

□ 기술애설 □

전화망에서의 음성인식기술의 활용현황

한국통신 김재인

1. 서 론

전화는 가장 편리하고 값싼 단말기라고 할 수 있으며, 오늘날에는 전화를 사용하여 많은 유용한 정보를 얻을 수 있다. 그러나 처음에는 음성정보서비스와 같이 기계에서 사용자가 원하는 정보만을 얻기 위해서는 전화기에 달려있는 전화버튼을 이용하여 그 내용을 선택하여야만 해서 불편하였으며 이 역시 전자식 교환기에 연결되거나, 휴대용단말기(휴대폰, PCS)를 가진 경우만 가능하였다. 이러한 불편은 음성인식기능을 가진 전화정보시스템이 개발되어 어느 정도 해결되어가고 있으며, 좀 더 다양한 서비스들이 개발되어 사용자들을 편리하게 하고 있다. 본 논문에서는 전화망을 통해서 음성인식기능을 이용하여 사용할 수 있는 서비스에 관하여 알아보기로 한다.

2. 외국의 서비스 현황

음성인식에 대한 연구는 1952년 미국의 벨 연구소에서 숫자음인식에 대한 논문이 발표된 이후 이제까지 여러 나라에서 활발히 진행되어 오고 있다. 이 기술이 전화망에 사용된 것은 1986년 영국에서 cell-phone에 적용한 이래 1990년에 들어서면서부터 서비스 개발이 활발해져서 현재까지 세계각국에서 시험서비스 또는 상용서비스를 개발하거나 또는 제공하고 있다. 여기서는 유럽과 미주 그리고 일본에 서비스 현황에 대하여 차례로 알아보고 마지막으로 우리 나라에 대하여 설명하겠다.

2.1 유럽

유럽은 유럽통합에 따라 동일한 서비스를 동시에 여러 언어로 제공해야 하기 때문에 하나의 공통된 서비스에 대하여 몇개 나라가 공동으로 개발하면서 각자의 언어에 대한 문제를 해결해주는 연구들이 활발히 진행되고 있다.

2.1.1 프랑스

1992년 이후 몇가지 시험서비스가 개발되었지만, 대부분 CNET의 automatic speech recognition(ASR) PHIL90의 기술을 사용한 것으로, 이 기술은 화자독립, 단어모델을 사용하고, 안내방송 중에 음성입력이 가능한 barge-in, 인식대상단어가 아닌 경우에도 틀린 인식결과를 내는 것을 방지하는 rejection, 그리고 입력음성 중에서 인식대상단어만 찾아내는 핵심어 추출기능을 지원하며, 화자종속과 allophone model도 지원한다. 현재 여러 분야에 상용서비스가 제공되고 있다[1].

가. FT Public Voice-Mail Service

Yes, no, listen, record, delete, add와 announce 등 7단어를 인식하며 incoming call만을 처리한다. Transvox에 ALCATEL/TITN platform에서 CNET 구현되었으며, 3개 도시(240회선)에 지역주민만 서비스 가능하고, 1997년안에 전국적인 서비스를 할 예정이다.

나. Information Service for People on the Move

유럽의 5개국 교환원(FT, BT, DT, TI, Telefonica)이 협동으로 지원하는 다국어 서비스이며 5개국어의 4단어의 음성명령이 가능하다.

[가]의 경우와 같은 인식기와 H/W를 사용하고 있으나, 45port로 서비스하고 있다. 1995년 7월에 서비스가 개시된 이래 1995년 여름에서 가을까지 매월 25,000 call을 처리했다.

다. France Telecom(FT) Sales Agency Service

FT의 제품 및 서비스들을 근무시간 이외에 제공하는 것으로, 1991년부터 4port 정도의 소규모로 시험서비스 중이다. 인식단어는 23개이며 하루에 120통화를 처리한다.

다음은 FT가 아닌 다른 사업자가 제공하고 있는 서비스이다.

라. "Les Baladins" Service

1986년부터 Lannion근교에서 cinema 프로그램들에 대한 정보를 알려주고 있다. 1port로 일주일에 400통화를 처리한다.

마. "MACIF" Service

MACIF는 프랑스에서 가장 큰 자동차보험 회사로 고객에게 24시간 내내 상담이나 보험처리결과를 알려주고 있다. 인식단어는 30개("insurance," "vehicle," "car registration number," "robbery," etc)로 PC로 된 server는 동시에 100통화를 처리할 수 있으며, 하루에 수백 통화를 처리하고 있다.

다음은 FT가 현재 field trial 또는 lab testing단계에 있는 서비스들이다.

바. Voice Dialing Service

사용자 한사람당 10~30명의 이름을 등록하여 사용할 수 있다.

사. Automatic Directory Service

CNET사의 Lannion 고용자들의 성명을 인식하며, 인식이 어려운 경우 spelling을 사용한다. 1995년 7월부터 시험서비스 중이다.

아. CINEZOOM

Video on Demand에 대한 예약서비스를 제공한다.

자. Rail Travel Information Service

European RailTel project에 일부로 개발되어 시험서비스를 제공중에 있다.

2.1.2 영국

영국의 BT는 10년이상 음성처리기술을 전화망에 적용해 왔으며, 그 결과 다양한 서비스를

제공하고 있다. 1986년 세계 최초로 handfree speech dialing cellphone인 "Topaz"를 개발하였으며, 1988년에 "Topaz II"와 1991년 "Azure"로 개량되었다. BT에서는 두 종류의 main platform을 사용하여 모든 서비스를 개발하고 있다. 그중에 대부분은 BT에서 설계하고 현재 Ericsson에서 만들어내고 있는 interactive Speech Application Platform(ISAP) [2]이 있으며, MAP(Minor Application Platform)은 PC based platform으로 main network이외에서 사용될 서비스를 위해 이용될 수 있다.

가. CallMinder

영국 내에서 음성인식기술을 이용한 서비스 중 가장 큰 것으로 사용자가 전화선을 사용하고 있을 때, 대신 응답을 해준다. 음성인식을 이용하여 사용자가 메시지를 검색하거나, 시스템을 사용자가 원하는 대로 변환할 수 있다. 예를 들어 CallMinder가 받기까지의 ring의 횟수를 조정할 수 있다. 1995년 6월부터 서비스되기 시작하여 전국적으로 19개 site에서 50만 이상의 사용자를 확보하고 있다.

나. Automatic Directory Enquiries

1994년 East Suffolk에서 25,000명의 이름에 대해 대어휘인식과 spelt 입력이 가능한 directory enquiry service를 처음 제공한 이래 현재에도 몇개의 시스템이 시험서비스 중에 있다. BT의 한 연구소에서도 4,000명의 근무자의 전화번호를 사람이름으로 찾을 수 있는 서비스를 제공하고 있다.

다. Payment Line

이 서비스는 home shopping이나 고객관리를 위해서 개발되었으며, 개인번호, 이름 주소를 인식하기 위해 대어휘 연속음성인식기술과 확인을 위해 음성합성기술을 사용하였다. 이는 몇 개의 우편주문회사에서 시험 중에 있다.

라. Medical Line

의사들이 원격지에 있는 데이터베이스 환자들의 진료기록을 찾아볼 수 있도록 하는 것으로 음성합성기를 같이 사용하고 있다.

2.1.3 이탈리아

지난 몇 년 사이에 음성인식과 합성기술을

이용한 서비스 개발이 활발히 진행되어 왔다. 그 결과 여러 서비스들 중 RaiITel은 Euro-speech '97에서 실시된 음성인식시스템 평가회에서 10여개의 시스템 중에서 1등을 차지하기도 하였다.

가. 1412Reverse Directory Service

전화번호를 입력하면 가입자의 이름과 주소를 알려주는 것으로 1994년 처음에는 pulse dialing이나 DTMF로 번호를 입력하였으나 1996년부터 음성입력과 (yes/no)의 입력이 가능한 시험서비스가 실시되고 있다.

나. Automatic Alternate Billing Services

시의전화나 국제전화의 방법을 선택하는 것으로 연속숫자음과 (yes/no)를 인식할 수 있으며 1996년부터 부분적으로 또는 전부 자동화하였다.

다. Customer Care Service

이 서비스는 12만 Telecom Italia 직원에게 생산품의 기술적인 정보를 전달해 준다. 입력은 DTMF나 음성입력이 가능하며, 8자리의 개인번호와 문제의 종류를 5개의 범위중에서 선택하여 입력한다. 1995년 시작하여 내부적으로 운영중에 있다.

1996년 말부터는 연속 숫자음 인식과 보다 자세한 정보를 입력할 수 있는 인식기를 사용하고 있다.

라. NOMINA-Voice Dialing on PABX

이 서비스는 원하는 사람의 이름을 말하면 자동적으로 그 번호를 dialing해준다. 물론 인식결과를 TTS를 사용하여 들려준다. 회사전화 번호부에 의해 인식어휘가 자동적으로 생성되며, PC-Dialogic Antares platform을 이용하여 개발되었다. 이는 1994년부터 시작되어 몇 군데서 시험운용중에 있으며, 약 300명에서 1,200명의 이름을 인식할 수 있다.

마. VAD-Voice Activated Dialing

이는 개인적으로 20명까지의 이름과 전화번호를 등록할 수 있으며, 일반전화망에서 이름만으로 dialing을 가능하게 해준다. 화자종속인식기를 사용하였으며 1996년 전반기에 108명의 사용자를 대상으로 시험을 하였다.

바. DEMOS-Automatic Telephone Interview System

이 서비스는 교환원없이 DTMF 또는 ASR을 이용하여 면접을 하는 것이다. 현재 1996년 말부터 시험서비스를 개시하였다.

사. Gas Meter Reading by Phone

전화선을 통해서 고객 코드(10~11자리)에 대한 음성입력과 gas meter를 읽는다. 두 개의 gas회사에서 서비스 중이다.

아. Fax Yellow Pages

원하는 광고주에 대한 새로운 정보를 fax로 전송해주는 서비스로 광고주 전화번호와 fax 전화번호에 대한 숫자입력을 DTMF나 ASR을 이용하여 받는다. 1994년부터 시험을 시작하였다.

자. INFORMACITY Service

Venice에 대한 설명을 해주며, 60회선 규모로 1995년부터 운영해 오고 있다.

차. RAILTEL-Italian Train Timetable Service[3]

이탈리아 주요 도시간에 철도연결에 대한 정보를 제공한다. 500개의 도시명과 날짜, 숫자 등 고립단어를 인식하며, 4회선규모의 PC-Dialogic Antares platform을 사용하여 개발되었다. 1995년 9월부터 운영하기 시작했다.

타. DIALOGOS-Italian Train Timetable Service

차와는 달리 자연스런 언어를 사용할 수 있게 개발되었다. 1990년 첫 시스템이 개발되었으며 3,500단어를 인식할 수 있다. DEC workstation에서 Dialogic telephone board를 사용하여 한 개의 채널을 수용할 수 있으며 TTS와 연속음성인식, 그리고 담화기술들이 적용되었다. 1995년말 500명의 사용자를 대상으로 서비스되고 있다.

2.1.4 독일

Deutsche Telekom에서는 ASR 기술을 이용하여 다음과 같은 서비스를 제공하고 있다.

가. SCALL-a new paging service

이 서비스는 1995년에 시작되었으며, 사용자는 숫자정보를 특수한 호출기로 보낼 수 있다. 숫자음과 제어를 위한 명령어를 인식할 수 있다.

나. T-Card-an international virtual calling card service

이는 걸려오는 전화를 ASR기능을 이용하여 제어 할 수 있는 음성응답시스템이다. 사용자의 신분확인을 위하여 calling card number와 PIN code를 입력하여야 한다.

다. FAUST-a directory assistance demonstrator

현재 5,000명의 가입자와 25개 도시에 대한 전화번호안내를 해 준다. 이 시스템은 핵심어 추출기능, alphabet 인식과 음성합성기술을 사용하고 있다. 1996년부터 시험중에 있으며, 실제의 전화번호안내서비스에서 40개의 도시이름을 인식하는 반자동 시스템이 사용될 예정이다.

라. Tarifinfo

Deutsche Telekom의 통화요금에 대한 정보를 제공한다. DTMF입력과 정해진 어휘를 인식하며, 교환원과 연결가능하다.

2.1.5 스페인

현재 몇가지 서비스가 시험운용중이며, 1997년에 상업적인 서비스를 개시할 예정이다. 이미 서비스중인 것중에는 "CAMPIN" 시스템이 있는데, 사용자가 Telefonica의 calling card의 PIN을 변경할 수 있게 해 준다. 현재는 숫자음 인식을 하지만, 자연스러운 연속숫자음 인식기능을 추가할 예정이다. 그 외에는 voice dialing, answering machine, automatic directory assistance 등 앞에서 설명한 국가들이 제공하고 있는 서비스들에 대하여 다각적인 시도를 하고 있는 것으로 보인다.

2.2 미국

AT&T는 주로 전화교환원의 일을 자동화하는데 음성처리기술을 적용하고 있다.

가. VRCP-Voice Recognition Call Processing

1992년 다섯 단어{collect, calling card, third number, person, operator}를 인식하는 시스템으로 시작하여 현재 이 다섯 단어에 대한 spotting과 연속단어인식기술이 추가되어 미국 48개 주에서 사용되고 있으며, 일년에 수십억 call을 처리하고 있다. 인식률을 90%이다.

나. Universal Card 24Hr Customer Services(1-800-423-4343)

카드소유자에 대한 신용정보를 조회하는데 연속숫자음 인식기술을 이용하여 카드번호, 비밀번호, 서비스선택 등을 할 수 있다. 현대 한 달에 400만 call이상을 처리하며 97%이상의 인식률을 보이고 있으며, call의 57%를 자동으로 처리하고 있다. 숫자입력방식에서 DTMF보다 음성입력을 더 선호하고 있는 것으로 나타났다(45%@55%).

다. VoiceDialing[5]

NYNEX에서 1992년 12월 처음 개발된 이후 93년 중반부터 실제 서비스에 들어갔으며, 94년 New York와 New England에서만 1만5천 가입자를 확보하였다. 이 서비스는 30, 50, 70명까지 개인의 전화번호 및 인명을 등록하여 사용할 수 있다. On line help기능을 가지고 있으며, DTW를 사용한 화자중속음성인식방법을 사용하고 있다. 1995년부터는 DTW에서 continuous density multi-gaussian mixture HMM방식으로 변환을 시작했으며, 화자독립 명령어와 화자중속이름인식기를 동시에 사용할 수 있도록 개선하고 있다.

2.3 일본

가. ANSER[4]

NTT가 1981년 개발한 시스템으로 은행업무에 대한 정보를 제공한다. 초기에는 voice response기능과 DTMF 수신기능만이 있었고 후에 숫자인식기능이 결합되었다. 1990년까지 일본내 601개의 은행에서 사용하고 있다.

나. Voice Dialing

KDD에서 개발하여 1995년 11월부터 사용하고 있는데 PBX에 환경에서 이름을 말하면 전화를 걸어주며, 5,000명의 이름을 인식한다.

다. Operator Assistance

국제전화에 대한 사용자의 문의사항에 대답하므로써 KDD의 교환원을 도와주기도 하며, 다른 종류로는 핵심어의 존재여부를 조사하여 외국인의 국제장난전화를 차단하는 역할을 하고 있는 것도 있다.

라. 날씨정보 안내

일본내 각 지역 명이나 휴양소를 말하면 그

곳의 날씨를 알려준다.

마. 새소리 안내

일본 내 야생조류(189종류)의 소리를 들려 주며, 1995년 8월부터 서비스를 시작하였다.

바. Home Banking

원하는 은행에 돈을 송금할 수 있게 해주며, 계좌번호, 암호, 이체금액 등을 인식하며 아직 시험중에 있다.

2.4 우리나라

국내에서의 음성인식연구는 KT, ETRI, KA-IST와 일부 대학연구소에서 지속적으로 연구되고 있다. 전화망에 관련기술을 적용한 경우는 사람이름이나 부서명을 인식하여 전화번호를 안내해 주는 시스템들이 시험적으로 몇 군데서 개발된 적이 있으나, 아직 일반인들을 대상으로 상용서비스를 제공하고 있지 못하고 있다. 다행히 한국통신(KT)에서는 음성인식증권정보서비스, 음성다이얼링서비스 등과 관련 시스템들을 개발하였고, 현재 시험서비스중에 있으며, 또한 여기에 적용된 기술들을 국내시장을 활성화시키기 위하여 업체로 이전하였고, 관련서비스를 개발할 수 있도록 도와준 결과 LG정보통신, 삼성전자, 삼보정보통신 등에서 음성인식전화정보시스템을 개발하여 증권 정보서비스를 구현하였으며, 이를 바탕으로 다양한 서비스를 구상하고 있다. 그리고 컴퓨터 회사에서는 PC용 음성인식기로 개량하여 응용제품을 개발할 예정이다.

가. 증권정보서비스[6]

1995년 11월 9일부터 PC-base로 된 시스템을 2회선 규모로 시험서비스를 시작하였으며, 대용량 회선을 수용할 수 있는 VME-base 시스템을 개발하여 이것으로 시험서비스를 준비중에 있다. 이 시스템은 화자독립, 어휘독립기술이 적용되어 인식단어를 추가 및 변경이 자유로우며, barge-in과 echo cancellation 기능이 있다. 약 300개의 CDP를 사용하여 1,000개의 회사명을 실시간으로 인식할 수 있다.

또한 PCS회사 중에서는 한국통신의 음성인식기술을 이용하여 개발된 시스템을 삼성전자로부터 도입하여 증권정보서비스를 1997년 12

월 10일부터 제공하고 있으나, 통신망이 불안하여 입력된 말의 일부가 잘리는 경우가 발생하여 인식률이 낮은 형편이며 망이 안정되면 이 문제도 해결되리라 기대하고 있다.

나. Voice Dialing Service

화자중속기술을 적용하여 휴대통신사업자인 SK-telecom과 신세기이동통신등에서 외국의 기술과 장비를 이용하여 1997년 초에 상용서비스중에 개시하였다. 한국통신에서는 화자독립기술을 이용하여 1997년 2월에서 10월말까지 서울지역에서 시험서비스를 실시하였으며 150개의 정해진 단어중에서 선택하여 전화번호를 등록, 사용할 수 있으며, 위치독립과 위치중속서비스가 가능하다. 이 서비스 역시 화자독립인식기술이 적용되었다.

다. 이름인식 부서안내서비스

1997년말에 개발된 것으로 lab-test를 중에 있다. 1000명정도의 이름을 인식할 수 있지만 이름의 특성상 인식률이 다른 서비스보다 낮은 편이어서 연구소 내에서 시험서비스를 하면서 개선시킬 예정이다. 이 서비스는 사람이름을 말하면 해당사람의 전화번호와 부서를 안내해 주며, 사설교환기에 연결하면 전화도 걸어줄 수 있다.

3. 결 론

본 논문에서는 음성 인식기술을 전화망에 적용한 서비스에 대하여 살펴보았다. 전반적으로 인식단어수는 많지 않으면서 90%이상의 높은 인식률을 가지고 있었으며, 숫자음 인식기술을 많이 적용되었다. 또한 사용자들이 보다 편리하게 사용할 수 있게 barge-in이나 핵심어 추출기능 등에 기술이 적용되고 있었다. 우리나라도 1997년 초에 비록 국내기술을 이용한 것은 아니지만 음성다이얼링서비스가 시작되었으며, 한국통신이 개발한 두 가지 정도의 서비스가 시험서비스중에 있다. 1998년에는 한국통신으로부터 이전 받은 음성인식기술을 바탕으로 관련시장에 서비스개발이 더욱 활성화될 전망이다.

참고문헌

[1] C. Sorin et al. : "Operational and experimental French telecommunication services using CNET speech recognition and text-to-speech synthesis", *Speech Communication*, Vol 17, pp. 275~286, November 1995.

[2] F.A. Westall, R.D. Johnston, A.V. Lewis, (Editors) : "Speech Technology for Telecommunications", *British Telecom Technical Journal*, Vol 14, No 1, January 1996.

[3] R. Billi et al. : "Field Trial Evaluation of Two Different Inquiry Systems"; *Proc. this IVTTA '96 Workshop*.

[4] R. Naktsu, "Anser : an application of speech technology to the Japanese banking industry," *IEEE computer*, Vol. 23, No. 8, pp. 43~48, Aug. 1990.

[5] George 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.

[6] 김재인, 구명완, "음성인식 증권정보시스템의 개발 및 시험운용결과 분석," *음성통신 및 신호처리 워크샵 논문집*, 제13회, pp. 185~191, 1996.

[7] L. Rabiner and B. H. Juang, *Fundamentals of speech recognition*, Prentice-Hall, NJ, 1993.



김 재 인

1981 고려대학교 전자공학과 졸업(공학사)
 1986 고려대학교 대학원 전자공학
 학과 졸업(공학석사)
 1986 음성전기연구소(연구원)
 1988~현재 한국통신 멀티미디어 연구소 음성언어 팀 선임연구원
 1996 고려대학교 대학원 전자공학과 졸업(공학박사)
 관심분야: 음성인식, 화자인식, 음성합성 등 음성처리분야

● HCI '98 학술대회 ●

- 일 자 : 1998년 2월 18일(수)~20일(금)
- 장 소 : 피닉스 파크 컨벤션센터
- 주 최 : HCI연구회
- 문 의 처 : 고려대학교 컴퓨터학과 이성환 교수

Tel. 02-3290-3197, E-mail : swlee@image.korea.ac.kr