

통신서비스의 고도화를 위한 음성언어처리기술

李 勇 柱

韓國電子通信研究所 室長

I. 서론

정보통신기술의 개발 방향을 일본의 NTT는 VI & P, 즉 가시화(Visual), 지능화(Intelligent), 개인화(Personal)를 목표로 하고 있고, 우리 나라에서도 전자통신연구소에서 IMPH, 즉 지능화(Intelligent), 복합화(Multimedia), 개인화(Personal), 인간화(Human)를 통해 “누구든지 원하는 다양한 정보를, 자연스럽게 자유로운 형태로, 언제, 어디서나, 누구와도 편안하고 쾌적하게 주고 받을 수 있는 정보통신 서비스”의 개발을 목표로 하고 있다.^[1] 이런 관점에서 볼 때 음성언어정보는 전달해야 할 정보 그 자체이기도 하지만 그 처리 기술은 인간에게 정보의 전달, 가공, 처리 및 이용이 손쉽도록 도와주는 즉, 인간중심의 친숙한 서비스를 구현하는 중요한 기반기술의 하나이다.^[2] 이러한 기술을 응용한 새로운 통신서비스들은 통신기술과 컴퓨터기술, 통신망과 단말기술이 한데 어우러져야 비로소 달성될 수 있다. 요즘의 급속한 소자기술의 발달과 이에 따른 컴퓨터기술의 진보로 통신망과 단말간의 기능 분담 및 역할 배분에 대한 융통성이 커지고 있다. 통신회선 업체들은 가능한 많은 기능들을 통신망으로 끌어 들여 회선의 이용율을 높이려 하고 있고, 단말기기 제조업체들은 가급적이면 싼값에 고도의 기능을 단말기에 부여하여 상품의 가치를 높이려 하고 있다. 따라서 어떤 기술이 통신망에서만 가능한지 아니면 단말기에도 가능한지를 구분 짓는 것이 무의미해지는 경우도 있다. 본고에서는 단말기의 고기능화를 포함한 통신서비스의 고도화를 위한 음성 및 언어처리기술에 대해 논하려 한다. 음성 및 언어 관련 기술의 상세한 해설은 본지

93년 5월호의 “음성합성 및 인식” 특집에서 이미 많이 다루어졌으므로^[3] 여기서는 고도화의 관점에서 “IMPH”를 중심으로 음성 및 언어처리기술의 응용현황 및 전망에 대해서 정리해 보고자 한다.

II. 음성 및 언어처리기술의 현황

1. 음성처리기술

음성인식 및 합성, 즉 음성 입출력기술은 코스트 및 사이즈의 절감을 위해 앞으로 그 필요성이 더욱 커질 것으로 보인다. 반도체 소자기술의 발달로 컴퓨터는 더욱 싼값에 소형화 되어 가겠지만 키보드나 디스플레이는 사이즈나 코스트 양면에서 장애가 될 것이고, 이를 해결할 수 있는 유력한 수단의 하나가 음성입출력 기술이다. 음성입력은 키보드 사용이 미숙한 사람도 쉽게 사용할 수 있고, 정보 입력속도도 빠르며, 입력시 손은 다른 일을 할 수 있을 뿐만 아니라 키보드나 마우스등의 입력수단과 함께 사용할 수 있다는 이점이 있다. 아울러, 특별한 단말기가 필요 없이 일반전화기로 정보입력이 가능하고 시각장애자도 쉽게 사용할 수 있다는 나름대로의 강한 장점이 있다. 또한, 음성출력도 눈으로 다른 일을 하면서도 귀로 정보를 얻을 수 있고 전화기를 출력단말로 사용 가능하고 시각장애자도 사용 가능하다는 것이 장점이다. 이러한 이점 때문에 사용자의 편리성 향상을 위해 음성처리기술이 활용이 점차 확대 되어 가는 추세이다. 아직 고가의 기술이라는 문제는 있으나 앞으로는 저가 및 소형화 경쟁에 의해 결국 이러한 기술을 전면적으로 채용해야만 할 시점이 오게 될 것이고, 그때쯤 되면 반도체 칩과 마이크, 스피커, 전지로 구

성된 초소형의 컴퓨터 또는 단말기가 일반적으로 사용되고 지금과 같은 키보드나 디스플레이가 붙어있는 장치는 고가의 고급기로 남게 될 것으로 보는 사람도 많다.^[4] 또, 휴대전화 기능과 결합하므로써 인간끼리의 대화 뿐만 아니라 인간과 기계, 기계와 기계간의 대화가 장소에 상관없이 가능하게 될 것으로도 예상된다. 이를 가능하게 해주기 위한 요소기술로서 음성 입출력 기술의 현황은 아직도 여러 면에서 풀어야 할 과제가 많다.

1) 음성인식

음성인식을 가장 어렵게 만드는 요인은 우선 환경에 따른 음성패턴의 변동을 들 수 있다. 또한 어휘수가 많다는 점, 음성은 시간적으로 연속된 형태라는 점, 화자가 불특정이라는 점, 잡음의 영향을 받는다는 점, 실시간에 정확하게 인식해야 한다는 점 등을 들 수 있다. 현재 외국에서는 불특정화자의 단어인식은 수십 단어 수준, 특정화자의 경우는 수천 단어 수준의 시스템이 상품화 및 서비스화 되고 있으나^[5] 국내에서는 특정화자의 수십 단어급의 음성다이얼전 화기가 단말기로서 상품화 되어 있고, 아직까지는 통신망에서 제공하는 서비스는 없다.^[7] 음성인식의 응용 분야를 확대하기 위해서는 단어 단위의 발성에서 문장 단위의 자연스러운 발성, 소음중의 음성도 정확하게 인식해 내는 방식, 화자에 의존하지 않는 인식법의 연구가 필요하다. 최근에는 개인차 및 발성조건 등에 따른 변동을 통계적으로 모델화하는 화자적응법, 연속음성인식에 적합한 HMM(Hidden Markov Model)에 의한 인식, 훈련에 의해 식별기준을 스스로 만들어주는 신경망에 의한 인식, 음성을 음소레벨로 정확하게 인식하는 음향처리기술 등이 빠른 속도로 개발되고 있어 21세기 초에는 불특정화자의 경우 수천 단어, 특정화자의 경우 1만어 정도의 문장인식이 가능할 것으로 보고 있다.

2) 화자인식

크게는 음성인식의 범주에 속한다 할 수 있으나 발성의 내용을 인식하는 음성인식에 비해 화자인식은 발성자를 인식하는 기술이다. 현재의 기술수준은 등록된 발성자들의 음성중에 임의의 발성자의 음성이 포함되어 있는지를 확인해 주는 화자확인 기술의 일부가 출입자 통제시스템들에 응용되고 있으나 실용의 면에서는 화자인식기술에만 의존하기 보다는 얼굴, 지문, 암호번호 등을 동시에 이용하여 식별율을 높이고 있다. 아직까지는 발성시의 개인차, 주변환경이나

회선특성변동 등에 쉽게 영향을 받으므로 높은 수준의 인식율은 아직 확보되지 않고 있으나, 전화망을 통한 본인확인 등 화자식별 기술의 응용성은 크다.

3) 음성합성

음성합성의 방식에는 여러 가지가 있으나 임의의 내용을 합성하는 규칙합성, 특히 텍스트를 음성으로 변환해주는 텍스트-음성 변환기술은 응용면에서 매우 중요하다. 이러한 합성기술은 텍스트로부터 직접 음성을 합성해야 하므로 문장해석 및 의미 이해에 의해 억양, 길이, 세기, 끊어 읽기 등 소위 운율정보를 얼마나 자연스럽게 만들어내느냐 하는 것과 임의의 음성을 고품질로 만들어내기 위한 합성단위 및 방식의 양부에 따라 합성음의 품질이 결정된다. 각국에서 자국어 합성시스템들이 많이 개발되어 응용되고 있어서, 인식에 비해서는 그 응용의 속도가 빠른 편이나 현재까지는 명료성이 확보된 수준이고 자연성은 아직 충분하지 않다.^[8] 국내에서도 몇몇 시스템 들이 개발되어 상품화되어 있고 음질 개선을 위해 노력하고 있다.^[9, 10, 11] 외국의 경우는 발성속도 뿐만 아니라 음색의 가변을 거쳐 궁극적으로는 개인성의 부여, 다양한 감정표현의 연구가 진행되고 있다.^[12]

4) 음성부호화

좁은 대역폭으로 음성을 전송하거나 적은 메모리에 효율적으로 음성을 저장하기위한 연구는 광케이블로 대표되는 통신채널의 대형화로 일시 그 효용이 과소 평가되기도 하였으나 제한된 전파자원을 이용해야 하는 이동통신, 영상전화나 화상회의 등의 음성의 고품위전송, 음성의 비화통신, 합성 및 인식을 위한 효율적인 특징추출 및 저장방식으로써 최근에는 그 중요성이 더욱 커지고 있어서 각종표준화기구에서 표준 방식들이 정해지고 있다.

2. 언어처리기술

자연언어처리의 기술개발은 주로 기계번역 및 자연어 인터페이스를 중심으로 연구되어 왔다. 특히 번역 시스템 중에는 외국의 경우, 상품화되어 있는 것들도 있다.^[13, 14] 우리 나라의 경우에도 각 대학 및 연구소 들을 중심으로 개발한 시스템들이 하나 둘 선보이기 시작하고 있고, 일한 기계번역의 경우 성능이 제한적이지만 국내외 공동개발 또는 외국에서 개발된 것이 상품화 된 바 있다.^[15] 번역방식에는 대역방식과 중간언어방식(PIVOT)방식이 있는데, 대역방식은 소스언어(예를 들면 한국어)를 직접 대상언어(예를 들면 영

어)로 변환하는 방식이며 중간언어방식은 중간언어라 부르는 비교적 자연언어와는 독립된 계산기의 내부표현을 통해 대상언어로 변환하는 방식이다. 중간언어방식은 다국어번역에는 유리한 방법이나 자연어로부터 중간언어로 변환하거나 그 반대의 경우의 규칙을 만들기가 어렵고 따라서 중간언어의 설계자체가 어렵다. 자연언어처리의 가장 큰 문제는 문맥의 이해이다. 소위 행간을 읽어야 하는 기술로써 고도의 지식처리가 요구된다. 각국에서 이를 위해 스크립트모델, 디폴트추론, 상황의머론 등 문맥을 이루는 의미적, 논리적, 인과적인 정합성을 형식추론과 그 모델이론 기법에 의해 구체화하려 하고 있으나 아직은 초기단계라 할 수 있다.

Ⅲ. 지능화와 음성언어처리기술

지능화된 서비스란 융통성이 있고 영리한 통신망(예를 들어 지능망)을 통해 고부가가치의 정보를 효율적으로 제공해주는 것이라 할 수 있다.

음성언어정보처리기술을 이용하여 가능한 몇가지 서비스 예를 보기로 한다.

1. 번역, 요약, 지능적 정보관리

1) 번역통신 서비스^[6]

번역통신은 통신에서 거리와 시간의 벽 뿐만 아니라 언어의 벽도 극복해 주는 서비스이며 이 서비스의 중심기술은 기계번역기술이다. 현재의 기계번역기술은 입력된 원문을 직접 고품질의 외국어로 번역해 주는 수준에는 도달하지 못하였고 현 수준에서의 실용화를 위해서는 인간이 원문을 미리 기계번역이 쉽도록 문장을 다시 써 주는 전편집 과정과 기계번역 후에 손으로 수정해주는 후편집 과정이 필요하다. 현재 수준으로 가능한 번역통신서비스로 외국에서 검토되고 있는 것으로는 DB검색용 단문 번역 및 정보의 초역통신(수정 없는 초벌번역), 입력문 전편집에 의한 텍스트메일의 번역통신 등이 있다. 이 정도의 수준은 국내에서도 실현이 그다지 어렵지 않을 것으로 보인다.

2) 자동통역전화^[7]

음성의 인식과 합성 그리고 언어간의 번역기술이 통합되면 음성번역(speech translation)이 가능하고

전화망 상에서 이를 양방향으로 구성한 것이 자동통역전화이다. 이러한 시스템은 언어의 장벽을 극복해 주는 지능적인 통신 서비스를 가능하게 하는 것으로, 음성언어정보처리연구의 궁극적인 목표의 하나라 할 수 있다. 이에 대한 연구는 미국, 영국, 독일, 일본 등에서 연구가 활발하며 우리 나라도 1991년부터 ETRI가 한국통신과 함께 관련연구를 시작하였다.^[18]
^[9] 현재는 자동통역전화의 요소기술개발 및 서비스의 가능성타진 단계이며 2000년대의 실용화를 목표로 하고 있다. 개발 과정중의 연구결과들은 번역통신, 음성언어 이해에 의한 정보검색 등 관련 서비스 구현에 파급효과가 크다.

3) 문서요약 서비스^[20]

전자메일의 내용을 요약하거나 원하는 내용을 선택하는데는 요약, 선택 및 지능적인 검색기능이 필요하다. 일반적으로 문서를 이해하고 이를 요약하는데는 다음과 같은 과정이 필요하다.(그림1) 입력된 메시지의 내용을 문법규칙에 따라 구문해석하고 문법규칙과 대상영역의 상식을 기초로 단어간 또는 문장간의 의미관계를 추출하고 메시지 작성자의 의도를 이해한다. 이러한 해석 결과를 토대로 대상영역에서 중요하다고 생각되는 개념을 추출하고 문법규칙에 따라 그 개념을 연결하여 문장을 생성한다. 외국의 예를 보면 이러한 기술은 대상영역을 한정시켰을 경우 어느 정도 가능성을 보이고 있다.

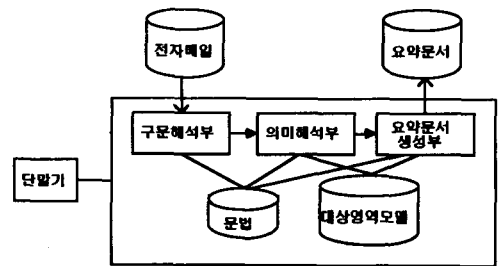


그림 1. 문서요약시스템의 개념도([20] 에서 인용)

4) 지능적 정보검색 서비스

지정된 특정한 키워드에 매칭하는 문서를 선택해주는 것을 말한다. 먼저 그림 2와 같이 검색조건을 음성인식 및 문자인식기술에 의해 자연언어 형태로 입력한다. 입력된 검색조건에서 키워드를 추출하고 검색의도를 이해한다. 효율적인 검색식을 생성하고

선택트리 등을 제시하여 상세한 검색요구를 상세히 파악한다. 검색에 해당하는 문서를 추출한다. 국내에서도 학계에서 이에 관한 연구가 수행되고 있다.^[21]

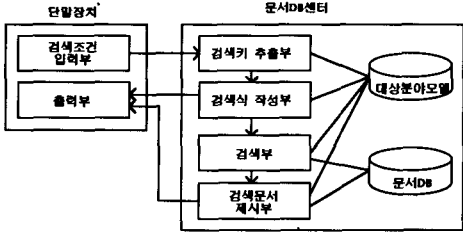


그림 2. 지능적 정보검색 시스템 개념도 ([22] 에서 인용)

2. 미디어간의 변환서비스

통신망 상에 접속된 다양한 단말기 사이에 미디어에 국한하지 않는 자유로운 통신이 가능하도록 하기 위해서는 미디어변환기능의 실현이 필요하다. 특히 음성언어를 중심으로 한 미디어변환은 가장 보편적인 단말기인 전화기와 타 매체간의 통신을 가능하게 해주는 기능이 된다.

1) 텍스트-음성변환

데이터터미날로 액세스하도록 만들어진 텍스트 형태의 정보DB를 일반전화기를 통해 액세스할 수 있게 하려면 텍스트를 읽어서 음성으로 출력해주는 텍스트 음성 변환기능이 필요하다. 이 기술은 음성합성의 가장 진전된 형태로 각국에서 일기예보, 뉴스 등의 서비스가 시도되고 있다. 우리 나라에서도 ETRI에서 통신처리시스템의 서브기능으로 적용을 시도하고 있으며^[23] 현재는 음질의 개선이 진행되고 있다.^[23] 이러한 서비스는 시각장애자의 정보검색에도 매우 유용한 수단이 된다. 또한 전화기에 의한 전자메일의 수신에도 응용될 수 있다. 약간의 코드변환 및 프로토콜 변환 기술이 여기에 가미된다면 팩시밀리로 수신할 텍스트데이터를 전화기로 수신할 수도 있을 것이다.

2) 음성 -텍스트 변환

기술적으로는 고도의 음성인식기술이 필요하며 이러한 기능은 전화기로부터 데이터단말이나 팩시밀리로 통신할 수 있게 한다. 이는 음성타자기와 같이 일종의 Dictation 시스템을 구현하는 방법인데 음성의 속성상 상위정보를 이용하다더라도 한자 한자 정확하

게 인식하는데는 아직 한계가 있다. 음성이해의 차원에서 개략적으로 의도를 이해하는 제한적인 형태로 응용될 가능성이 많으며 내용을 요약해서 출력해 주는 자동 속기의 형태도 가능성이 있다.

3. 전자비서

마치 통신망 또는 단말기 내에 비서가 있는 것과 같은 기능을 수행해 주는 서비스를 전자비서(Electronic secretary) 또는 망비서라 부른다. 전자비서의 기능으로 여러 가지를 생각해 볼 수 있지만 주로 스케줄링, 지능형 음성다이얼링(예를 들면 어느 부의 무슨 과장, 어떤 부서의 무슨팀 전원에게 전화 연결), 정보조회(예를 들면 인사기록 조회), 예약 등 일상의 비서가 지적인 활동을 대행하는 기능에 해당한다. 이러한 기능은 단말기나 구내교환기 차원에서 구현할 수도 있고 지능망의 IP (Intelligent Pheriperal)를 통해 공중망의 교환기에서도 구현될 수 있을 것이다.^[24] 여기에는 음성인식기술, 음성메일기술, 교환기술, 지능형 정보검색기술이 필요하다. 음성인식 및 정보검색에 얼마나 고도의 지식처리가 가능한가에 따라 효용성이 판가름나게 될 것이다.

IV. 복합화와 음성언어처리기술

매체의 다양화를 통해 자연스럽고 자유로운 형태로 정보를 주고받을 수 있는 복합화 기술은 보고, 듣고, 느끼는 서비스의 구현을 위한 기술이다.

여기서는 특히 음성언어정보의 전달을 중심으로 전달효율의 향상을 위한 미디어간의 융합에 관하여 살펴보기로 한다.

1. 미디어관련 용어의 정의

먼저 미디어와 관련된 몇가지 용어의 정의를 참고하기로 한다.^[25]

1) 멀티미디어(multimedia)

미디어는 정보의 전달매체이다. 예를 들면 인쇄용지, 비디오, 오디오 및 비트맵디스플레이 등은 미디어이다. 요즘의 컴퓨터업계에서는 단순히 여러 개의 출력미디어를 가지고 있는 경우에도 멀티미디어컴퓨터라 부르고 있다. 신문지에는 텍스트, 사진 그림이 섞여 있어도 멀티미디어라고 부르지 않으므로 화면상

의 텍스트와 이미지를 멀티미디어라 부르는 것은 적당하지 않다. 입력의 경우도 키보드, 마우스, 마이크, 카메라 및 그 밖의 다른 미디어를 복수로 사용할 경우에 멀티미디어라 부른다.

2) 멀티모달(multimodal)

모달리티(modality)란 정보가 지각되는 감각을 나타낸다. 따라서 TV는 시각 및 청각정보를 가지고 시각과 청각을 동시에 사용하므로 멀티모달이라 할 수 있지만 신문은 모두 시각정보이므로 멀티모달이 아니다.

3) 멀티채널(Multichannel)

멀티채널이란 하나의 모달리티라도 동시에 복수의 정보를 액세스할 수 있는 것을 의미한다. 예를 들어 2대의 TV를 동시에 볼 경우에 2채널이라 하며, 멀티윈도우도 멀티채널을 실현하는 예이다. 하나의 디스플레이상에서 양손에 의한 키보드 입력도 멀티채널의 예이다.

2. 음성언어정보의 멀티모달 인터페이스^[26]

통신서비스의 다양화 및 효율화를 위해서는 단순히 복수미디어의 사용(멀티미디어)뿐만 아니라 이들 미디어의 융합에 의한 상호보완을 통해 정보전달의 효율성을 극대화시키는 문제는 매우 중요하다. 특히 인간의 언어지식을 이용한 음성정보의 전달, 가공 처리는 그 표현형태에 따른 많은 애매성을 포함하고 있어서, 타 미디어정보를 동시에 활용함으로써 이를 해결 하고자 하는 시도들이 점차 활발해 지고 있으며 이른바 음성언어정보의 멀티모달 인터페이스라 할 수 있다.

그 예를 몇가지 들어 보기로 한다.

1) Put that There^[26]

MIT가 1980년에 개발된 시스템으로 음성과 제스처어를 이용하여 휴먼인터페이스를 대폭적으로 향상시킨 예로 알려져있다. 그 구성을 그림3에 보였다. 대화는 미디어룸의 대형스크린을 상대로 하고 지시장치는 손에 장치한 위치센서를 이용하며, 키보드 대신 음성인식장치를 이용한다. 사용자가 화면상의 어떤 대상물을 가리키며 "Put that"이라 발성하면 시스템은 사용자가 무엇에 대해 말하고 있는가를 안다. 그리고 사용자가 다른 한점을 가리키며 "There"라고 발성하면 시스템은 이를 알아듣고 대상물을 지시한 장소로 이동시킨다. 즉 "that" 이나 "there" 같은 지시대명사나 장소를 나타내는 부사의 애매성을 제스처어에 의한 지시를 통해 해결할 수 있다.

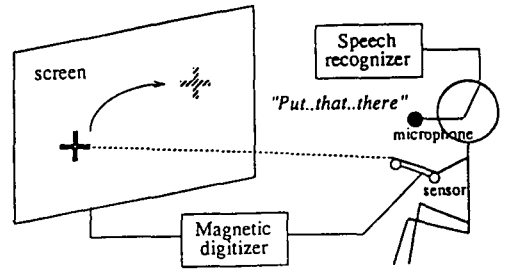


그림 3. "Put that There" 시스템 ([25] 으로부터 재인용)

2) 음성대화시스템 TOSBURG^[26]

일본의 토시바가 시험적으로 제작한 불특정화자를 대상으로 한 실시간 음성대화시스템이다. TOSBURG (Task-Oriented dialogue system Based on speech Understanding and Response Generation) 는 햄버거집에서의 주문시스템을 상정하여 불특정의 사용자가 일상의 언어로 계산기와 자연스럽게 대화하는 것을 목표로 하였다. 그림 4에 그 구성을 보인다. 키워드 검출부는 발성내용의 이해에 필요한 키워드를 연속적으로 검출해 낸다. 음성이해부는 음성을 한자 한자 인식하지 않고 키워드를 중심으로 발성내용 및 의도를 이해해서 의미표현을 얻어낸다. 또 대화제어부는 대화상황과 대화이력정보를 이용하여 사용자의 발성이해와 적절하게 응답할 의미표현을 생성하여 사용자 주도형의 대화를 실현한다. 응답생성부는 응답의미표현으로부터 합성음성에 의한 응답문, 그래픽에 의한 멀티모달 응답을 생성 출력한다. 합성음성으로는 대화상황에 따른 즉각적인 메시지를 청각적으로 전하고 동시에 응답문을 시각적으로 표시함으로써 합성음성에 의한 일과성의 결점을 보완하고 있다. 또 확인을 위한 응답음성과 동시에 주문내용도 아이콘에 의해 화면상에 시각적으로 표시하고 있다. 또 대화 상대자(가상)의 표정을 화면에 애니메이션에 의해 표시하여 사용자가 안심하고 대화할 수 있도록 유도하고 있다. 또 바닥에 플로어 매트를 설치하여 대화의 시작 및 인물의 겹침에 이용한다.

이밖에도 음성관련의 멀티모달 인터페이스로 음성 미디어와 시각미디어에 의한 독순(lip reading)의 통합연구가 있다. 이러한 연구는 입모양에 의한 자모음 인식, 세그먼트이션 성능의 향상, 잠음환경에서의 음성인식에 응용이 기대되고 있다. 또한 눈의 움직임(시선)과 음성인식의 통합연구도 시도되고 있으며

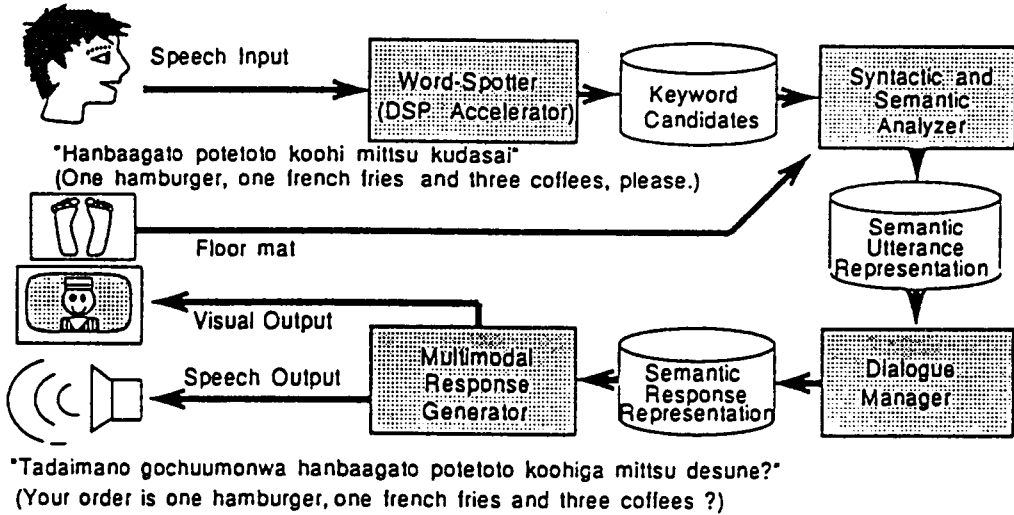


그림 4. 음성대화시스템 "TOSBURG"의 구성 ([26] 으로부터 인용)

[27], 차세대 번역시스템으로 멀티모달음성번역 통신이 국내외에서 제안되고 있다. [28]

V. 개인화와 음성언어처리기술

통신의 개인화는 개인휴대 이동통신을 먼저 생각할 수 있으나 좀더 넓게 생각하면 각 개인의 상황에 맞는 통신서비스의 제공이라고 볼 수 있을 것이다. 개인의 상황을 인지하고 이에 적합하도록 한다는 면에서는 지능화, 인간화와 깊은 관련이 있다. 음성에 의한 개인통신에서 필연적으로 대두되는 것은 프라이버시의 보호 등 보안성의 확보문제이다. 이를 위해서는 음성부호화에 의한 비화기술, 본인임을 확인하기 위한 화자인식기술이 응용될 수 있다. 또한 앞에서 언급한 바와 같이 이동형 단말기의 소형화를 위해서는 통신망내에서의 음성다이얼링 기능을 통해 단말기의 기계적인 부분을 대폭 줄여나가는 데도 음성합성 및 인식 등 음성입출력기술이 유용하다.

VI. 인간화와 음성 및 언어정보처리

인간과 기계(컴퓨터 또는 통신기기)간의 원활한 통

신수단의 확보는 통신의 효율성을 높이는데 있어서 매우 중요한 문제일 뿐만 아니라 인간중심의 통신서비스 구현에도 필수적이다. 이 문제의 해결을 위해서는 HCI(Human Computer Interaction)이라는 관점에서 사용의 중심이 되는 인간의 행동특성을 검토하여, 인간의 지능 및 감각기능에 더 부합되고 사용이 용이한 방법들이 계속 개선 연구되어야 한다. 즉 인감의 5감의 기능을 최대한 활용할 수 있는 방향의 연구가 필요할 것이다. 특히 음성언어를 포함한 청각 정보는 사용의 용이성 면에서 볼 때 그 중요성은 매우 크다. 생각을 전달하는 방법에는 말, 몸짓, 표정 등 다양하지만 가장 구체적이고 효율적이며 인간만이 그 기능성을 최대한 누리고 있는 '말'은, 생각의 내용을 언어적 체계로 구성하여 청각정보인 음성이나 시각정보인 글자를 통해 전달된다. 이 중에서 음성은 단순한 소리가 아닌 고도의 언어적, 인지심리학적 정보들이 사람의 목소리에 실린 것으로, 이러한 기능을 기계와의 대화에 얼마나 잘 활용할 것인가 하는 것이 HCI연구중의 하나의 커다란 과제가 될 것이다. [29]

현재의 통신에서는 통신기기의 기능 고도화에 따라 사용경험이 있거나 숙련된 사람을 제외하고는 완벽하게 사용하는데 어려움이 있고, 사용한다 하여도 기능의 일부에 만족하는 경향이 있다. 또 통신시스템 위주의 사용법에 사용자가 억지로 적응해 가야만 한다. 인간위주의 통신서비스는 이러한 면을 개선하기위한 여러 조치를 필요로 하고 있다. 음성 및 언어처리기

술은 인간에게 있어서 가장 자연스러운 의사전달 수단인 음성언어에 의해 기기의 조작의 편리성을 향상시킬 수 있다. 조작을 자연언어로 지시하고 상황에 따라 조작설명을 해주는 help기능의 실현에는 단말기 레벨에서 실현이 용이한 단어레벨의 음성인식이 우선적으로 응용되고 있고, 아직 초보 단계이나 자연언어의 이해, 대화 및 담화이해 기술이 활발하게 연구되고 있어서 이러한 기능의 고도화 가능성을 높혀주고 있다.

장애인들의 의사전달 수단을 참고하면 정상인의 정보조작, 처리, 표현의 간편성 신속성을 높이는데 좋은 아이디어를 제공해 줄 수도 있다. 장애인들의 의사전달 및 통신을 위해서는 장애형태에 따라 음성, 문자, 수화, 점자, 속기 등이 쓰일 수 있다. 정상인과 장애인간의 언어정보전달은 음성 또는 문자에 의한 전달이 자연스러우나 장애의 형태에 따라서는 음성 및 문자언어를 사용할 수 없는 경우도 많다. 즉 시각장애인들에게는 시각언어인 문자정보를 그대로 전달하기 어렵고 음성으로 읽어 주거나 점자로 변환하여 촉각을 통해서 전달할 수 밖에 없다. 또 청각장애인의 경우는 음성언어 그자체로는 무의미하고 문자 또는 수화에 의한 의사표현방법이 주로 쓰이게 된다. 발성장애자는 글자나 속기, 심볼 키의 입력에 의해 음성을 합성해주는 음성 발생기가 필요하다. 이와 같이 음성, 문자, 속기, 점자, 수화 등 여러 가지 미디어 간을 그림 5와 같이 상호변환하는 기술을 통해 장애인의 통신을 지원할 수 있을 것이다.

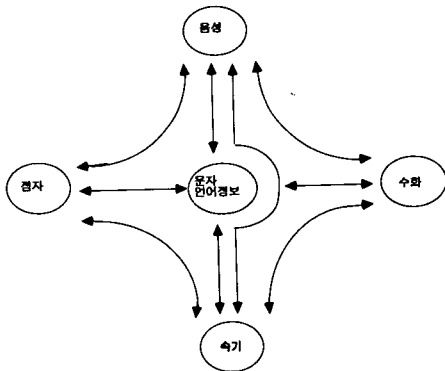


그림 5. 장애인의 의사전달 미디어간의 변환

여기에는 언어이해기술, 음성합성 및 인식기술,

문자인식기술, 속기기호 인식 및 합성기술, 수화영상 인식 및 합성기술, 입모양 인식 및 합성기술이 필요하다. 각 미디어간의 변환에 필요한 요소기술 및 응용시스템의 예를 그림 6에 보였다.

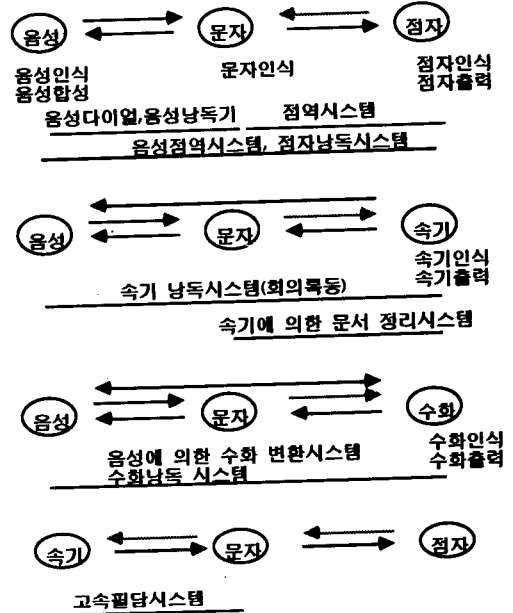


그림 6. 장애인을 위한 미디어변환 및 응용시스템 (← 표시는 기술적으로 가능하나 응용성희박)

VII. 결론

지금까지 통신 서비스의 고도화를 위한 음성 및 언어정보처리기술의 응용에 대해서 IMPH를 중심으로 정리하여 보았다. 이러한 서비스들은 IMPH의 어느 한 항목에 국한하지 않고 복합적으로 구현되는 것이 대부분이다. 따라서 각 항목별 분류는 어떤 면에서는 무리한 일인지도 모른다. 중요한 것은 통신기술의 개발방향으로써의 IMPH, 그리고 이를 실현 시키기 위해 필요한 요소기술으로써의 음성 및 언어처리기술의 중요성의 인식이다. 본고를 통해 응용 예를 간단히 소개하므로써 이 분야 기술개발을 위한 작은 자극이 되었으면 하는 바램이다.

參 考 文 獻

- [1] 양승택, “정보통신 기술개발 추진방향” 한국정보과학회 1993년도 춘계학술발표회 특강 요약집, 1993년 4월.
- [2] 이용주, 강철희, “음성, 언어 및 음향정보처리 연구현황” 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, 1993년 4월.
- [3] 대한전자과학회, 한국어 음성인식 및 합성기술 특집 전자공학회지 제20권 제5호, 1993년 5월.
- [4] 한국전자통신연구소 역, 2005년의 정보통신기술 한국전자통신연구소, 1991년 12월.
- [5] 이용주, 김경태, “음성정보처리 장치의 동향” 전자통신동향 분석, 1988년 3월.
- [6] 이용주, 김경태, “음성이해 연구의 동향” 전자통신 9권 1호, 1987년 3월)
- [7] 김순협, “국내의 음성인식기술의 연구동향 및 전망” 제9회 음성통신 및 신호처리워크샵 논문집, 한국음향학회, 1992년 8월.
- [8] 이용주, “음성합성 기술의 현황과 과제” Korea-Japan Joint Workshop on Advanced Technology of Speech Recognition and Synthesis, Jul. 1991.
- [9] 이양희 음성합성기술의 현황 및 전망 제9회 음성통신 및 신호처리워크샵 논문집, 한국음향학회, 1992년 8월.
- [10] J.C.Lee, Y.J.Lee,H.Hahn,E.B.Kim, C.J.Kim, K.T.Kim“Speech Synthesis Using Demisyllables for Korean: A Preliminary System” ICSLP 90, Kobe, Japan, Nov. 1990.
- [11] 최운천 외, 고품질의 한국어 문장음성변환시스템: 글소리Ⅱ 제9회 음성통신 및 신호처리워크샵 논문집, 한국음향학회, 1992년 8월.
- [12] 이용주, 이정철, “음성 응답 기술 동향” 전자통신동향 분석, 한국전자통신연구소, 1989년 4월.
- [13] Engeliem, Mcbryde, Natural language markets: Commercial strategiesOvum Ltd, London, 1991.
- [14] Ehara, “Present state of commercial machine translation system in Japan” Seminar of IEICE, Oct. 1989.
- [15] 박동인, “기계번역 어디까지 왔나” 한국어정보처리연구회 소식지, 제1권 제1호, 1992년 4월
- [16] Sakama, sakai, “Current Status and Future Applications of natural language processing” NTT R&D vol.40 no.7, 1991
- [17] 이용주, “자동통역전화의 기술현황 및 과제” 대한전자공학회지 제20권 제5호, 1993년 5월.
- [18] 한국전자통신연구소, 자동통역전화를 위한 요소기술 개발(I)(II)한국전자통신연구소 연구보고서, 1991년 12월, 1992년 12월.
- [19] 이종락, “한국통신의 자동통역연구현황” 제9회 음성통신 및 신호처리 워크샵 논문집, 한국음향학회, 1992년 8월.
- [20] 전기통신에의 인공지능 기술도입에 관한 조사 연구회 편, 인공지능에 의한 전기통신의 고도화전략, 전기통신정책총합연구소, 일본, 1990년 4월.
- [21] 한국과학기술원, 지능형 정보검색에 관한 연구 (I)(II) 연구보고서, 한국통신, 1991년, 1992년.
- [22] 최훈, 김동원, 이용주, 김상중, 이경준, 이성경, “통신처리시스템 개요”한국통신학회 1990년도 하계종합 학술 발표회 논문집, 1990년 8월.
- [23] 이승훈, “Audiotex 음질개선 연구” 제1회 ETRI 음성, 언어 및 음향정보처리 워크샵 논문집, 1993년 4월.
- [24] 최고봉,이영호 외, “고도지능망을 위한 지능형 정보제공 시스템”전자공학회지 제20권 제2호, 1993년 2월.
- [25] Takebayashi, “Human-computer dialogue using multimedia undestanding and synthesis functions” Technical report of IEICE, SP 92-37, 1992.
- [26] Takebayashi et al, “A real-time speech dialogue system using spontaneous speech understanding” ICSLP 92 Proceedings, Oct. 1992.
- [27] 이용주, “일본 ATR연구소의 첨단연구 소개”

한국전자통신연구소 전자통신동향분석, 1993년 1월.

[28] Kurematsu, "Perspective view of multimedia cross language communication" Proc. International workshop on Advanced communications and applica-

tions for high speed networks, Germany, March 1992.

[29] 이용주, "HCI를 위한 음성, 언어 및 음향정보처리" HCI '93 논문집, 한국정보과학회, 1993년 2월. (㉠)

筆者紹介



李 勇 柱

1954年 1月 17日生

1976年 2月 고려대학교 전자공학과(학사)

1987年 8月 고려대학교 대학원 전자공학과(석사)

1992年 8月 고려대학교 대학원 전자공학과(박사)

1976年 3月 ~ 1980年 7月 공군통신장교 근무

1985年 7月 ~ 1986年 2月 일본 토호쿠대학 응용정보학연구소(연구원)

1980年 8月 ~ 현재 한국전자통신연구소 자동통역연구실실장(책임연구원)

주관심분야 : 음성, 언어 및 음향정보처리, HCI 등