

KT 증권정보 서비스 이용 실태 및 인식 결과 조사

최영재, 김재인, 구명완

한국통신 연구개발본부 멀티미디어연구소 음성언어연구소

A Research on the state of the utilization of the stock-information-retrieval-service

Young-Jae Choi, Jae-In Kim, Myoung-Wan Koo

Spoken Language Research Division, Multimedia Technology Research Laboratory, R&D Group, Korea Telecom

Email : {yjchoi, jikim, mwkoo}@smm.kotel.co.kr

요 약

한국통신에서는 PC로 된 프로토타입 시스템을 이용하여 음성인식 증권정보 서비스를 1995년 11월부터 1998년 초까지 5채널에 대해 시험운용을 해왔으며, 상용서비스를 위해 120명이 동시에 서비스 받을 수 있는 시스템을 개발하였다. 개발된 시스템의 전반적인 문제점을 파악하기 위하여 개발된 시스템을 사용하여 1998년 3월 16일부터 30채널규모로 일반인들에게 시험서비스를 제공하고 있다. 음성인식 전화정보 서비스를 현재보다 훨씬 더 활성화시키기 위해서, 서비스의 이용 행태에 대한 분석을 통해, 어느 부분이 어떻게 개선되어야 할지를 연구하여, 초보 사용자라도 이용하기 쉬운 형태로 서비스의 시나리오를 개선해 나가고 있다.

본 논문에서는 사용자 특히, 처음 사용자의 여러 가지 이용 실패 요인을 분석하였다. 또한, 음성인식 증권 정보 서비스가 정식으로 서비스되기 이전과 그 이후의 일시별 인식률을 통계 조사하고, 이용자가 동일 대상 단어를 연속으로 발음하는 경우, 동일 대상 단어에 대한 인식률을 조사하였다. 조사결과 문제점은 4가지로 분류될 수 있었으며, 드러난 문제점을 해결하기 위하여 노력하고 있다.

1. 서 론

음성인식기술에 대한 연구는 오래전부터 수행되어 왔으며, 이제 우리생활의 주변에서 관련기술을 이용한 제품이나 서비스들을 보거나 이용해 볼 수 있게 되었다. 외국에서는 이미 전화망에서 AABS⁽¹⁾(Nortel)나 VRCP(AT&T), 음성다이얼링서비스(AT&T, Nynex, Bell Atlantic)와 같은 다양한 서비스가 제공되고 있고, 요사이는 Personal Computer(이하 PC)를 이용하여 IBM의 "Via Voice", MicroSoft(이하 MS)의 "Whisper" 그리고 Dragon의 "Dragon Voice Dictator"등의 음성타이프라이터나

명령어를 키보드나 마우스가 아닌 음성으로 입력할 수 있는 "음성명령기"등이 선을 보이고 있다. 최근에는 MS가 자사의 차세대 OS에 음성인식기능을 기본으로 채용하려고 준비하고 있으며, 벌써 일반 S/W에서 음성인식기능을 이용할 수 있도록 하는 SDK(Software Development Kit) 베타버전을 소개하고 있다. 한편 국내에서는 한국통신이 1995년 11월에 소규모이기는 하지만 전화망 통한 음성인식증권정보서비스를 시험적으로 제공하였으며, 1997년초 음성다이얼 서비스를 SK Telecom이나 신세기통신 같은 무선이동통신 사업자가 국내최초로 음성다이얼 서비스를 상용서비스로 제공하였다. 이후 음성인식기능이 내장된 무선통신단말기가 공급되기 시작하여 국내의 일반인들도 음성인식기술에 대한 편리성을 몸으로 느끼기 시작하고 있다. 한편 PC에 대해서는 "음성명령어"나 "음성타이프라이터"등에 적용하려 하고 있지만 우리의 음성인식기술을 이용한 제품의 개발은 아직 시작 단계에 있다고 하겠다.

한국통신에서 PC를 이용하여 개발된 프로토타입 시스템으로 음성인식 증권정보 서비스를 1995년 11월부터 1998년 초까지 5채널에 대해 국내 최초로 시험운용을 하면서, 상용서비스를 할 수 있는 시스템을 개발하기 위하여 노력하였다. 그 결과 97년 말 120명이 동시에 음성인식을 이용한 서비스를 제공할 수 있는 시스템을 개발하였다. 이를 이용하여 우선 1998년 3월 16일부터 30채널 규모로 일반인을 대상으로 시험서비스를 시작하였다. 이 논문은 시험서비스를 하면서 사용자와 시스템간에 당화에 대한 분석 및 경토에 관한 것이다.

음성인식서비스가 현재보다 훨씬 활성화되기 위해서는, 우선 서비스의 이용 행태에 대한 분석을 통해, 어느 부분이 어떻게 개선되어야 할지를 연구하여, 사용자가 이용하기 쉬운 시스템으로 개선해 나가야 한다.

본 논문의 구성은 다음과 같다. 제 2장은 고객 만족도 제고를 위해 사용자 이용 실태 분석에 대해

기술하였다. 여기에서는 처음 사용자의 여러 가지 이용 실패 요인을 살펴보고, 담화시스템 개량을 위해 노력해야 할 부분을 간략하게 고찰하였다.. 3장에서는, 음성인식 증권 정보 서비스가 정식으로 서비스되기 이전과 그 이후의 일시별 인식률에 대한 통계 조사, 각 일시별로, 음성인식 증권 정보 서비스를 이용한 사람의 수 등을 조사하였다. 이외에도, 여러가지 운용 결과를 정리하였다. 4장에서는 앞으로 해야 할 일에 대해 간략하게 기술한다.

2. 서비스 이용 실태 분석

새로 개발된 시스템을 이용하여 서비스를 시작한 것은 98년 1월 중순 부터였다. 이때부터는 PC로 된 프로토타입 시스템과는 전화번호가 다르지만, 서비스는 거의 동일한 서비스가 PC로 된 프로토타입과 병행하여 서비스 제공되었다. 그리고 프로토타입 시스템 이용자에게는 새로운 전화번호를 안내해 주어, 이전 이용자들이 새로운 시스템으로 자연스럽게 이동할 수 있도록 하였다. 서비스 이용 실태 분석을 위하여 1998년 3월 13일부터 3월 16일간 약3일반(도요일, 일요일제외) 정도의 기간에 표본적으로 약4시간 분량의 이용 상황을 모니터링하였다. 신문에 서비스관련 기사가 나오기 전까지는 이전장비를 사용하던 사람들이 새 서비스를 사용하였기 때문에 사용범위속에 따른 실수가 거의 없었다. 하지만 언론에 전화번호가 알려진 직후의 인식률은 사용범위속에 따른 실수가 증가되었다.

서비스 이용에 실패한 경우를 살펴보면, 크게 두 가지로 구분된다. 하나는 제대로 음성이 입력되었음에도 음성인식이 잘못되어 틀린 인식 결과를 내는 경우이며, 다른 하나는 이용자가 음성을 잘못 입력하여 인식이 잘못된 경우이다. 조사 결과, 음성이 제대로 입력되었는데 인식 결과가 틀린 경우는 그다지 많지 않았으며, 대부분이 이용자가 제대로 입력을 잘못하여 인식이 잘못되었다. 특히, 처음 사용하는 이용자의 경우, 사용법을 잘 몰라, 서비스가 실패하는 경우가 많았다. 이용자가 음성입력을 잘못된 경우는 다음과 같이 9가지로 구분할 수 있었다.

- (1) 입력 음성의 레벨이 낮아, 시스템이 이를 음성 입력으로 받아들이지 못한 경우(이용자가 작은 목소리로 음성 입력한 경우와, 전화기 송화기를 입에서 멀리하고 말한 경우)
- (2) 입력 음성 전 후에, 음성이 아닌 다른 소리가 끼어 입력된 경우(목청 가다듬는 소리, 주변의 프린터 소리, 송소리, 전화기 달가닥 소리, DTMF

신호, 기침 소리, 감탄하는 소리, 입바람 소리)

(3) 단어를 제대로 말하였으나, 발음이 애매모호한 경우(허스키한 음성, 입바람소리가 많이 섞인 음성, 유성음이 전혀 섞이지 않게 말한 음성, 사투리 음성)

(4) 인식 대상 단어명은 맞으나, 음성입력시 앞으로 인식 대상 단어의 일부 또는 전체를 반복하여 말한 경우

(5) 인식 대상 단어명을 몰라, 비슷한 단어를 음성입력한 경우

(6) 인식 과정을 알리는 음악소리 도중에 음성을 입력하여 입력된 음성이 어떤 것인지 잘 모르게 된 경우

(7) 전혀 인식 대상 단어가 아닌 음성이 입력된 경우(이용자가 전화기를 들은 채 다른 사람과 대화할 때 입력된 음성, 또는 원하는 정보가 아닌 정보가 나왔을 때 화가나서 한 말이 음성이 입력된 경우)

(8) 2개 이상의 인식명칭을 한번에 입력한 경우

(9) 인식 대상단어 다음에 '부탁합니다'를 붙여 입력한 경우

현재 서비스중인 시스템은 화자독립 단어인식시스템이기 때문에 위와같은 대부분의 문제를 시스템측면에서 해결할 방법은 없기 때문에 사용자들에게 서비스사용법을 어떻게하면 쉽고 빠르게 알려줄 수 있을 것인가에 주안점을 두어 서비스 시나리오를 수정하였다.

3. 인식을 분석

음성인식서비스를 하는데서 어려운 점중에 하나는 고른 인식률을 가진 시스템을 만드는 것이라 생각된다. 컴퓨터 시뮬레이션으로 시스템에 대략적인 인식률은 나오지만, 서비스가 시작되면 이것도 별 도움이 되지 않는다. 왜냐하면 사용자는 그러한 인식률은 모르고 있으며 본인이 사용해서 잘 동작되면, 인식률이 좋다고 느낄 것이며, 그렇지 않으면 아주 나쁘다고 생각하게 된다. 어떤 사람이 어떤 단어가 인식이 잘 안 된다고 하여 시험해 보면 시험한 당사자에게는 문제가 없는 경우가 거의 대부분이기 때문이다. 그래서 인식이 잘 안 되는 단어에 대한 것은 서비스를 하면서 찾아 원인을 분석하여 개선하는 단계가 필요하다. 연구실에서는 1998년 2월 23일부터 3월 39일까지 8차례에 걸쳐, 서비스의 인식률을 조사하였다. 조사결과 총 입력횟수는 1,240회이었으며, 총 입력 대상단어는 783개이지만, 한번이라도 입력된 단어는 304단어만이어서 인식대상 단어별로 인식을 통계를 내기에는 데이터양이 부족하여 추가적인 데이터를 더 모으고 있다.

조사에서는 입력이 5회 이상인 인식 대상 단어에 대해서만, 인식 대상 단어별 인식률 통계를 내었다. 인식률이 낮은 인식 대상명에 대해서는 컴퓨터 시뮬레이션 결과와 비교하여 차이가 많이 난다면 그 이유를 분석하여 시스템을 보완해야 된다.

5회 미만의 인식 대상 단어라도 3회 또는 4회의 입력 동안에 아무런 오인식이 없는 경우 해당 인식 대상 단어의 인식률은 안정적이라고 생각 할 수 있다.

한번도 인식이 성공하지 못한 인식대상명의 경우에 대한 분석은 중요하다. 초보 사용자가 첫 번째 입력이 성공되지 않는다면, 사용자가 계속해서 입력 실패할 가능성이 높고, 특히, 처음 이용자의 경우에는 이 시스템에 적응 못하고, 추후 이용할 마음이 생기지 않게 하는 요인이 될 수가 있다.

입력 횟수가 3회 이상인 경우에 1회도 인식하지 못한 단어에 대해서도 원인을 분석하였다. 원인은 주로 발음잘못이 가장 많았고, 다음으로 목소리가 너무 굵거나 허스키한 경우와 같이 특이한 목소리입력인 경우 인식이 잘 되지 않았다.

일시별 음성인식률 변화는 다음과 같다. 본 서비스가 신문에 공표하기 이전인 98년 2월23일경의 인식률은 88.30%로 비교적 서비스가 안정되었다고 판단하였다. 그러나, 서비스를 처음으로 이용하는 사용자가 생기기 시작한 98년 3월13일부터 3월 16일까지의 통계 결과를 보면, 인식률은 70.92%까지로 떨어졌다.

입력 실패의 원인을 유형별로 분류하여, 통계 조사한 결과, 사용자의 음성 입력 레벨이 낮아 무입력 처리되는 경우와 인식 대상단어를 사용자가 모르는 경우가 71%를 차지하였다. 그 이외에 주변 잡음, 송소리, 탈가닥 소리, 기침소리 등이 실패의 원인이었다.

서비스 사용자는 자신이 말한 단어가 인식 안 된 경우, 그 단어를 바로 다시 말을 한다. 이와 같이 동일 대상 단어를 연속으로 발음하는 경우 대한 누적 인식률을 조사하였다. 이러한 인식률은 사용자 측면에서, 서비스에 대한 최소한의 만족도를 나타낸다. 조사 결과, 2회 연속 발음 시까지의 누적 인식률은 90%를 훨씬 상회하는 것으로 나타났다.

4. 추진계획

음성인식기술을 이용하여 증권정보안내 서비스를 하면서 문제가 되는 것은 크게 네가지로 나누어진다. 먼저 사용자의 음성이 너무 작아서 입력이 되지 않는 경우이다. 실제로 입력실패의 원인중 상당부분이 이것에 의한 것이었다. 주로 여자 사용자에게서 이러한 것을 자주 볼 수 있었다.

그래서 원래 선천적으로 목소리가 작은 사람들을 위해 끝점검출기의 성능을 개선할 필요가 있다고 생각된다. 두번째로 증권정보와 같이 한번에 여러 번의 입력을 하는 경우 사용법에 대한 빠르고도 충분한 설명을 들려줄 수 있는 시나리오 구성이 중요하다. 자주 사용하는 사람들도 만족스럽게 사용하지 못하고 있는 것이 자주 발견되며, 처음의 이러한 잘못된 사용 방식을 그대로 고수하는 사람들도 많이 발견된다. 예를들어 “한국통신”을 “한 국 통 신”과 같이 띄어 읽는 다든지 입력단어를 한번이 아닌 여러 번 반복해서 입력한다든지 목소리를 아주 크게 입력하는 경우가 있다. 이러한 것은 사용법 안내를 한번만 들어보면 해결될 수 있지만 강제로 들어 보게 할 수 있는 방법이 없는 것 같다. 실험실에서는 시나리오의 변경을 통해 해결하려고 하고 있다. 세번째로 입력단어 앞뒤에 잡음이 들어가는 경우는 keyword spotting 기술이 적용된다면 어느정도 해결될 것으로 생각된다. 하지만 현재는 rejection기능을 구현하여 우선 엉뚱한 결과가 방송되지 않도록 적용시험중에 있다. 네번째로 입력단어에 대한 기준 설정이 문제이다. 회사 이름이 4자이상어 되면 축약해서 입력하는 경우가 많은데 이를 무시하고 입력단어를 확장형으로 고수할 수도 없는 경우이다. 연구실에서는 이문제점과 발음상의 변화를 같이 수용하여 인식을 할 수 있는 방법을 연구중에 있다.

앞으로 11월 말까지 시험서비스를 계속하면서 나타난 기술적인 문제를 해결하기 위하여 새로운 알고리즘을 적용하여 시험할 예정이며, 아울러 시스템운영측면과 시스템의 안정성 검증 및 서비스 확산에 가장 큰 걸림돌이 되고 있는 시스템 가격을 내리기 위한 H/W의 최적화 연구를 계속할 예정이다.

참 고 문 헌

1. W. E. Logenbaker, "Automation of operator services: A successful application of speech recognition technology," Proc. IVTTA94, pp161-164, Sep., 1994.
2. M. W. Koo et al., "KT-STOCK : A speaker-independent, large-vocabulary speech recognition system over the telephone," in Proc. 1994 Int. Conf. on Spoken Lang. Processing, pp. 1387-1390, Sep., 1994.
3. M. W. Koo et al., "An experimental field trial of a large vocabulary, speaker independent recognition system," in Proc. Second IEEE Workshop on Interactive Voice Technology for telecomm. Applications, pp. 33-36, Sep., 1994.

4. M. W. Koo et al., "A STOCK INFORMATION SYSTEM OVER THE TELEPHONE NETWORK," in Proc. The International Conference on Signal Processing Applications & Technology, Vol.2, pp. 2039~2043, 1995
 5. S.J.Park, J.I.Kim, "The architecture of a speech recognition system for cost reduction," in ICSP'97(International Conference on Speech Processing), 1997
 6. J.I.Kim, H.K.Kim, "The considerations of name recognition," in ICSP'97(International Conference on Speech Processing), 1997
 7. S.K.Park, J.I.Kim, "Development of multichannel information retrieval system using speech recognition," in ICSP'97(International Conference on Speech Processing), 1997
-