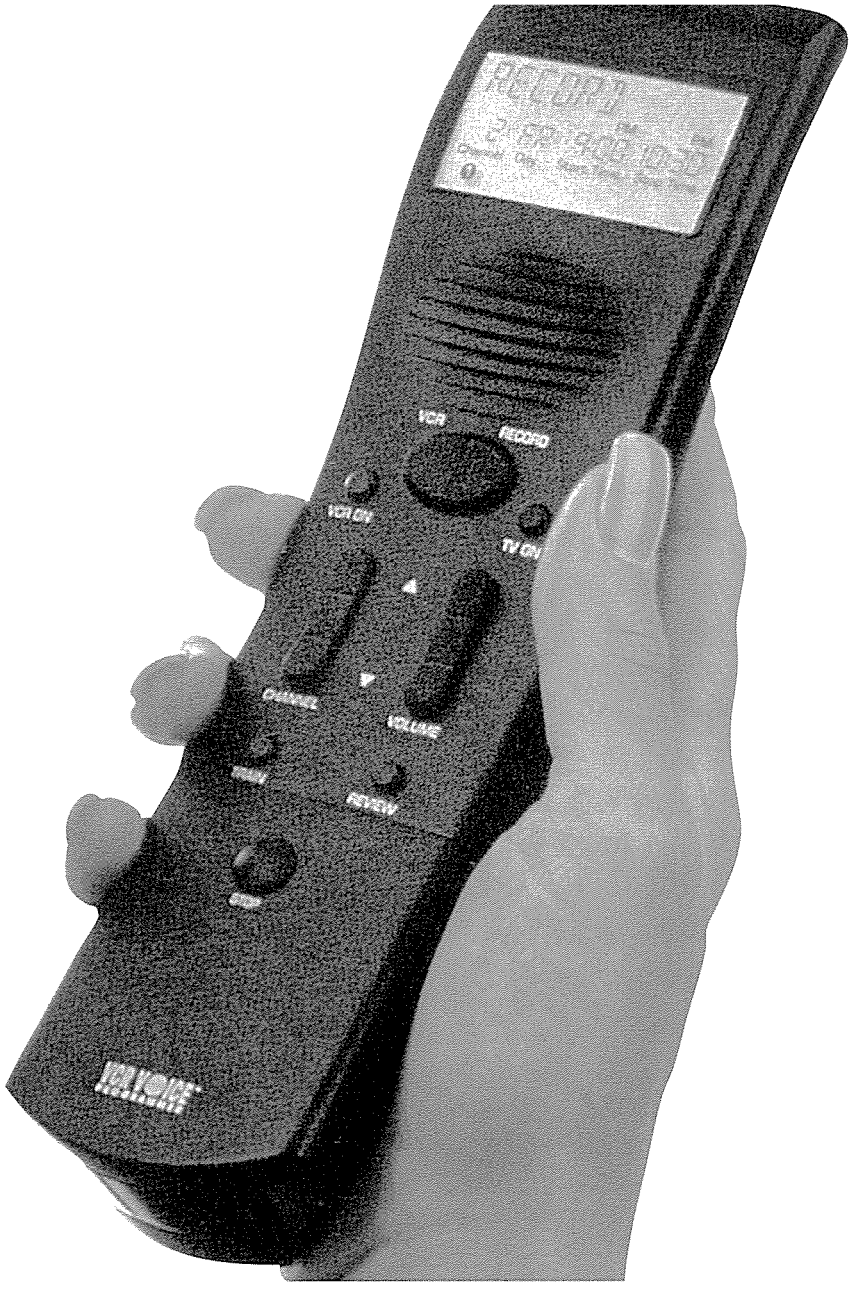


◀ 음성명령으로 VTR을 프로그램할 수 있는 음성 인식 프로그래머.

첨단과학기술현장

음성인식기술의 오늘과 내일

기계가 말하고 알아 듣는 시대가 빠른 걸음으로 다가오고 있다. 21세기 초에는 외국어를 모르는 사람들도 외국인과 대화할 때 거추장스럽게 통역관을 내세우지 않아도 된다. 통역용 소프트웨어가 내장된 전화기가 등장하는가 하면 포켓용 통역장치가 그때 그때 대화내용을 우리말과 외국어로 옮겨 합성소리로 알려 준다.

玄 源 福 〈과학저널리스트/본지 편집위원〉

인간과 기계의 결합

몇해 전부터 컴퓨터의 글자판이나 전화기의 다이얼에는 손을 대지 않고도 컴퓨터에게 문서작성을 시키거나 지정한 사람에게 전화를 걸 수 있는 음성인식시스템들이 줄줄이 등장하기 시작했다. 최근에는 미국의 드래건시스템즈, IBM 그리고 런아웃호스파이사(L&H)가 제1세대의 '연속식' 음성인식시스템을 내놓았다. 이런 시스템을 내장한 워드프로세서는 비교적 정확하게 보통하는 말을 옮겨 쓸 수 있어 속기사들의 일자리를 위협하고 있다. 2~3년 전부터 국내에서도 일본어-한국어 번역 소프트웨어들이 쏟아져 나와 불꽃튀는 판매전을 벌이고 있다.

1998년에는 L&H사가 차량항법장치(카 네비게이션시스템)에 음성인식기술을 도입하기 시작한다. 그래서 예컨대 서울 올림픽대로를 달리고 있는 운전자가 "반포대교에서 김포공항까지의 거리를 알고 싶다."고 말하면 내장된 컴퓨터가 지체없이 인공음성으로 답변해 준다. 또 이 기업이 개발중인 휴대용 번역기는 예컨대 북경의 거리를 걷고 있는 여행자가 지나가는 행인(중국인)에게 영어로 가장 가까운 호텔의 위치를 물으면 이것은 중국어로 동시통역이 되고 이에 대해 행인이 중국어로 답변한 것은 영어로 동시통역이 된다.

2001년에는 미국에서 팔리는 거의 모든 컴퓨터가 음성인식기능을 내장할 전망이다. L&H사가 개발한 음성인식용 소프트웨어인 '보이스코맨드'를 사용하면 인공지능전문가들이 말하는 이른바 자연언어(한 집단의 사람들이 모국어로 사용하는 언어)를 사용하여 어떤 일이든 컴퓨터에게 일을 시킬 수 있다. 마이크로소프트 워드 97이나 7.0에 대해서만 가동되는 이 소프트웨어는 예컨대 "3번째 파라그래프를 선택하라. 볼드체로 나타내라. 이탤릭체로 바꿔라. 중심으로 옮겨라"고 말로 명령하면 말이 떨어지기 바쁘게 '워드'의 복잡한 메뉴나무와 상의할 것 없이 시키는 대로 실행한다. 그런데 보이스코맨드의 어휘는 한정되어 있기 때문에 받아쓰기 시스템처럼 사용자가 발음 훈련없이도 처음부터 이 프로그램을 사용할 수 있다.

▲ 1997년 20년간에 걸친 연구개발 끝에 내놓은 음성인식프로그램 '내츄럴리 스피킹'으로 각광을 받고있는 제임스와 제넷 베이커부부.

'정보바다'의 길잡이

한편 우리가 보통하는 말을 그 자리에서 글로 옮길 수 있는 음성인식 시스템 개발을 20년 전에 착수한 제임스 및 제넷 베이커부부는 1997년 마침내 최초의 연속식 받아쓰기 소프트웨어 '내추럴리 스피킹'을 내놓았다. 미국 매서추세츠주 뉴턴시 소재 드래건시스템사가 선보인 이 음성인식용 소프트웨어를 이용하면 컴퓨터가 분당 1백50단어의 말을 95%의 정확성을 가지고 글로 옮길 수 있다. 또 컴퓨터가 "이것을 복사하라!"든가 "이것을 지우라!"는 말을 알아듣고 그대로 실행한다.

그런데 IBM이 개발중인 음성인식 기술은 예컨대 병원에서 종래 글자판을 두드리거나 마우스를 사용하여 X선 사진이나 컴퓨터단층사진 등 필요한 자료를 데이터베이스에서 불러 냈으나 앞으로는 말로 명령만 하면 된다. IBM은 또 생활의 편의나 업무의 효율을 끌어 올리기 위한 음성인식기술을 개발하여 21세기 초에는 줄줄이 선보일 계획이다. 예컨대 공항종합안내용 음성인식시스템은 여행자들이 시스템 앞에 서서 마이크에 대고 원하는 여객기, 가고자 하는 곳, 주머니 사정(여행예산) 등을 이야기 하면 기계가 여행자의 요구와 사정에 알맞는 스케줄을 보여주는 것은 물론 예약서비스까지 제공한다.

그런데 IBM은 패키지로 된 소프트웨어보다는 음성을 이용하는 제품개발에 장기적인 노력을 기울이고 있다. 이 기업이 최우선으로 착수한 연구과제는 21세기에 조(兆)비트단위로 데이터베이스에 저장될 '정보의 바다'를 어떻게 다룰 것인가 하

는 문제다. 예컨대 50년간의 지구 화폐 및 이자율에 관한 데이터에서 지구상에 생존하는 모든 동식물종의 DNA(유전자를 구성하는 분자화합물)부호에 이르기까지 아찔할 정도로 방대한 양의 정보를 저장한 데이터베이스에서 키펀치나 마우스를 이용하여 정보를 찾아내기 어려운 시대가 10년 내에 도래할 것으로 보인다. 그래서 IBM은 2백여명의 음성인식엔지니어들을 동원하여 이런 시스템을 다루는 데이터베이스관리 소프트웨어와 하드웨어를 개발하여 고객들이 말로 명령하여 인터넷이나 또는 다른 데이터베이스에서 필요한 정보를 불러내는 길을 열어 줄 계획이다.

떠오르는 스마트전화

미국의 주요한 전화회사들은 수년 전부터 음성인식기술에 많은 노력을 기울여 왔다.

예컨대 루센트, AT&T, 노던텔레콤 그리고 GTE사는 각기 자체의 음성인식기술을 개발하여 제품에 이용하고 있다. AT&T의 경우는 벨연구소의 음성인식소프트웨어로 신용카드 호출과 수금을 함으로써 지난 6년간 수십억달러의 비용을 절약했다. 한편 노던텔레콤사는 벨 캐나다사에게 월간 4백만회의 전화번호 호출서비스를 제공할 수 있는 시스템을 제공하고 있다. 노던텔레콤의 소프트웨어는 자연어를 다룰 수 있는 프로그램을 갖추고 있어 호출자가 더듬거리면서 전화번호를 물어도 어렵지 않게 호응할 수 있다. 최근 우리나라에서도 삼성전자와 LG정보통신이 각각 2가지 모델의 음성인식 휴대폰과 전화기를 내놓아 빠른 걸음으로 보급되고 있다. 이 전화는 사용자가 '본사'나 또는 '우리집'이라고 말하면 전화 속에 내장된 음성인식칩이 알아 듣고 자동으로 지시한 곳으로 전화를 걸어준다. 현재 음성인식률은 85~95% 정도로 알려졌으나 차의 창문이 열렸거나 시속 1백km 이상의 경우는 80% 정도로 떨어지는 것으로 알려져 있다.

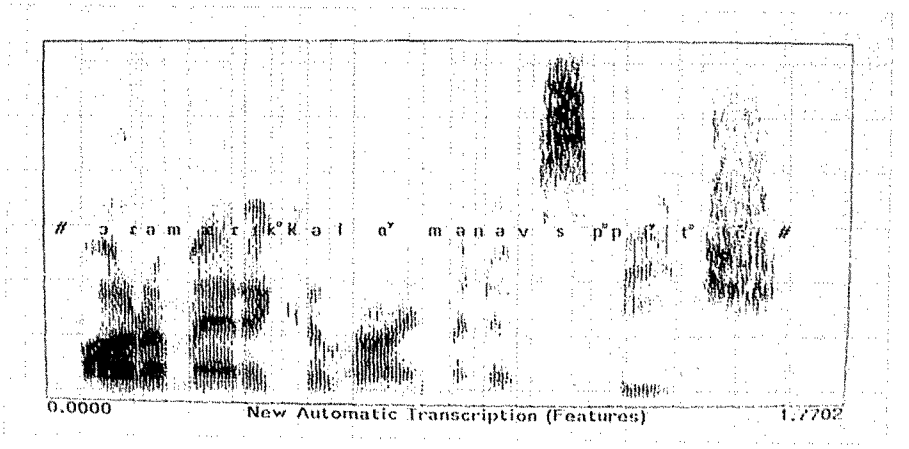
미국 매서추세츠공대(MIT) 컴퓨터과학연구소의 부소장 빅터 주에는 미국에서 인터넷과 접속하고 있는 인구는 3천만인데 비해 9천만가구가 전화를 갖고 있다고 전제하고 "미래의 정보접촉은 전화로 이루어져야 한다"고 주장하고 있다. 루센트 벨연구소의 연구는 이런 지적을 뒷받침하고 있다. 퍼시픽 벨 모바일 서비스사는 와일드화이어 커뮤니케이션사가 개발한 음성활성화이동시스템을 실험하고 있다. 이 시스템을 이용하면 운전자들은 핸들에서 손을 떼지 않고도 전화호출을 하거나 음성우편 메시지를 검색할 수 있다.

주에가 개발한 '쥬피터' 시스템은 소비자들이 800번(무료전화번호)을 돌려 세계 5백개 도시의 기상정보를 얻을 수 있다. 쥬피터는 기상에 관한 화제라면 이용하는 사람이 어떤 낱말을 택하든 상관없이 모두 호응할 수 있다. 예컨대 "북경은 더운가?" 또는 "내일 뉴욕에는 비가 올 것인가 알려 달라"고 물으면 쥬피터는 적절한 회답을 한다. "런던의 기상은 어떤가?"고 물으면 쥬피터는 "영국의 런던말인가 또는 미국 켄터키주의 런던말인가?"고 되묻기도 한다. 쥬피터의 소프트웨어는 하루에도 여러번 웹과 연결되어 각각 2개의 미국 국내 및 국제기상컴퓨터서버에서 최신 기상정보를 읽는다.

상식갖춘 기계

지난 10년간 MIT, 카네기멜론대학 그리고 GTE와 같은 연구기관에 해마다 1천만달러에서 1천5백만달러의 음성연구비를 지원한 미(美)국방부 첨단연구사업국은 앞으로는 대화와 관련된 연구에 더 많은 무게를 실을 것으로 알려져 있다. 오늘날 기계와의 대화는 대부분 한번 또는 두차례의 대화로 마감된다. 예컨대 주식의 값이나 영화에 관한 질문을 던지면 기계는 질문의 내용을 분명하게 밝혀 달라고 요구한다. 다시 한두가지의 정보를 제공하면 그것으로 컴퓨터는 거래를 마감한다. 그러나 앞으로는 10차례 이상의 대화를 할 수 있는 시스템을 개발하여 여행자들이 예컨대 5일간, 3개의 다른 도시를 여행하는 항공표를 사는데 도움을 줄 수 있는 시스템을 개발할 계획이다.

그런데 컴퓨터가 이런 기대를 충족시키기 위해서는 세상이 어떻게 돌아가는가에 관한 많은 지식을 갖출 필요가 있다. 이것은 인공지능전문가들이 오랜 세월에 걸쳐 당면하고 있는 것과 같은 문제다. 예컨대 텍서스주 오스틴 소재 사이코사의 사장이며 인공지능전문가인 더글러스 르넷은 지난 10년간 컴퓨터가 세상을 이해하는데 도움이 될 상식적인 사실과 그 관계를 컴퓨터에 주입하고 있다. 「사이」(백과사전이라는 뜻의 영어에서 따온 말)라고 부르는 그의 시스템은 현재 새 정보를 독자



▲ MIT의 빅터 주에교수는 8년간 소리의 스펙트로그램(오른쪽 그림)을 연구하여 음성인식기술 수준을 한단계 끌어 올렸다.

적으로 발견할 수 있을 정도는 되었으나 인간이 교환하는 상식을 갖추기에는 아직도 많은 세월이 필요하다. 지금까지 약 4천만달러의 연구비를 사용하여 예컨대 어머니는 언제나 딸보다 더 늙었다거나 새들은 털을 가졌다거나 하는 등의 2백만건의 상식에서 나온 약 50만개의 규칙을 사이에게 입력했다. 사이는 이어 신문과 책과 과학저널을 읽고 스스로 배우기 시작하여 8~9년 내에 대학원생 수준의 일을 할 수 있을 정도로 지능이 높아지면 2020년경에는 스스로의 연구실을 운영하면서 새로운 지식을 탐지할 수 있는 수준에 이른다는 것이다.

장애자의 벗

음성인식기술은 수천만의 장애자들에게 새로운 삶의 길을 열어 주기 시작했다. 실명, 난독증(難讀症) 그리고 실어증과 같은 마비상태로 고통을 받고 있는 사람들은 음성인식기술을 마스터키로 사용하여 커뮤니케이션과 자립의 터전을 잡고 있다. 예컨대 난독증환자들은 컴퓨터에게 글을 읽히게 할 수 있어 다른 사람의 도움을 청할 필요가 없다. 귀가 먼 사람들은 컴퓨터 시뮬레이션을 사용하여 낱말을 발음할 때 입을 어떻게 움직이는가를 정확하게 알 수 있다. 음성인식기술이 미치는 경제적인 영향은 막대하다. 예컨대 장애보조비를 받는 사람이 2천6백만에 이르는 미국에서는 연방정부가 연간 지출하는 비용이 2천억달러에 이른다. 보다 우수한 음성인식기술을 개발하면 많은 사람들에게 보다 생산적인 일을 제공할 수 있어 수십억달러를 절약할 수 있게 된다.

영국의 이론물리학자이며 베스트셀러작가인 스티븐 호킹은 장애인이 사회에 대해 헤아릴 수 없이 큰 공헌을 할 수 있는 보기가 되고 있다. 근위축성측색경화증(ALS)이라는 퇴행성 뇌질환을 앓고 있는 호킹은 미국의 디지털 이큅먼트사(DEC)가 개발한 컴퓨터합성소리를 사용하여 커뮤니케이션을 한다. 첨단기술의 결정적인 도움을 받아 호킹의 저술과 강의는 사람들에게 우주와 블랙홀의 성격을 이해시키는데 중요한 역할을 하고 있다. 퍼스널 컴퓨터의 역량이 강력해지면서 음성인식시스템은 데스크톱의 값을 끌어 내리고 있다. 호킹의 DEC토크 시스템의 값은 1980년대 초에는 4천5백달러나 했으나 현재 홑 2백달러로 떨어졌다.

음성인식기술은 최근 인터넷으로 진출하여 장애자의 벗이 되고 있다. IBM저팬사의 맹인연구자인 사사카와 치에코를 포함한 연구팀은 맹인들에게 월드 와이드 웹(WWW)을 개방하는 브라우저를 개발했다. IBM홈페이지 독자들은 컴퓨터합성의 남성 음성을 사용하여 일본어로 정상적인 글을 읽을 수 있다. IBM이 현재 개발중인 영문판은 1998년 후반에 선을 보인다. 음성인식기술은 또 귀먹은 사람들을 돕고 있다. 미국 오리건주 포틀랜드의 터커-맥슨학교의 귀먹은 아동들은 '볼디'라고 부르는 컴퓨터를 이용한 말하는 두뇌로부터 낱말의 발음 방법을 배우는 신기한 경험을 하고 있다. 볼디의 가장 관심을 끄는 특징은 자기가 사용하는 낱말에 어울리는 얼굴 표정을 재생하는 능력이다. 학생들

은 볼디를 이용하여 전통적인 발음 훈련을 하고 있다.

1백여년 전 알렉산더 그라함 벨은 전화를 발견할 때 귀먹은 자기 부인을 위해 낱말을 영상으로 옮기려고 애를 썼다. 장차 장애자를 도울 수 있는 혁명적인 기술에는 어떤 것이 있을까? 말을 글로 옮길 수 있는 장치는 외국어를 동시번역하는데 도움을 줄 수 있을 것이다. 컴퓨터스크린의 커서는 육체적으로 불구가 된 사람들이 컴퓨터를 사용하는 것을 돕기 위해 만든 것이다. 그러나 앞으로는 마우스의 대용품이 등장하는가 하면 귀먹은 아동들이 말을 배우고 눈먼 사람들이 웹을 훑는 방법을 배우는 길이 열릴 것으로 기대된다.

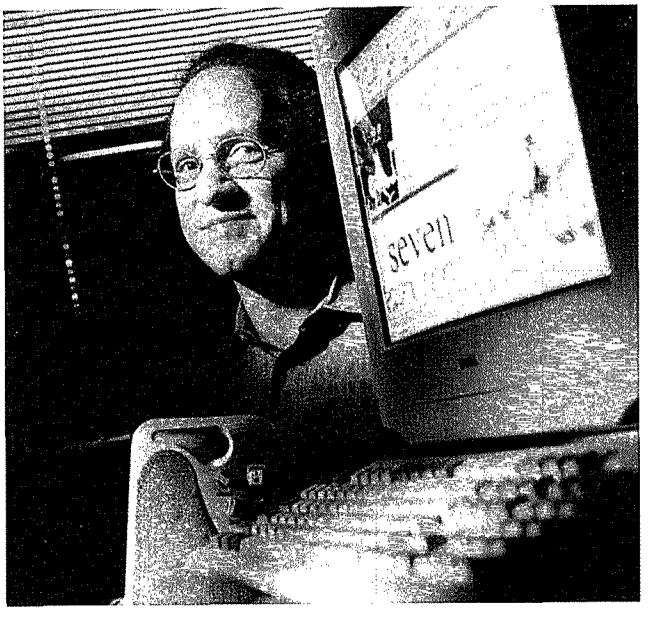
치열한 개발경쟁

현재 벨기에 남부지방에서 건설중인 플랜더즈 랜기지 밸리(FLV)는 5억달러를 투자하여 인간과 기계사이의 틈새를 메울 수 있는 소프트웨어 개발에 전념하게 된다. 벨기에 음성인식계의 대표적인 기업인 런아웃호스파이(L&H)사가 주축이 되어 개발중인 FLV는 마이크로소프트사의 지원으로 유럽 음성기술의 메카로 육성된다. 그러나 벨기에가 유럽 언어기술의 유일한 보금자리는 아니다. 유럽대륙에는 언어번역하는 기계를 개발하는 대기업의 연구소와 대학의 모험기업들이 도처에서 활동하고 있다. 예컨대 비엔나에서는 필립전자회사가 전화를 통해 질문을 하고 철도정보를 이해하는 독일어 프로그램을 완성중이다. 다임러 벤츠사 연구자들은 여러 나라 언어로 말하는 명령(예컨대 '열을 올려라!' 등)에 호응하는 첨단자동차 계기판을 개발중이다. 영국 케임브리지대학의 과학자들은 음성인식기술에 관한 미 국방부 심사에서 최고 평가를 받았다.

그러나 음성인식장비 시장에서는 벨기에의 L&H사가 다른 경쟁사들보다 한발 앞서고 있다. 9천9백만달러의 연간 매출고를 가진 이 기업은 음성인식기술의 선발기업인 미국의 커웨일사와 멘제즈사를 흡수 통합함으로써 듣고 말하고 번역하는 광범위한 기본기술을 축적했다. 현재 1백8종의 하드웨어와 소프트웨어 제품을 팔고 있는 L&H사는 마이크로소프트사와 함께 신형 오토PC 차량 항법장치용의 음성인식프로그램을 개발하고 있다. 이밖에도 인터넷용 문장번역프로그램과 1백달러의 받아쓰기 프로그램 개발에 착수했다. '보이스 익스프레스'라는 이름의 이 프로그램은 말을 글로 옮겨쓰고 말로 하는 명령에 복종할 수 있다. 한편 인터넷 번역프로그램은 1998년 후반기에 선보일 예정이다. 그런데 가장 뛰어난 음성인식기술들은 케임브리지대학과 그 산하기업인 엔트로픽 그룹에서 나오고 있다. 1995년에 창립된 엔트로픽그룹은 대화용 키트와 구술의 필기를 포함하여 여러 가지의 음성인식프로그램을 팔고 있다.

한편 한자를 바탕으로 하는 언어는 너무 복잡하여 중국과 일본과 같은 나라들이 디지털시대의 장점을 이용하는 길을 막고 있다. 중국은 4천에서 6천자의 글자를 사용하는데 글자 하나가 획이 10여개나 되는 것이 있다. 일본은 이보다 적은 2천자 안팎의 한자를 사용하고 있으나 로마자와 '가나'라고 하는 2가지의 음표문자 등 3가지의 글자를 병행하여 사용하고 있다. 그래서 해결책은 글자판을 회피할 수 있는 음성인식기술에 달려 있다고 하겠다. 실제로 1997년부터 이런 언어장벽을 허무는 길을 열어 줄 제품들이 조용히 나타나기 시작했다. 예컨대 일본전기사(NEC)는 1997년 사용자가 소리로 된 전자우편을 글로 표기할 수 있는 80달러짜리의 제품(헤드셋과

▲ 호킹박사가 사용하는 합성소리장치의 값은 1980년대에는 4천5백달러나 했으나 이제 기술개발 덕에 2백달러로 떨어졌다.



◀ 난독증(難讀症) 환자용으로 '읽는 기계'를 개발중인 레이 커즈웨일은 최초의 음성인식장치를 개발한 이 분야의 선구자다.

마이크로폰)을 내놓았다. 한편 IBM의 '바이어보이스'(99달러)는 1997년 9월부터 중국에서 저품위 PC와 묶여 시판되기 시작했다. 그러나 중국어나 일본어는 소리는 같지만 뜻이 틀리는 이른바 동음이의어(同音異義語)로 뒤섞여 있다. 그래서 음성인식프로그램은 중국어의 다른 음색과 방언을 가려 낼 수 있어야 한다. 엔지니어들은 말하는 기계에다 보다 빠른 반도체칩과 광범한 언어샘플을 제공함으로써 이런 장애를 극복할 수 있다고 생각하고 있다.

한편 우리나라의 한국통신과 한국전자통신연구소가 일본의 국제전신전화(KDD)와 공동으로 개발중인 자동통역전화가 완성되면 한일 양국간의 전화통화는 상대국의 언어를 동시 통역해 주기 때문에 언어의 장벽을 허물 것으로 기대되고 있다. 그러나 자동통역전화에서는 음성인식·언어번역·음성합성이 리얼타임에 가까운 속도로 이루어져야 하기 때문에 아직도 몇가지 기술적인 벽을 넘어야 한다. 일본의 국제전기통신기초기술연구소(ATR)는 한국전자통신연구소 외에도 독일의 칼스루에 공대와 독일인공지능연구소, 미국의 MIT, 링컨연구소와 스탠퍼드연구소, 프랑스과학기술연구소, 이탈리아의 기계공학·기술과학연구소 등과 제휴하면서 일어, 한국어, 영어, 독일어를 대상으로 1999년에 다(多)언어음성번역 국제실험을 거친 뒤 2000년 초에는 음성번역기능을 가진 전화기를 선보일 계획이다.

장벽을 넘고

인간과 기계의 결합은 공상과학소설이나 영화에서 가장 흥미를 끄는 대상이다. 예컨대 '2001 : 스페이스 오디세이(2001년 우주여행)'에 등장하는 할이나 '스타 워즈(별들의 전쟁)'에 나오는 C3PO와 같은 로봇을 볼 때 기계에게 말을 시키는 일은 어렵지 않은 것처럼 보인다. 그러나 실제로 컴퓨터에게 말을 시킨다는 것은 엄청나게 힘든 일이다. 낱말을 구성하는 소리는 수십개의 얽히고 설킨 주파들로 되어 있을 뿐 아니라 이런 주파들은 발언자가 얼마나 빨리 또는 크게 말하는가에 따라 수시로 변한다. 더욱이 똑똑하지 않거나 빨리 발음할 때 주파의 패턴은 완전히 바뀐다. 컴퓨터는 저장된 말의 음향학적 모델을 참조하고 통계기법을 이용하여 이런 문제를 해결한다. 기계는 또 구문론과 문법의 규칙을 배울 수 있다. 그러나 인간은 문법대로 말하지 않는 일이 흔하다. 또 문법대로 말한다고 해도 사투리나 농담이나 생략하여 이야기할 때 기계는 과연 이해할 수 있을까?

최근 선을 보인 드래건사의 '내추럴 스피킹'이나 IBM사의 '바이어보이스'와 같은 받아쓰기 프로그램이 95%의 정확도를 달성할 수 있다고 하지만 실은 이것도 이상적인 환경에서만 가능하다. 여러 사람들을 한방에 집어 넣고 활발한 토론의 불을 지른 뒤 받아쓰기 프로그램을 가동시키면 별안간 오류율은 10%에서 50%로 뛰어 오른다. 이것은 곧 하나 건너씩 낱말이 틀린다는 것을 말한다. 음성인식기술은 미국어를 독일식이나 스페인식 또는 미국 남부지방의 액센트로 발음하건 또는 이동전화나 공항 또는 스피커폰으로 이야기하건 상관없이 작동할 수 있어야 한다. 음성인식기술이 해결해야 할 과제는 한두가지가 아니지만 우리에게는 말이 가장 쉽고 풍부한 커뮤니케이션의 방법이다. 일단 컴퓨터와 자연스럽게 말을 할 수 있게 되면 사람과 기계사이를 가로 막는 거대한 장벽은 사라지고 만다. ST