음 성 인 식

김 형 순 (부산대학교 전자공학과)

요 약

반도체 및 컴퓨터 응용기술 등의 급속한 발전과 더불어 인간의 가장 자연스러운 의사전달수단인 음성을 인간과 기계 사이의 의사소통의 매개체로 사용하기 위한 음성인식기술에 관한 연구가 활발히 진행되고 있으며, 일부 상품화된 음성인식 시스템들이 다양한 응용분야에 걸쳐 등장하고 있다. 본 고에서는 지금까지 개발되어온 음성인식기술의 방법론 및 기술적으로 해결해야 할 과제들을 살펴보고, 음성인식기술에 관한 국내외 연구동향과 음성인식의 응용분야, 그리고 앞으로의 전망에 대해 논의한다. 이 과정에서 음성학 및 언어학적 지식이 음성인식에서 차지하는 중요성에 대해서도 일부 언급한다.

1. 서 론

음성이 인간의 가장 자연스러운 의사전달 수단이라는 데에는 이론의 여지가 없을 것이다. 이 음성을 이용하여 각종 기계나 도구들을 쉽게 조작하고자 하는 것은 인류의 오랜 꿈 중의 하나였다. 최근 컴퓨터 및 전자공학의 급속한 발전은 인간의 음성을 알아듣고(음성인식) 인간에게 음성으로 응답하는(음성합성) 기능을 가지는 기계의 출현이 꿈이 아닌 현실로 다가오고 있음을 보여주고 있으며, 일부 상품화된 제품으로까지 등장하기에 이르렀다. 이러한 음성인식기술은 키보드와 같은 번거로운 수단을 사용하지 않고도 컴퓨터를 조작할 수 있게 해 주고, 멀리 떨어진 곳에서 전화를 이용하여 컴퓨터에 수록된 정보를 꺼낸다든지 또는 각종 작업지시를 명령한다든지 하는 일들을 가능케 해준다. 그러나, 아직까지는 공상과학영화에 나오는 것처럼 사람과 유창하게 대화를 나눌 수 있는 컴퓨터가 개발되지 못하고 있으며, 앞으로도 가까운 장래에 현실화될 것 같지는 않다.

음성인식의 목표가 잡음이 있는 실제적인 환경에서 임의의 화자가 어휘에 제한없이 자연스럽게 발음한 연속음성을 실시간에 인식 및 이해하는 것이라고 할 때, 음성인식이 지니는 기술적인 어려움들은 크게 다음의 다섯 가지 항목으로 나누어 볼 수 있다. 그 중 첫째가 화자 독립성(speaker independence)에 관한 것으로서, 특정한 한 사람의 음성을 인식하는 것에 비해 연령과 성별, 그리고 방언이 다른 임의의 사람의 음성을 인식하는 것이 훨씬 어렵다. 두번째는 발음방법 또는 속도와 관계되는 것으로서, 각 단어와 단어를

또박또박 띄어 발음하는 것보다 자연스럽게 연결시켜 발음하게 되면 단어들 사이의 상호 조음현상으로 인해 인식이 어려워지며, 이 현상은 발음속도가 빨라질수록 심각해진다. 세 번째로 인식대상어휘의 난이도를 들 수 있으며, 일반적으로 인식하고자 하는 어휘 규모가 커질수록 혼동되기 쉬운 단어나 구절들이 많아지고 따라서 오인식의 가능성도 커진다. 물론 동일한 어휘 수라고 할지라도 단어들의 음성학적 유사성에 따라 인식 난이도는 다르게 된다. 네번째는 언어의 문법구조 및 주제와 관련된 것으로서, 사람들의 일상적인 언어는 컴퓨터 언어와 같이 문법구조에 인위적인 제약을 둔 경우와는 달리 컴퓨터로 해득하기가 쉽지 않으며, 특히 회화체 음성언어는 더욱 인식하기 어렵다. 그리고, 임의의 주제를 대상으로 할 경우 특정한 주제로 발언내용을 제한할 때보다 음성인식의 난이도가 높아진다. 마지막으로 음성통신 환경도 음성인식에 중요한 영향을 끼칠 수 있는데, 예를 들어 전화음성과 같이 주파수 대역폭이 제한되거나, 배경잡음이 있는 경우 동일한 음성인식시스템이라도 인식율이 떨어지게 된다.

그러나, 특정화자의 음성만을 대상으로 하여 수십 단어 정도의 제한된 어휘를 또박또박 띄어 발음한 음성을 인식하는 것과 같이 제한된 목표를 설정할 경우, 매우 우수한 성능을 나타내는 음성인식시스템들이 개발되어 상품화된 제품으로 나와 있다. 그리고, 앞서 언급한 문제점들을 해결하기 위하여 많은 연구가 진행됨에 따라, 부분적으로나마 이러한 제한성을 극복하는 결과들이 보고되고 있다. 예를 들어, 방음실에서 발음한 열 개의 숫자음 인식의 경우 임의의 화자에 대해서도 99.7%의 인식율을 얻고 있으며, 1000단어 어휘의 연속음성인식에서의 단어 인식율도 95%를 상회하는 시스템도 개발되고 있다. 본 고에서는 현재까지 개발되어온 음성인식기술의 방법론을 살펴보고, 국내외 연구동향과 응용분야, 그리고 앞으로의 전망에 대해 논의하고자 한다.

2. 음성인식기술

지금까지 다양한 형태로 개발되고 발전되어온 음성인식기술을 짧은 지면을 통해 다 소개할 수는 없으므로, 본 고에서는 음성인식을 위한 공학적인 방법론에 대해 개괄적으로 기술하도록 한다. 그림 1에 일반적인 음성인식 시스템의 구성도가 나타나 있으며, 그 기본적인 동작은 다음과 같다. 마이크를 통해 입력된 음성은 컴퓨터에서 처리할 수 있는 디지탈 신호로 변환되어 음성인식시스템으로 들어오게 되며, 음성인식의 첫 단계인 음성특징분석을 통해 음성학적 특징을 잘 표현해 줄 수 있는 음성특징계수들을 추출하게 된다. 추출된 음성특징계수들은 패턴인식과정으로 넘겨져서 미리 저장된 단어 또는 음소들의 모델과 비교하게 되며, 그 결과는 일련의 후보단어 또는 후보음소들의 형태로 언어처리과정에 전달된다. 언어처리과정에서는 후보단어 또는 후보음소들의 정보를 토대로 하여, 인식대상어휘 및 문법구조, 그리고 특정 주제에의 부합 여부를 판단하여 최종 인식된

문장을 출력시키게 된다. 경우에 따라서는 언어처리과정에서 새로운 후보단어나 후보음소를 추정하여 패턴인식과정에 전달하여 이를 확인해보도록 지시하기도 한다. 그림 1과 같은 구성도가 모든 음성인식방식에 적용된다고는 말할 수 없지만, 대부분의 음성인식시스템이 이러한 구성도를 통해 설명이 가능하므로 이하에서는 이 구성도에서의 각각의 부분에 대해 보다 상세히 설명하기로 한다.

가. 음성특징분석

음성신호에는 누가 어떤 환경에서 어떤 말투로 어떤 내용을 이야기했는가 하는 정보가 모두 포함되어 있다. 따라서 음성신호의 모든 내용이 음성인식과 관련되는 것은 아니며, 음성인식을 위해서는 입력된 음성신호로부터 말한 사람이 누구인지, 어떤 배경잡음이 있는 상황인지, 그리고 어떤 감정적인 상태에서 말했는지 등에 상관없이, 말하고자 하는 내용이 무엇인지를 파악해낼 수 있는 음성학적인 특징들을 추출하여야 한다. 시간영역에서 관찰되는 음성신호의 파형 그 자체만로는 이러한 음성특징이 명료하게 드러나지 않는 경우가 많으므로, 일반적으로 음성신호의 시간영역 및 주파수영역에서 음성학적 특징을 잘 표현해 주는 특징계수들을 찾아내게 된다. 이러한 음성특징분석 방법으로는 음성발생기관의 모델에 근거를 둔 방법과 음성청취기관의 모델에 근거를 둔 방법의 두 가지 접근방안이 주로 사용된다.

음성발생기관의 모델에 근거를 둔 방법의 대표적인 예로는 선형예측부호화(Linear Predictive Coding(LPC)) 방법을 들 수 있다. 이 방법은 인접한 음성신호 사이의 연관관계를 선형적으로 표현한 것으로서, 음성발생기관 중 성도(vocal tract)의 모양을 디지탈 필터로 모델링한 효과를 가지기 때문에 음성학적으로 의미있는 특징, 특히 포먼트(formant) 주파수에 관한 정보를 잘 나타낸다. 그리고, 음성청취기관의 모델로는 달팽이관 내부의 기저막(basilar membrane)에서의 주파수 선택특성에 기초한 대역 필터군(bandpass filter bank)이 사용되며, 경우에 따라서는 청각기관에서의 비선형적 특성들을 포함시키기도 한다.

이들 음성특징분석 방법들은 실제 음성인식시스템들에 성공적으로 사용되고는 있으나, 매우 단순화된 모델에 근거를 두고 있기 때문에 여러 가지 한계점들을 드러내기도 한다. 예를 들어 LPC 방법의 경우 모델의 특성으로 인하여 비음의 묘사에 한계가 있는 것으로 알려져 있으며, 대역 필터군의 경우에도 계산소요시간을 고려하여 실제 청각기관과는 달리 단지 수십 개의 필터 만으로 특징분석을 수행하는 경우가 대부분이다. 뿐만 아니라, 달팽이관에서 전기적인 신호로 바뀌어 신경조직을 통해 두뇌에 전달된 음성신호가 두뇌에서 어떻게 처리되는지에 대해서는 아직 알려진 사실이 별로 없기 때문에 실제로는 음성청취기관의 전반부의 일부만 모델링한 셈이 된다. 앞으로 음성인식성능의 향상을 위해서는 음성발생기관 및 청취기관의 보다 정교한 모델을 토대로 한 음성특징분석 방법이 계속 연구되어야 할 것으로 판단되며, 음성신호의 실험음향학적인 지식도 더 많이 확보

되어 이를 통해 보다 우수한 음성특징계수들이 개발되어야 할 것이다.

나. 패턴인식과정

패턴인식과정은 음성특징분석과정에서 추출된 음성특징 계수들과 가장 잘 부합되는 언어적 표현을 찾아내는 과정이라고 말할 수 있다. 이를 위해서는 먼저 패턴인식을 하기 위한 음성의 기본 단위(단어, 반음절, 음소, 변이음 등)를 정한 다음, 훈련용 음성 데이타로부터 미리 이들 음성단위에 해당하는 각각의 대표패턴을 구해서 저장한다. 그 다음으로 인식하고자 하는 입력음성의 특징패턴이 분석되면 이들과 저장된 특징패턴을 비교하여 가장 가까운 패턴들에 해당하는 표현들을 정리하여 인식된 단어 또는 음소의 후보로 결정하게 된다.

패턴인식과정의 가장 기본적인 방법은 소위 template matching이라는 방법으로서, 이 방법은 미지의 입력특징패턴의 시간열(time sequence)과 저장된 음성단위들의 시간열을 직접 비교하는 것이다. 이 경우 발음속도의 차이에 따른 영향을 보상해 주기 위하여 보통 Dynamic Time Warping(DTW)이라 불리우는 기술이 사용된다.

두 번째는 은닉 마르코프 모델(Hidden Markov Model(HMM))이라는 방법으로서, template matching 방법이 음성의 기본단위에 해당하는 패턴들의 대표값(들)을 저장하여 이를 비교 수단으로 삼는 데에 반하여, 이 방법은 이들 패턴들의 통계적인 정보를 확률모델 형태로 저장하고 비교하는 방법이다. 실제로 HMM 방법에서는 음성신호를 두 단계에 걸친 확률과정으로 모델링하며, 음성신호로부터 모델의 파라메터들을 추정하고 추정된 모델과 새로 입력된 음성과의 유사도를 측정하는 일련의 과정들이 명확하게 정리되어 있어서 현재로서는 가장 효과적인 패턴인식과정의 방법론으로 알려지고 있다

세번째로 최근 패턴인식의 새로운 접근방법으로 각광을 받고 있는 인공신경망에 의한 방법을 들 수 있으며, 이 방법은 부분적으로나마 인간의 두뇌모델에 기초하여 학습이 진행됨에 따라 점차적으로 정보분류능력이 향상되는 신경망 구조를 이용하고 있다. 그러나, 이 방법은 template matching이나 HMM 방법에 비해 시간영역에서의 동적인 처리능력면에서는 아직까지 뒤떨어지는 것으로 알려지고 있으며, 따라서 상기 방법들에 의한 패턴인식 이후의 후처리 과정에서 정적인 패턴분류작업에 더 많이 사용되고 있다.

마지막으로 지금까지의 접근방법과는 전혀 다른 방법으로서 규칙에 의한 패턴인식방법을 들 수 있다. 이 방법은 인공지능공학에서의 전문가 시스템의 일종으로 생각할 수 있으며, 음성특징분석 결과로부터 음성의 기본단위들로 분류하는 과정에서 일련의 음성학적인 규칙들을 선별적으로 적용하여 최종 인식결과를 도출해내는 방법이다. 이 방법은 음성분류에 적용될 음성학적 지식이 체계적이고 완벽하게 준비될 수 있다는 전제 하에서는 가장 이상적인 방법이겠지만, 실제로 이러한 지식 및 이들 지식들을 운용하는 방법론에서 한계가 있는 현 상황에서는 앞에서 설명한 방법들에 비해 오히려 성능면에서 저조한 결과를 나타내고 있는 실정이다.

다. 언어학적 처리

음소를 음성인식의 기본단위로 삼을 경우, 상위계층의 지식을 동원하지 않은 상태에서 현재의 기술로 도달할 수 있는 불특정화자 음소인식율은 70% 정도에 불과하다. 그리고, 실제로 사람의 경우에도 무의미 단어를 나열한 문장의 청취도는 이보다 그리 높은 수준이 못되는 것으로 알려져 있다. 이러한 사실은 음성인식에서의 상위계층의 지식, 즉, 어휘론적(lexical), 구문론적(syntactic), 의미론적(semantic), 그리고 실용론적(pragmatic) 지식을 효과적으로 활용하는 것이 매우 중요함을 시사하고 있다. 물론 수십 단어 정도의 고립단어인식이 목표인 응용분야에서는 단어를 음성인식의 기본 단위로 삼으면 되고 별다른 언어학적 지식을 동원할 필요도 없다. 그러나, 어휘 수가 많아지고 문장형태의 음성인식이 불가피한 상황에서는 단어보다 작은 음성단위를 기본 인식수단으로 택할 수 밖에 없으며, 어떠한 형태로든지 언어학적 지식을 동원해야만 인식시스템으로서의 기능을 구현할 수 있게 된다.

음성인식에서 언어학적 모델을 설정하는 데에는 다음 두 가지 관점이 검토된다. 첫째는 패턴인식과정에서 추정한 단어들 또는 음소들의 순서열(이들 가운데는 상당 수의 오류가 포함되어 있음)로부터 언어학적으로 올바른 문장을 도출할 수 있는 언어 모델을 어떻게 기술할 것인가에 관한 것이고, 두번째는 공학적인 문제로서 이러한 언어 모델이 주어졌을 때 최적의 문장을 찾아내는 효과적인 구현 방법은 무엇인가에 관한 것이다. 초기의 음성인식연구는 컴퓨터 언어와 같이 매우 제한적인 문법구조를 갖는 문장들을 대상으로 수행되었던 반면에, 현재의 연속인식연구는 특정 주제에 대한 대용량의 문장 데이타베이스로부터 추출한 확률론적 언어 모델을 이용함으로써 제한된 주제 하에서는 상당히 우수한 성능을 나타내고 있다. 이 때의 확률적 언어 모델로는 두 개 또는 세 개의 연속된 단어 사이의 연관관계를 확률적으로 모델링한 bigram 및 trigram 언어 모델이 주로 사용된다.

그러나, 주제에 제한을 두지 않거나, '어, 저' 등과 같은 비문법적인 표현들이 섞여있는 자연스러운 회화체 음성을 인식하기 위해서 어떠한 언어 모델을 사용하고 이를 어떻게 구현할 것인가에 관해서는 아직도 해결해야 할 많은 문제점이 남아있는 실정이다. 그리고, 의미론적 지식과 음성의 운율 정보를 인식에 어떻게 활용할 것인가 하는 점도 연구가 되어야 할 분야이다.

3. 음성인식기술의 연구동향

가. 국외의 연구동향

미국을 비롯한 선진국들의 음성인식 분야의 연구는 대략 40년의 역사를 지니고 있으

며, 1952년에 Bell 연구소에서 단일화자에 대한 고립 숫자음 인식시스템이 처음으로 개발된 것으로 알려져 있다. 미국의 경우, 1971년부터 국방성의 주도아래 음성이해시스템의 개발을 위해 당시 1500만불의 예산을 투입하여 연구가 추진되었다. 이 연구과제의 목표는 1000 단어의 어휘를 갖는 연속음성의 인식 및 이해로서, 단일화자의 음성에 대해 의미론적으로 90% 이상의 인식율을 얻는 것이었다. 과제종료 시점인 1976년에 상기 요구조건을 만족시킨 것은 패턴인식 기법에 기초를 두고 문법구조 등에 많은 제약을 둔 한 시스템 뿐이었으며, 의욕적으로 음성학 및 언어학적인 전문가적 지식을 인공지능 기법으로 구현한 방식들은 기대수준의 성능을 얻지 못하였다.

1986년부터 재개된 미국방성 주도하의 음성인식연구는 음성언어(spoken language) 프로그램이라는 이름으로 추진되고 있는데, 이 프로그램은 대용량 어휘의 연속음성인식과 대화에 의한 문제해결을 목표로 하는 음성언어의 이해라는 두 가지 주요 과제를 대상으로 하고 있다. 이들 과제는 모두 불특정화자(speaker independent) 또는 화자적응(speaker adaptive) 방식으로 실시간처리가 가능해야 한다는 전제를 달고 있으며, AT&T, BBN, Brown 대학, Boston 대학, Carnegie-Mellon 대학, IBM, MIT, Stanford 연구소(SRI), Texas Instrument사 등 주요 음성연구기관들이 대부분 참가하고 있다. 1987년부터 1992년 사이에는 낭독체의 질의 또는 명령들로 구성된 1000단어 어휘의 자원관리 데이타베이스를 대상으로 한 대용량 어휘 음성인식에 촛점이 맞추어졌으며, bigram 문법을 사용했을 경우 최고 96%의 단어인식율을 얻었다. 1991년부터는 항공여행에 관한 정보문의를 주제로 하는 회화체의 음성으로 구성되어 난이도가 훨씬 높은 항공여행정보시스템 데이타베이스에 대한 인식연구가 수행되고 있으며, 일부 기관들은 자원관리 데이타베이스에 대한 인식율에 육박하는 성능을 가진 시스템을 workstation 상에서 실시간 시범시스템으로 구현하였다.

일본에서는 1982년에 시작된 지능을 가진 제5세대 컴퓨터 프로젝트의 일환으로 음성인식연구가 추진되었으며, 1986년 우정성 주관아래 관민합동으로 설립된 ATR(Advanced telecommunication research institute) 산하의 자동통역연구소에서 간단한 일상회화나 국제무역업무를 위한 자동통역전화 과제를 추진하면서 음성인식에 대한 활발한 연구가 진행되고 있다. 7년간에 걸친 1단계 연구의 결과로 지난 1993년 1월에 일본, 미국, 독일간의 세계 최초의 자동통역전화실험을 주관하여 수행한 바 있으며, 현재 2단계의 연구가 진행되고 있다.

유럽에서는 범국가적인 정보통신분야의 연구개발 프로그램으로 1984년에 시작된 ESPRIT(European strategic program for research and development in information technology) 과제에서 다국어 음성을 이용한 정보통신 서비스를 겨냥한 음성인식 연구가 추진되고 있으며, 이 외에도 영국, 프랑스, 독일 등에서는 국가 규모의 대형 과제를 통해 대어휘 음성인식 및 자동통역을 목표로 하는 연속음성인식기술의 개발이 진행되고 있다.

나. 국내의 연구동향

국내에서의 음성인식연구는 1980년대 초반부터 본격적인 시작이 이루어졌으며, 여러 대학, 정부출연연구소 및 기업체 부설연구소 등에서 연구가 진행되고 있다. 초기의 연구는 매우 제한된 어휘수의 고립단어 인식 및 연결 숫자음 인식 등에 국한되었으며, 1986년에 고립단어인식기술을 이용한 음성 다이얼링 전화기의 시제품이 처음으로 선보였다. 1990년대에 들어서 한국과학기술원과 한국전자통신연구소(ETRI), 및 한국통신 소프트웨어 연구소 등을 중심으로 제한된 task domain하에서의 연속음성인식 연구가 추진되고 있으며, 그 중 한국과학기술원에서는 1993년에 백 단어 이상의 어휘를 대상으로 하는 연속음성인식시스템을 개발하여 시범운용을 한 바 있다. 한국전자통신연구소와 한국통신 소프트웨어 연구소는 자동통역전화를 최종 목표로 하는 음성인식연구를 추진해 오고 있으며, 현재 호텔예약이라는 주제에 대한 수백 단어 어휘의 연속인식시스템의 개발이 진행되고 있다. 그 외에도 국내 여러 대학 및 대기업 부설연구소에서도 음성인식연구가 진행 중이며, 일부 업체에서는 전화기나 VCR 등 일부 가전제품에 음성인식 기능을 구현한 시제품을 발표하기도 하였다.

그러나, 지금까지의 연구는 고립단어나 연결숫자음의 인식 범주를 벗어나지 못하거나 연속음성인식의 경우에도 문법구조 등에서 매우 제한된 것이 대부분으로 한국어 고유의 음성학적 지식이나 언어학적 지식을 효과적으로 사용하지는 못하고 있는 실정이다. 또한 표준 음성 데이타베이스가 구축되어 있지 못한 이유로 인하여, 개개의 연구기관들의 연구개발 결과들을 객관적으로 평가하고 기술결과를 공유할 수 있는 여건이 아직 조성되지 못하고 있다. 최근 음성학 및 언어학을 연구하는 연구진들과 음성공학을 연구하는 연구진들이 공감대를 형성하여 함께 일할 수 있는 분위기가 점차적으로 조성되고 있으며, 이는 국내 음성인식기술의 진보를 앞당길 수 있는 청신호라고 여겨진다.

4. 음성인식기술의 응용

사람과 자연스럽게 대화할 수 있는 컴퓨터가 개발된다면 그 응용분야는 무궁무진할 것이다. 그러나, 현재의 음성인식기술은 사용 어휘 면에서 제한이 있을 뿐만 아니라 제한된 어휘에 대해서도 필연적으로 어느 정도의 오류를 나타낼 수 밖에 없으며, 이러한 상황은 조속한 시일 내에 해결될 성질의 것은 아니다. 따라서, 이러한 시점에서 음성인식의 응용분야로 접근이 시도되는 분야들은 어느 정도의 인식오류를 감내할 수 있어야 한다는 조건이 따르게 된다. 이러한 제약에도 불구하고 여러 분야에서 음성인식기술의 실용화가 추진되고 있는데, 그 중 첫번째로 정보통신분야를 들 수 있다. 음성인식은 사람이 전화를 이용하여 컴퓨터와 정보를 주고받는 것을 가능케 해준다. 예를 들어, AT&T를 비롯한 미국의 전화회사들은 증권시세를 비롯한 각종 생활정보 조회, 은행업무 등을 전화음성인

식에 의해 수행하는 서비스를 제공하기 시작하였다. 이러한 서비스는 기존에는 버튼식 전화기로만 가능했던 기능들을 다이얼식 전화기를 가지고도 이용할 수 있게 해줄 뿐만 아니라, 전화기의 버튼입력으로는 곤란한 회사이름 등을 음성을 통해 컴퓨터에 전달하여 관련 정보를 조회할 수 있게 한다. 그 외에도 전화교환원이 수행하던 서비스를 자동화하는 데에도 음성인식기술이 사용되고 있는데, 전화번호 안내 서비스나 전화통화의 수신자 요금부담 여부확인 등이 그 대상으로 부각되고 있다.

음성인식의 두번째 응용분야로 일반 가정용 또는 산업용 기기들에 음성인식기능이 부가되는 것을 들 수 있다. 사람이 전화번호 또는 상대방의 이름을 말하면 자동적으로 전화를 걸어주는 음성인식 전화기, 음성으로 방송 채널을 선택하는 카 스테레오, 음성으로 예약녹화를 지시하는 비디오기기(VCR), 그리고 음성명령에 의해 작동하는 퍼스널 컴퓨터 등이 이미 등장하고 있으며, 장난감이나 오락기기에도 음성인식기능이 적용되고 있다.

세번째로 음성인식에 의한 원고작성(voice dictation)을 들 수 있으며, 영어의 경우 최근 2만 단어에서 5만 단어 정도의 어휘를 인식할 수 있는 제품들이 상품화되었다. IBM, Dragon Systems, 및 Kurzweil AI사 등에서 개발된 이들 제품들은 사용자가 자신의 목소리로 수 분에서 수십 분 가량 말을 하여 인식시스템이 자신의 목소리에 익숙해지도록 하고, 각 단어와 단어 사이를 또박또박 띄어 읽어야 한다는 등의 제약조건이 있기는 하지만, 분당 50 단어 이상의 입력이 가능하다. 뿐만 아니라 인식오류도 3%에서 5% 사이로 비교적 낮은 편이고, 오류가 발견될 경우 음성명령에 의해 즉시 수정할 수 있는 기능을 가지고 있어서, 방사선과 의사들의 X선 사진 판독보고서 작성 등의 응용분야에 활용되고 있다.

이 외에도 음성인식기술은 신체장애자들에게 많은 도움을 줄 수 있으며, 음성합성 및 기계번역기술과 결합되어 만들어지는 자동통역전화 서비스도 현재 여러 나라에서 많은 연구가 진행되고 있는데, 제한된 주제에 대해서나마 성공될 경우 지대한 파급효과가 기대되는 분야이다.

5. 향후 전망 및 결론

음성인식기술의 향후 전망과 관련해서는 섣부른 낙관이나 비관, 그 어느 쪽도 할 수 없는 상황이라고 보여진다. 1970년대 음성인식연구가 본격적으로 추진되기 시작한 시점만 해도 1980년대 후반이면 음성인식분야의 시장이 크게 형성될 것이라는 예측이 주류를 이루었지만, 1990년대 중반에 접어드는 현재 시점에서 보더라도 음성인식기술의 실생활에의 응용은 아직까지 시작단계라고 밖에 볼 수 없으며, 사람과 자유롭게 대화하는 컴퓨터가 과연 언제 출현할 수 있을지에 대해서는 예측조차 하기 어려운 실정이다. 또한 일부의 비관론자들은 음성인식이 너무나 어려운 과제이므로 이 분야의 연구에 투자하는 것

은 낭비에 불과하다는 주장을 했지만, 오늘날 전화음성을 통해 컴퓨터에 수록된 정보를 조회하고 음성인식에 의해 원고를 작성하는 일 등이 현실 세계에서 이미 진행되고 있다.

실제로 미국 등의 선진국에서는 특정 응용분야를 대상으로 하는 불특정화자 대용량 어휘의 음성인식시스템이 금세기 안에 실용화될 것이라고 전망하고 있다. 그리고 전화음성을 이용한 정보조회 및 음성명령에 의한 컴퓨터 조작 등은 더욱 가까운 장래에 일반화될 것으로 예상된다. 그러나, 현재의 기술수준, 특히 회화체 언어처리에 관한 기술수준으로 볼 때, 사람과 자유롭게 대화할 수 있는 컴퓨터의 개발은 다음 세기 중에나 실현될 수 있을 것으로 보인다. 따라서, 앞으로의 음성인식연구는 회화체 음성의 인식을 위한 언어학적 처리에 큰 비중이 두어질 것으로 판단되며, 현재 실험실 환경에서 진행되어온 음성인식기술의 실용화를 위한 작업들, 예를 들면 잡음환경에서의 음성인식 및 새로운 화자에 대한 적응방식연구 등에도 많은 관심이 기울여질 것이다.

한국어 음성인식기술이 이러한 흐름을 뒤따라 잡기 위해서는 한국어의 음성학 및 언어학적 지식에 대한 기반기술이 확보되고, 이를 공학적으로 구현하는 작업이 뒷받침되어야 할 것이다. 특히 한국어 음소들의 음향학적인 특성 연구를 통해 효과적인 음성특징분석 방법 및 음소들의 통계적 모델링 방법이 정립되어야 하며, 한국어 음성언어처리 방법에 대해서도 실제적인 접근방법이 개발되어야 할 것이다. 또한 음성학적으로 균형잡힌(phonetically balanced) 단어 및 문장들에 대한 표준 음성 데이타베이스 구축이 이루어져서 국내 연구진들이 동일한 평가기준을 토대로 공동연구를 수행할 수 있는 여건이 조속히 마련되어야 하리라고 본다.

음성인식과 음성합성, 그리고 음성부호화 등 음성신호처리기술이 다가오는 정보통신시대의 주요 핵심기술 중 하나가 될 것임에는 이론의 여지가 없다. 그 중에서도 여러 학문분야에 걸쳐 가장 복합적인 기술인 음성인식기술의 발전을 위해서는 음성학, 언어학, 심리학을 비롯하여 반도체, 신호처리, 컴퓨터공학에 이르기까지 제 분야의 연구인력들의 협력연구와 더불어 이를 뒷받침할 투자와 지원이 지속적으로 이루어져야 할 것이다.

참 고 문 헌

[1] 김형순, 김희동, 임병근, 은종관, "음성처리기술의 현황과 전망," 한국통신학회지, 제8권 제6호, 1991년 6월.

[2] L. R. Rabiner and B. H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[3] D. B. Roe and J. G. Wilpon, "Whither speech recognition: the next 25 years," IEEE Communication Magazine, Vol.31, No.11, Nov. 1993.

[4] 구명완, "음성인식기술의 현황과 전망," 대한전자공학회지, 제20권 제5호, 1993년 5월.

그림 1. 음성 인식 시스템의 기본 구성도