


 특 집

음성정보처리기술 응용서비스

구 명 완* 김 재 인**

(목 차)

1. 서 론
2. 유/무선 통신망 기반 응용서비스
3. 내장형 단말 기반 응용서비스
4. 데스크 탑 PC 기반 응용서비스
5. 결 론

1. 서 론

음성정보처리 기술은 사람의 말을 음향, 언어, 심리학 및 공학적인 측면에서 연구하여 사람과 기계 사이의 인터페이스를 자연스럽게 하는 것을 목표로 하고 있으며, 음성인식, 음성합성 및 언어처리 기술로 이루어져 있다. 음성인식기술이란 사람의 말을 이해하는 것뿐만 아니라 화자를 식별하고 인증하는 기술도 포함하고 있으며, 음성합성기술이란 문자로부터 음성을 생성하는 기술을 의미한다. 그리고 언어처리 기술은 음성인식, 음성합성기술 속에 포함될 수 있으나 최근 마크업 언어를 활용하여 음성인식, 합성 등을 제어하는 경향이 도래함에 따라 언어처리 기술을 따로 분류하기도 한다[1][2].

기술응용분야를 크게 분류하면 유/무선 통신망 환경에서의 응용, 단말기 기반 응용 및 데스크 탑 PC에서의 응용 등으로 분류된다. 유/무선 통신망 환경에서의 응용서비스가 가장 활발한 음성정보처리 기술의 응용서비스 시장으로 알려져 있으며 국내, 외에서 다양한 응용서비스가 개발되고 있다[3]. 일반적으로 유/무선 단말기에서는 음성을 처

리하지 않고 통신망의 상위에 구축된 음성정보처리 서버를 통하여 음성인식, 합성을 수행한다. 그러므로 고가의 장비를 사용하며 성능에 초점을 맞추어 서비스를 제공한다[4]. 단말기 기반 응용서비스는 단말기에 음성인식, 합성기능을 제공하여 단말 사용자들에서 편리한 서비스가 되도록 하는 것을 목표로 한다. 반면 단말기의 성능에 한계가 있으므로 음성인식, 합성기술이 저가의 성능에 맞도록 설계되어야 한다. 최근에는 휴대폰, PDA의 성능이 개선되고 있으므로 과거에 비해 음성인식, 합성의 품질이 향상되고 있다. 데스크 탑 PC기반 응용서비스는 PC 사용자들에게 편리한 인터페이스를 제공하는 것을 목표로 하며 음성인식 받아쓰기, 영어 발음교정 서비스 등 단순한 인터페이스를 넘어 특정한 기능을 제공하는 방향으로 발전하고 있다[5].

본고에서는 음성정보처리 기술을 이용한 응용서비스 현황에서 대하여 먼저 2장에서는 유/무선 통신망 응용서비스에 대하여 소개하고, 3장에서는 내장형 단말기 기반 응용서비스에 대하여 기술한다. 그리고 4장에서는 데스크 탑 PC기반의 응용서비스에 대하여 알아보기로 하겠으며 마지막으로 5장에서 결론을 맺도록 하겠다.

* KT 서비스개발연구소 음성언어연구팀장

** KT 서비스개발연구소 음성기반연구실

2. 유/무선 통신망 기반 응용서비스

유/무선 통신망 기반 응용서비스의 최근 동향에 대한 상세 내용은 2003년 7월호에 실려 있으며[3] 여기서는 이전에 언급하지 않은 내용을 중심으로 기술한다. 유/무선 통신망 기반 응용서비스는 크게 음성인식 전화번호 안내 서비스, 음성다이얼 서비스, 음성인식 콜센터 서비스 및 보이스포탈 서비스로 분류할 수 있다.

2.1 음성인식 전화번호 안내 서비스

미국 중심의 음성인식 전화번호 안내 서비스는 음성인식 기능의 한계를 고려하기 때문에 음성인식 기능만을 활용하여 완전히 자동화하는 것을 피하고 현재의 음성인식 기술로 가능한 영역만 자동화하고 나머지 서비스를 위해 안내원으로 자동 연결하도록 하여 서비스가 지속적으로 이루어지도록 한다. 예를 들면 전화번호를 안내하는 서비스인 경우에 도시 이름만을 먼저 음성인식을 하고 다음에는 담당 안내원으로 연결해 주는 방향으로 진행되고 있다.

반면 유럽 중심의 음성인식 전화번호 안내서비스는 저가의 서비스 요금을 받는 음성인식 기반 전화번호 안내서비스와 고가의 서비스 요금을 받는 안내원 기반 전화번호 안내서비스로 나누어진다. 음성인식 기반 전화번호 안내 서비스는 저가인 대신 음성인식 성능에 의한 품질 저하를 고객이 감수해야 한다는 단점이 있다. 예를 들면 핀란드의 020200 서비스, 덴마크, 스웨덴의 전화번호 안내 서비스가 유사한 방식으로 제공되고 있다. 일본에서는 특정 회사의 휴대폰을 가입하는 가입자를 대상으로 무료로 음성인식 기반 전화번호 안내 서비스(#9887 서비스)를 해주고 있다.

2.2 음성다이얼 서비스

가장 대표적인 음성다이얼 서비스가 스포린트사

가 제공하는 음성다이얼 서비스다. 이 서비스는 단말기 기반 서비스와 네트워크 기반 서비스로 나누어진다. 단말기 기반 서비스는 단말기에 기본적인 음성인식 기능을 수행하도록 하는 단말기 기반 서비스다. 단말기는 전화번호 숫자음, 간단한 음성 명령이 가능하며, 네트워크 기반 서비스와 병행하여 사용될 수 있다. 네트워크 기반 서비스는 음성인식 기능이 서버에서 수행되는 서비스로서 서비스가 입자는 웹상에 전화번호와 대표이름을 개인별로 등록하여 사용한다. 요금은 서비스 개시 몇 개월의 무료 서비스에 이어 월 \$5 정도다. 개인별로 500개 정도의 이름을 등록할 수 있으며 각 이름 당 5대의 유선전화, 무선전화, 사무실 전화 등의 번호를 입력할 수 있다. 또한 음성다이얼 서비스 이외에도 증권, 날씨, 뉴스, 운세 등의 서비스도 제공하고 있다. 또한 기업체 내의 직원용으로 사용할 수 있는 기업용 솔루션도 제공하고 있다. 2001년 6월부터 음성인식 기반 정보 및 다이얼 서비스를 시작하였으며 2001년 10월부터는 현 위치에 가까운 식당을 안내해 줄 수 있는 식당 안내 서비스도 개시하였다. 이 서비스는 미국 내에 있는 12,000개 이상의 도시에 있는 260,000개 이상의 식당에 대한 정보를 제공해주며 사용하는 방법은 *TALK을 누른 다음 "Call Restaurants" 라고 말하면 된다. 멕시코에서도 Pegaso PCS회사에서 음성 페가소라는 이름으로 유사한 서비스를 제공하고 있다.

야후 바이폰("Yahoo! by Phone")은 음성 다이얼 서비스를 제공하여 주는 것 이외에도 전자 메일이나 음성 메일의 청취, 주식 정보, 날씨, 스포츠 스코어, 뉴스 등의 청취를 해 주는 서비스다. 이 서비스는 1-800-MY-YAHOO(1-800-699-2466)로 전화를 걸고 주소록에 있는 사람의 이름을 말하면 그 사람에 대한 정보를 들려주며, 직장이나 휴대폰, 집 전화번호, 팩스 및 다른 전화번호를 비롯하여 전자 메일 주소까지 알 수 있다. 현재 한 달에 \$4.95 서

비스 요금을 받고 있으며 웹 사이트(<http://phone.yahoo.com>)를 통하여 등록할 수 있다.

또한 유사한 서비스로 AOLbyPhone 이라는 서비스가 있다. 이 서비스는 1-800-AOL-1234로 전화를 걸어 음성 메뉴로 뉴스, 날씨, 전자 메일 확인 등을 할 수 있다. 웹 사이트(<http://www.aolbyphone.com>)을 통해서 서비스 등록을 할 수 있다. 30일간 무료 서비스 후에 한 달에 \$3.95 서비스 요금을 받고 있다.

미국 Information Services Extended(ISx)사는 FlexiQ라는 전화변화 자동안내 시스템 구현에 적합한 솔루션을 개발하여 France Telecom, MCI, SBC, Sprint 사 등에 제공하여 전화번호 안내서비스에 이용되고 있다. 특히 주, 혹은 도시 이름만 음성인식하고 안내원으로 연결시킬 수도 있으며 완전히 음성인식으로 자동화 할 수 있는 기능을 포함하고 있다.

국내에서도 KT가 Everlink 1582 음성다이얼 서비스를 개설하여 유사한 서비스를 제공하고 있다[7].

2.3 음성인식 콜 센터 서비스

음성인식 콜 센터 서비스란 콜 센터로 전화를 거는 사람들의 다양한 요구를 자동으로 인식하여 미리 설정한 업무로 자동 분류하여 담당자로 연결해주는 서비스다. 이러한 서비스의 목표는 안내원의 인건비를 절감하는 것이며, 이를 위해 간단한 질의에 대한 답변을 위해 음성인식 기술을 사용한다. 최근에는 다양한 질문이 이루어지더라도 정황을 이해하여 올바른 담당자로 연결시켜주는 알고리즘 개발에 대한 연구가 진척되고, 특히 인터넷 웹과 연동되어 유/무선 전화망과 동시에 고객을 관리하는 서비스로 진화됨에 따라 컨택트 센터라는 용어가 생겼다. 즉, 유/무선 전화망기반의 콜 센터와 인터넷 웹 기반의 홈페이지가 연결되는 컨택터 센터가 인기를 얻게 되었고 이러한 개념을 많은 기업에

서 도입하고 있다.

HMIHY("How May I Help You?") 서비스는 미국 AT&T사가 자사 고객센터를 위한 콜 센터에 도입한 음성인식 콜센터 서비스다. 이 서비스는 사용자가 마치 사람에게 말하는 것처럼 인사를 받고 요구를 하는 것처럼 느끼도록 하기 위하여 만들어졌다[6]. 음성인식 시스템이 하는 일은 사용자가 원하는 모든 것을 이해하는 것이 아니라 호를 적절한 곳으로 연결할 수 있을 만큼만 이해한다. 이러한 서비스가 되기 위해서는 사용자와 콜 센터 안내원과 많은 대화를 녹취하여 상관관계를 최대한 나타내어 주는 관련 확률을 찾아내어야 한다[8]. HMIHY는 대화 중의 내용을 62가지로 분류하여 담당자에게 호 전환하거나 자동안내 서비스를 제공하고 있다. 벨 캐나다에서도 310-BELL 이란 콜 센터에 음성인식 기능을 제공하고 있다.

미국의 Expedia 여행사는 1-800-EXPEDIA의 전화를 개설하여 음성인식 콜 센터를 개설하여 여행 관련 담당자로 연결해 주거나 관련 정보를 제공하여 주고 있다. 초기에는 호 분배(call routing)중심으로 진행하였으나 점차로 자동화 범위를 확대하고 있다. 또한 미국 Thrifty 자동차 렌탈 회사에서도 1-800-THRIFTY의 음성인식 처리 가능 전화를 개설하고 있다.

2.4 보이스포탈 서비스

보이스 포탈은 PC기반의 인터넷 포탈에 상응하여 핸드폰 혹은 유선전화를 통하여 인터넷 정보를 음성 인식, 합성기술을 사용하여 제공하기위한 방안에서 시작되었으나 현재의 음성처리 기술의 한계 때문에 성공하지는 못했다. 그 대신 인터넷 기반의 특정정보를 유/무선 통신망을 통하여 전달해주는 역할을 수행한다. 인터넷 기반으로 되어 있는 많은 정보를 유/무선 통신망과 결합시켜 사용자들에서 보다 고품질의 정보를 제공하는 역할을 한다.

최근에는 VoiceXML, SALT 등의 표준안이 제안되어 다양한 서비스를 쉽게 개발할 수 있다. 특히 최근에는 보이스포탈 서비스 플랫폼을 개발하여 ASP(application service provider)사업을 전개하고 있는 보이스포탈 서비스 회사가 많다.

텔미넷웍 회사는 VoiceXML 기반 플랫폼을 개발한 후에 원하는 기업을 대상으로 음성인식 정보 서비스를 ASP 사업화 하고 있다. 현재 주요 고객으로 멜릴 린치(Merrill Lynch), UnitedHealth 그룹, 페덱스, AT&T의 #121 보이스포탈 서비스 및 American Airlines의 분실 화물 안내 서비스 등이 있다. 이 회사는 포춘지의 500대 회사 중 20여개의 회사를 서비스 호스팅 고객으로 관리하고 있다.

보이스포탈 서비스를 위한 플랫폼은 주로 VoiceXML 기반 시스템이었으나 MS가 SALT기반의 보이스포탈 플랫폼을 최근 개발한 후부터 MS(MicroSoft)사의 솔루션과 VoiceXML 기반 솔루션과의 경쟁이 예상된다. Intervoice 사는 SALT기반의 MS 음성 서버를 이용한 보이스포탈 플랫폼을 상품화 하였다[6]. 이 플랫폼의 가장 큰 장점은 가격이 기존의 제품에 비해서 66-70% 정도 저가라는 것이다. 현재 이 제품을 사용한 데모 서비스가 +1-866-TRY-INTV에서 운용이 되고 있으며 중, 소형 기업의 대상(4-24 채널)의 시스템에 적합한 것으로 알려져 있다.

보이스포탈 서비스 플랫폼을 이용하면 다양한 전화정보 안내 서비스를 제공하여 줄 수 있다. United Airlines, Continental Airlines, Scandinavian Airline 등이 항공 예약, 출도착에 관련된 정보를 음성인식기능을 통하여 제공해 주고 있다. 이러한 서비스를 사용하면 상담원 상담 건수를 약 50% 정도 까지 줄일 수 있다고 한다. 또한 열차 정보안내 시스템에도 활용될 수 있다. 미국 Amtrack의 1-800-USA-RAIL 서비스, Union Pacific Railroad, 독일 VGN, 네덜란드의 OVR 등

의 철도 회사가 음성인식 전화정보 서비스를 운용하고 있다. 이외에도 금융, 증권 회사에서 인건비를 줄이는 기회로 음성인식 전화정보 서비스를 운용하고 있다.

싱가폴 우정국에서는 음성으로 건물이름이나 주소를 말하면 우편번호를 알려주는 서비스도 운용하고 있다(1-800-842-7678). 또한 미국 Office Depot 사무용품 전문 판매회사는 1-888-GO-DEPOT의 전화번호를 개설하여 사무용품의 재고 및 주문을 할 수 있도록 하였으며 McKesson사는 의약품의 재고 및 정보를 알려주는 전화정보 서비스를 운용하고 있다. 프랑스에서는 Traffic-One 이라는 서비스를 개발하여 파리근교의 교통정보를 음성인식기술을 통하여 교통상황과 이동시간을 알려주고 있으며 독일의 ADAC 미국의 511 서비스와 유사한 형태로 진행하고 있다.

국내에서도 농협, 철도예약, 증권 등에 음성인식 기반 전화번호 안내 서비스를 제공하고 있다.

3. 내장형 단말 기반 응용서비스

내장형 단말 기반 응용서비스는 단말 환경에서 음성처리기술이 설계 및 구현이 되어야 하므로 단말기의 주어지는 제약 환경에 대응하도록 알고리즘의 속도 및 메모리 등의 최적화가 이루어지도록 구현이 되어야 한다. 또한 각 단말기의 사용자 상황 및 주변 환경에 따라 필요한 주변 잡음 및 채널 환경에 대한 대책 등과 같은 요소 기술들이 동반 개발이 되고 결합이 되어야 한다. 내장형 단말 기반 응용서비스는 홈 네트워크 환경에서의 정보 가전 응용서비스와 휴대 단말기를 이용한 텔레매틱 응용서비스로 나누어 질 수 있다.

3.1 정보가전 응용서비스

가정 내 모든 디지털 가전 응용제품들은 네트워크와 연결되어 사용자와 항시 접속되어 제어될 수

있다. 이렇게 되면 지능적 부가 서비스를 실현 가능하게 하여 시간과 공간의 제약을 받지 않고 가정과 업무 환경 등의 일반 생활환경에서 가사, 학습, 여가, 업무의 질을 높일 수 있다. 또한 정보가전을 쉽게 사용하기 위해서 가전 단말기에 음성인식 및 합성기술을 구현하여 다양한 정보가전 응용서비스에 활용하고 있다.

Girl Tech사는 음성으로 조명, 스테레오 및 TV 등을 제어하거나 문을 여는 데 사용하는 기술을 개발하였으며 Innotech사도 Invoca라는 리모콘 제어를 개발하여 TV, VCR 및 DVD를 음성으로 제어할 수 있게 하였다[5]. 정보가전 응용서비스에 음성 인터페이스를 활용하는데 있어서의 장애는 생활 중에 나오는 소음과 정보가전을 제어하는 데 사용하는 명령어를 구분하기 위한 기술이 아직 개발이 되어 있지 않아서 리모콘을 사용해야 한다는 불편함이 있다는 것이다. 최근 KT에서는 가정 내의 가전을 VoiceXML기반 전화망을 통해서 외부에서 제어할 수 있는 서비스도 개발되었다.

3.2 텔레매틱 응용서비스

텔레매틱 응용서비스에 음성인터페이스는 매우 중요한 역할을 한다. 즉, 운전 중에는 휴대폰을 직접 조작할 수 없기 때문이다. 이러한 응용서비스를 구현하기 위해서는 앞에서 설명한 유/무선 통신망 기반의 응용서비스를 활용할 수 있으나 단말기에 음성인터페이스 기능을 장착하여 사용하는 경우도 있다. 특히 휴대 단말기의 기능이 PDA 뿐만 아니라 현재의 PC 수준으로 기술이 발전함에 따라 단말기에 음성 인터페이스를 구현하는 노력이 지속될 것이다.

삼성과 스프린트사는 최근에 개발한 VI660 핸드폰에 핸드폰 내의 신호세기, 배터리 수준 등을 음성 명령으로 알아낼 수 있으며 영어 숫자를 연속으로 말해서 전화를 걸도록 하는 기능을 내장하였다[6].

여기서 사용한 기술은 보이스시그널 사의 기술을 사용하였다. 파이오니아사는 DVD기반 자동 길안내 서비스에 음성인식 기능을 추가하여 출시하였다.

유럽의 ETSI(European Telecommunication Standard Institute)에서는 휴대폰의 성능이 좋아진다고 해도 다양한 단말 서비스가 제공되어야 하기 때문에 휴대폰에 간단한 단어를 인식하는 정도의 인식 기능만 제공하고 특징 추출한 후에 음성인식 서버로 전달하여 대용량 단어의 음성인식이 전용서버에서 이루어지도록 하는 DSR(Distributed Speech Recognition)방식을 표준으로 진행하고 있다. 또한 단말기에서 수행될 음성명령어의 표준화 작업도 병행하고 있다. 국내에서도 TTA를 통하여 PDA 음성 명령어 표준화 작업이 수년째 진행되고 있다.

4. 데스크 탑 PC 기반 응용서비스

Desktop PC 기반 응용서비스는 PC에 음성 명령을 해 주는 Command and Control 분야, 음성으로 문서를 작성할 수 있는 받아쓰기(dictation) 분야, 외국어 발음 교정을 비롯한 언어 교육 분야, 음성 제어 기능을 가지는 게임 분야, 멀티미디어 DB 내용을 음성으로 검색하도록 하는 음성 인덱싱 및 검색 분야 등으로 나눌 수 있다.

4.1 Command and Control

1993년도에 매킨토시 컴퓨터에 Casper 라는 음성인식 프로그램이 내장되어 음성 명령에 의해 각종 프로그램을 구동시키는 기능을 선보인 이래 다양한 제품들이 등장하고 있다. 대표적인 예로 Realize Software 사의 음성인식 Command and Control 소프트웨어인 Realize Voice는 윈도우 계열의 운영체제 하에서 음성명령에 의해 각종 응용 프로그램들을 수행시키고, 음성 매크로를 정의하여 음성으로 미리 정해진 키보드 입력들을 대행할 수

있으며, 주소록과 연계하여 음성 다이얼링 기능을 수행하는 등의 역할을 한다.

4.2 받아쓰기

연속음성인식에 의한 받아쓰기 프로그램은 1997년부터 시판되기 시작하였으며, Dragon Systems (L&H를 거쳐 ScanSoft에 인수됨)의 Naturally Speaking과 IBM의 ViaVoice 등이 대표적인 제품이다. NaturallySpeaking은 최고 음성입력 속도가 분당 160 단어로 대학 졸업자의 평균 타이핑 속도인 분당 50 단어보다 3배 이상 가능하며, 95% 이상의 단어 인식률을 나타내고 있다. 인식대상 어휘는 45,000 단어이며 총 23만 단어 사전으로부터 사전 권리도구를 이용하여 자유롭게 인식 대상 단어를 변경할 수 있으며, 새로운 단어의 추가도 가능하다. 작업 영역별로 사전을 구성하여 운영할 수 있으며, 미국영어 이외에도 영국영어, 불어, 독일어, 이태리어, 스페인어, 중국어 및 일본어를 지원한다. 이 소프트웨어의 주 고객으로 병원, 법률 회사, 장애인 및 일반 사용자로 고려하고 있다. ViaVoice는 분당 125-140 단어의 입력이 가능하다. 국내에서도 보이스텍에서 ByVoice라는 우리말 받아쓰기 제품을 출시하였다.

4.3 교육

음성 정보 기술의 발전과 더불어 언어학습에 이 기술을 이용하려는 다양한 시도가 진행되어 왔다. 크루츠와일(Kurzweil) 교육 시스템 사는 크루츠와일 3000이란 제품을 개발하여 4-6세의 어린 학생들이 스스로 영어를 읽고 쓸 수 있도록 교육 프로그램을 개발하였다. Colorado 대학 CSLR(Center of Spoken Language Research)의 Reading Tutor 프로젝트는 미국 초등학교를 대상으로 읽기 교육을 지도하는 컴퓨터 캐릭터 선생님을 만드는 것을 목표로 한다. 자연스럽게 얼굴을 마주보며 대화하는

애니메이션 캐릭터를 통해 학생들은 발음 이해 및 발성, 유창하게 읽고 이해하기 등을 배우고 훈련하며 학습 성과를 평가받을 수 있다. Reading Tutor는 음성인식 기술을 이용하여 학생들이 글을 읽을 때 각 단어가 정확하게 발음되었는지 확인하고 그 결과를 영상 피드백을 통해 실시간으로 제공한다. 그리고 읽은 내용의 이해를 확인하는 질문을 던진 다음 답을 인식하여 평가에 적용한다.

국내에서도 언어과학에서 Dr. Speaking 이라는 프로그램을 개발하여 판매 중인데, 이 제품은 음성인식 기술을 이용하여 사용자의 잘못된 영어 발음을 분석하여 교정정보를 제공하며, 특히 한국인이 영어를 발음할 때 자주 발생하는 습관적 오류를 데이터베이스화 하여 발음오류의 유형에 대한 정보도 함께 제공한다.

4.4 게임

포닉스(Fonix)사는 플레이스테이션 2 게임에 음성 인터페이스를 할 수 있는 팩키지를 개발하였다. 현재 이 음성명령어는 영어, 독일어, 프랑스어, 스페인어 및 일본어가 가능하며 이탈리아 및 한국어는 2004년 중순에 가능할 것으로 예상된다. 마이크로소프트는 음성인식 기술을 이용한 육성 시뮬레이션 게임인 NUDE@ (Natural Ultimate Digital Experiment @)를 공개했다. 이 게임은 사용자가 6개월 동안 음성입력 시스템을 통해 육성 대상인 캐릭터(PASS)를 키우는 것으로, 캐릭터가 성장하면 자신의 감정을 표현할 수 있게 되며, 사용자와의 대화도 가능해진다고 한다.

국내에서도 게임에 음성인식 기능을 첨가한 제품들을 출시한 바 있다. 보이스웨어에서 개발한 “보이스게임 디아블로2”는 “디아블로2” 게임의 각종 명령 키들을 음성으로 작동시킬 수 있도록 한 것으로서, 예를 들어 자신이 소유한 무기를 보려면 키보드에서 “I”키를 입력하는 대신에 “인벤토리”

라고 말하면 된다. 디앤 엠테크놀로지도 “스타크래프트” 게임을 음성으로 진행할 수 있는 “보이스 크래프트”를 개발하였으며, 키보드나 마우스로 공격 대상이나 원하는 유닛을 지정하면서 공격 아이템을 음성으로 바꿀 수 있어서 게임 속도 향상에 기여할 수 있다[9].

4.5 음성 인덱싱 및 검색

음성 인덱싱 및 검색이란 음성, 영상 등이 포함된 멀티미디어 정보를 내용 중심으로 검색하고자 할 경우에 사용되는 응용서비스다. 영상 정보를 이해하기가 현재 기술로 어려우므로 음성 정보를 중심으로 멀티미디어 내용을 인덱싱하고 데이터베이스화 하여 저장한다. 검색을 할 경우에는 텍스트 혹은 음성으로 검색한다. 음성 인덱싱 및 검색 기능을 포함하는 멀티미디어 정보검색 시스템의 개발 사례로는 Carnegie Mellon 대학의 Informedia 프로젝트, Cambridge 대학의 Video Mail Retrieval (VMR) 및 Multimedia Document Retrieval (MDR) 프로젝트, IBM의 CueVideo 시스템, 그리고 Compaq의 SpeechBot 시스템 등을 들 수 있다.

5. 결 론

음성정보처리기술을 이용한 응용서비스는 유/무선 통신망 응용, 내장형 단말 기반 응용 및 데스크탑 PC 기반 응용 등으로 나눌 수가 있다. 유/무선 통신망 응용 서비스로 음성인식 전화번호 안내 서비스, 음성다이얼 서비스, 음성인식 콜 센터 서비스 및 보이스포탈 서비스가 있다. 내장형 단말 기반 응용 서비스로 정보가전 응용 서비스, 텔레메틱 응용 서비스로 나눌 수 있다. 마지막으로 데스크탑 PC 기반 응용서비스로는 Command and Control, 발아쓰기, 교육, 게임 및 음성 인덱싱 및 검색 서비스로 나눌 수 있다.

음성정보처리 기술은 연구 기간과 다양한 응용

분야에 비해 히트작이라고 할 만한 서비스가 아직 출시되고 있지 않다. 최근에는 국내에서 진행되고 있는 9대 신성장 동력 기술 분야에서 음성 인터페이스는 중요한 인터페이스의 하나로 고려되고 있다. 이러한 현상임을 고려할 때 미래의 모든 제품에 음성 인터페이스는 현재의 GUI(graphic user interface)처럼 당연한 인터페이스로 여겨질 것이다. 하지만 이러한 목표를 위해서는 산, 학, 연의 공동연구가 필수적이며 아울러 정부의 연구지원도 꾸준히 지속되어야 할 것이다.

참고문헌

- [1] 구명완, “VoiceXML 기반 음성응용 시스템,” Telecommunications Review, 제 13권 5호 pp. 695-705, 2003년 10월
- [2] 김형순, “음성정보처리기술의 현황과 전망,” 대한전자공학회지, 제 30권 제 7호, pp. 700-708, 2003년 7월
- [3] 구명완, “유/무선 통신망에서의 음성인식 상용 서비스 개발 동향,” 대한전자공학회지, 제 30권 제 7호, pp. 719-724, 2003년 7월
- [4] 구명완, “최근의 음성인식 응용서비스,” 대한전자공학회지, pp. 587-595, 2001년 5월
- [5] 구명완, “음성인식기술의 현황과 전망”, 대한전자공학회지 제20권 제5호, pp.548-557 1993년 5월.
- [6] W. S. Meisel, Speech recognition update, 2001년 - 2004년 2월호
- [7] 구명완, 정영준 “통신망 기반 음성언어 응용기술”, 정보처리학회지, 본 특집호.
- [8] G. D. Fabrizio, et. al., AT&T HELP DESK, ICASSP 2002, pp. 2681-2684
- [9] 이용주외, 음성정보기술 로드맵, 대한음성학회, 2004, 발간 예정.

저자약력



구 명 완

1982년 연세대학교 전자공학과 (공학사)
1985년 한국과학기술원 전기 및 전자공학과(공학석사)
1991년 한국과학기술원 전기 및 전자공학과(공학박사)
1996-1997년 미국 벨연구소 방문연구원
2003년 서울대학교 KT-MBA 과정 수료
1985년-현재 KT 서비스개발연구소 음성언어연구팀장
(상무)
관심분야 : 음성정보처리 시스템개발 및 응용, 음성인식,
음성합성, 멀티모달, 음성통역
이 메 일 : mwkoo@kt.co.kr

김 재 인

1981년 고려대학교 전자공학과 (공학사)
1986년 고려대학교 전자공학과 (공학석사)
1996년 고려대학교 전자공학과 (공학박사)
1988년-현재 KT 서비스개발연구소 음성기반연구실
관심분야 : 음성정보처리 시스템개발 및 응용, 음성인식,
음성합성, 멀티모달
이 메 일 : jaeinkim@kt.co.kr