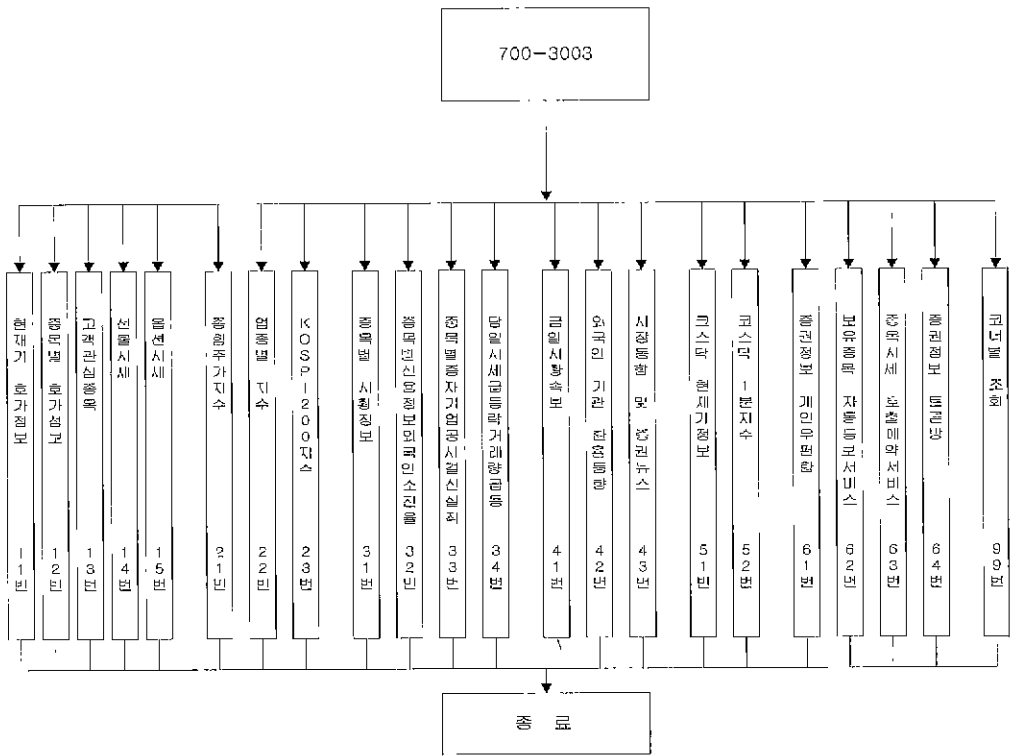
하지만 사용환경에 적합한 시스템을 고려할 수는 있다. 사람과 사람의 통신에도 100% 인식이 불가능하다고 한다. 실험에 의하면 인간의 고립되어 인식 능력은 98%라고 하는데 이 것과 비슷한 정도의 인식성능을 가진 시스템이라면 사용자가 불편함 없이 사용 가능하다고 볼 수 있으므로 상용화 할 수 있다. 그리고 인식대상 어휘도 사용 목적에 따라 다르므로 제한된 어휘 내에서 서비스 가능한 시스템을 구성할 수 있다. 즉, 시스템을 사용할 환경에 적합하게 구성하면 지금의 기술로도 상용화가 가능한 것이다. 특히, 서비스가 제한적인 전화의 경우는 조금만 고려하면 음성인식 서비스를 충분히 할 수가 있다.

3. 서비스 구성

3.1 자동응답 서비스의 적용

증권정보를 위한 자동응답 서비스(700 서비스)에 음성인식을 적용하였다

전화서비스의 장점은 원하는 정보만을 골라 서비스를 받을 수 있기 때문에 많은 양의 정보 속에서 원하는 정보를 찾는 수고를 줄일 수 있다. 그리고 이러한 전화서비스는 안내원 없이, 동시에 여러 사용자를 수용할 수 있는 자동응답 서비스의 형태가 대부분을 이루고 있다. 그러나 지금까지의 자동응답 서비스는 서비스마다 고유한 번호



수 있다는 장점이 있으나 사람이 아닌 프로그램에 의한 서비스이므로 제한이 많고 사용자의 입장에서 본다면 불편한 점이 많다. 우선, 원하는 서비스의 코드를 모를 경우 이를 알기 위해서 처음부터 안내를 들어야 하고 또 이러한 방법에 익숙하지 않은 사람은 실수를 할 경우가 많다. 이런 경우는 다시 처음부터 안내를 따라야 하는 경우가 많아 사용자는 매우 불편함을 느끼게 된다. 또한 회사마다 고유한 번호가 있어 알고자 하는 상장회사의 코드도 미리 알고 있어야 한다. 이러한 단점을 보장하는 방법으로 음성인식을 사용하였다. 회사의 고유한 코드가 아닌 회사명을 음성으로 인식하여 정보를 알려주는 것이다.

를 부여해 사용자가 원하는 정보를 얻기 위해선 반드시 해당 번호를 입력해야만 했다. 이러한 방법은 물론 이전에 안내원과 직접 통화를 하는 경우보다 훨씬 많은 양의 정보를 동시에 서비스할

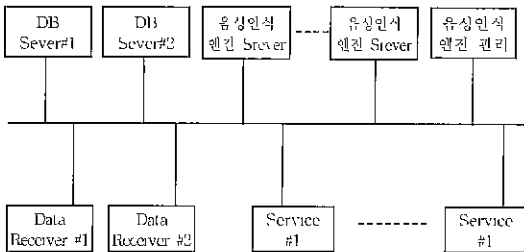
그러나 전화 상에서의 음성인식은 주변잡음 외에도 전화선상의 고유한 잡음이 있어 일반적인 마이크를 사용해 인식을 하는 경우보다 음성의 상태가 나쁘다. 또한 무선 전화기를 사용하는 경

우도 많으므로 음질의 상태가 더욱 나빠질 수 있다. 따라서 이러한 점을 고려하여 일반적인 음성 인식과는 다르게 인식 시스템을 구성하여야 한다.

전화에서 사용하는 음성을 입력받을 때 8KHz, 16bit로 sampling한 후, 잡음상태를 고려해 음성구간을 검출한다. 특징계수는 잡음환경에 강한 계수를 사용한다. 인식시스템을 훈련(training)시킬 때도 전화선상에서 채취된 음성으로 사용하여야 한다. 그리고 모든 서비스가 실시간으로 처리되어야 한다. 사용자가 발성한 뒤 바로 응답이 나와야만 이러한 서비스가 비로소 진가를 발휘하게 될 것이다.

3.2 시스템 구성

전체 시스템은 다음과 같이 구성하였다.(그림 2)



(그림 2) 전체 시스템의 구성도

4. 서비스 분석 및 전망

음성인식 증권정보 서비스는(700-3003) 1999년 1월 4일 시범서비스를 시작해서 3월 2일 인천에서 상용서비스를 먼저 시작했다. 그후 3월 8일부터 서울을 비롯하여 4월 6일 전국에서 서비스를 개시하였다. 처음 서비스를 개시한 3월에는 하루 평균 3만 여 통화가 있었으나, 4월부터는 5만 여 통화 이상이 되고 있다.

음성인식이 서비스 되고있으나 코드를 누르는

방법에 익숙한 사용자들이 많이 사용하고 있어 음성인식을 사용하는 빈도는 처음에는 30%정도였으나 이후 조금씩 증가하는 추세이다. 그리고 인식률은 사용자마다 사용하는 환경이나 전화기가 달라서 개인차가 심하지만 일반적으로 평균 80 - 90 %정도의 인식률도 보이고 있다. 그러나 몇몇 사투리를 심하게 쓰는 사용자나 몇몇 단어들어 매우 저조한 인식률을 보이고 있어 지속적인 보완과 수정이 요구되고 있다.

현재까지 음성인식의 장점을 충분히 누리지 못하는 사용자들이 많지만 음성인식의 기술이 널리 알려지기 시작하면서, 또한 사용자들도 이러한 서비스에 익숙해지면서 음성인식을 사용하는 빈도는 앞으로 계속 높아질 것으로 보인다.

자동응답 서비스의 음성인식 기술의 적용은 단순한 사용의 편의를 넘어 정보통신 분야의 산업을 더욱 발달시킬 것으로 보이며, 앞으로는 거의 모든 전화정보 서비스에 음성인식 기술이 사용될 것으로 전망된다.

5. 결 론

현대에는 매우 빠른 속도로 새로운 기술이 개발되고 있으며, 이로 인해 기계, 전자산업의 급격한 발달이 이루어지고 있고 매일매일 많은 정보가 쏟아져 나오고 있다. 급격한 과학의 성장과 수많은 정보의 바다 속에서 생활은 점점 편리해져가지만 그만큼 많은 기계와 시스템들을 접하고 살고 있는 것이다. 따라서 기계나 시스템들과 인간의 통신도 그만큼 많이 필요로 하며 중요한 분야로 자리를 잡았다. 복잡한 기계의 사용을 쉽게, 편리하며 빠르게 사용할 수 있는 방법 중 하나로 음성인식을 사용할 수 있다.

요즘 많은 사람들의 관심을 모으고 있는 증권정보 서비스에 음성인식의 기술을 사용하여 사용

의 편의를 도모하였다. 물론 부족함이 있으나 이 기술의 사용으로 전화의 가장 큰 장점인 음성통신을 시스템과 사람이 할 수 있게 되었다.

정보통신 산업은 이 기술의 사용으로 더욱 성장할 수 있으며, 음성인식의 기술은 앞으로 반드시 필요로 하는 기술이 될 것이다.



감 지 은

1996년 경원대학교 물리학과 (이학사)
1998년 한양대학교 물리학과 (음성인식 전공, 이학석사)
1998년-현재 범일정보통신(주) 음성기술연구소 연구원

관심분야 : 음성인식, 음성신호처리

참고문헌

- [1] L. Rabiner, R. W. Schafer, Digital Processing of Speech Signal, Prentice-Hall International, Inc.
- [2] 안영목, "차세대 UI 음성인식 기술", 마이크로소프트 1996.8
- [3] L. Rabiner, B. H. Juang, Fundamentals of speech recognition, Prentice-Hall International, Inc.



이 문 형

1978년 한국증권전산 입사
1978년-1993년 증권시스템부, 정보시스템부 개발부 근무
1994년-현재 한국증권전산 네트워크 사업부 부장

소프트웨어공학연구회지 제2권 제3호(통권5호) 논문 모집 (1999년 9월 발간 예정)

- 주 제 : 소프트웨어 신뢰도와 테스트
- 제 출 처 : 정혜정 교수(평택대학교)
- 논문마감 : 1999. 8. 14(토)
- 연 락 처 : jhjung@ptuniv.ac.kr

Tel:0333-655-3225

Fax:0333-658-3130