特輯

技術解說

음성 인식 기술 현황 및 실용화 전망

김 순 협
(광운대학교 컴퓨터 공학과)

Ⅰ. 서 론

음성에 관한 자연 과학적 연구에 박차가 가해진 것은 전기 통신이 시작 되면서 부터이다. 그 이후의 착실한 연구에 의해 Computer를 중심으로한 새로운 수단이 추가하여지고 정보통신 시스템의 발전에 따라 기계와 인간 사이에 정보 교환의 필요성이 증가하면서 인간에게 더욱 자연스럽고 친숙한 정보 전달 수단인 사람의 말, 즉 음성을 기계가 인식하는 음성 인식 연구가 현저한 발전을 해왔다. 음성인식이란 음성 속에 내재 되어 있는 언어 정보를 자동으로 추출하여 그 음성의 정보를 기계가 이해하는 과정을 말한다. 음성인식에 대한 연구는 수십년 전부터 수행되었으며 현재의 음성인식 기술수준은 인간의 능력에 비하면 보잘것이 없다. 이와같은 이유중의 하나는 아직 인간이 음성을 인식하는 정확한 방법을 모른다는 것이다. 음성이 인간의 귀에 도달하면 外耳, 中耳, 內耳를 거쳐 주파수 성질을 나타내는 신호로 바뀐다는 사실을 알고 있지만 이 신호가 뇌에 도달하여 어떻게 언어정보로 변하는 지는 아직 알려지지 않고 있다. 그렇지만 음성인식 분야 관련 연구가들은 인간의 음성인식 방식을 알아내려고 하고 있으며 또한 기존 방식을 이용하여 연구를 계속하고 있다.

Ⅱ. 음성 인식 시스템

일반적인 음성 인식 시스템은 그림 1에서 보는 것처럼 음성으로부터 음성 패턴의 특징을 추출하여 기준 패턴을 만든 후 미지의 음성이 입력되면 저장된 기준 패턴과 비교하여 가장 유사한 기준 패턴을 찾아

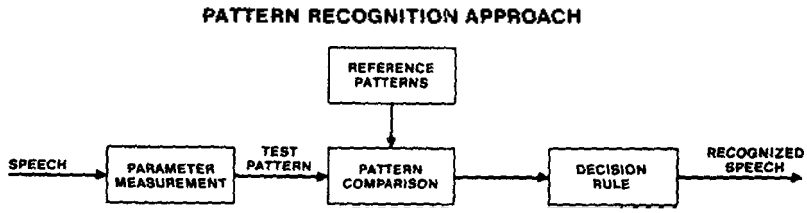

그림 1. 음성 인식 시스템의 개념도
Fig. 1 Diagram of speech recognition system

인식하는 과정으로 나누는데 이러한 알고리즘을 패턴 매칭 알고리즘이라고 부른다.

2.1 음성 인식의 분류

음성 인식 시스템은 음성의 형태, 화자의 수, 인식 방법등에 따라 다음 표와 같이 분류할 수 있다.

표 1. 음성 인식의 분류
Table 1. Classification of speech recognition

분류	종 류	비 교
음성의 형태	격리 단어	단어의 앞뒤에 묵음이 있다고 가정한 음성
	연결 단어	격리 단어가 자연스럽게 발음된 음성
	연속 음성	연속 단어가 자연스럽게 발음된 음성
화자의 수	화자 종속	훈련된 화자의 음성으로 test하는 실험
	화자 독립	훈련하지 않은 화자의 음성으로 test하는 실험
인식 방법	패턴 매칭	음성의 특징을 비교하여 인식하는 실험
	확률적 방법	음성의 발생 확률을 이용하는 방법

최근에 연속 음성 인식 시스템의 하나로 대화체 음성(Conversational speech)인식 시스템이 연구되고 있는데 이를 음성 이해 시스템이라고도 불리우며, 언어지식을 사용하는 것이 중요시 되고 있다. 대화체 음성인식 시스템은 매우 어렵기 때문에 문장내에서 필요한 단어만 선별하여 인식할 수 있는 단어선별(word spotting) 인식 시스템의 연구가 최근 진행되고 있다. 음성의 인식에 있어서 단어를 기준 패턴으로 인식에 사용하면 단어내의 연음(coarticulation) 현상을 고려할 필요가 없기 때문에 인식율이 높은 반면 인식대상 단어수가 많아질수록 메모리와 계산량이 증가한다는 단점이 있고 단어 사이의 연속음성에 내재되어 있는 연음현상을 표현할 수 없다는 단점도 있다. 그러나 음소를 기준 패턴으로 사용하게되면 대상 단어수가 늘어난다고 하여도 계산량 및 메모리 사용량이 많이 증가되지 않으며 훈련과정도 간단하며 또한 단어사이의 연음현상도 쉽게 표현될수 있다. 한편 음소 인식 과정에서는 음소의 발음규칙이 명확히 알려져 있지 않기 때문에 인식율이 떨어지는 단점이 있다.

2.2 특징 추출

음성 인식에 있어서 가장 기본이 되는 특징을 추출하는 것은 매우 중요하다. 기존의 가장 많이 사용되어 오고 있는 방법 가운데 하나는 음성파형을 그 특징으로 정하는 것이다. 그러나 음성파형은 시간에 따른 많은 변화를 갖고 데이타의 양도 많아 이를 주파수 영역으로 변환시켜 특징을 추출하는 방식을 사용한다. 음성이 구강(vocal tract)으로부터 발성된다는 사실을 근거로 구강의 형태를 필터(filter)로 가정하고 그 필터 계수를 음성의 특징으로 삼는 방법이 있으며, 인간의 귀가 음성을 분석하는 방식을 이용하는 auditory 분석 방식도 있다. 최근에는 시간 영역의 동적 특징(dynamic feature)을 주파수 영역의 특징(spectral feature)들과 함께 사용하기도 하여 최근의 연구 결과에 의하면 mel scale된 LPC cepstrum의 차(differential LPC cepstrum), energy 및 energy차를 음성 특징으로 사용하였을때 높은 인식율을 보이고 있다.

2.3 음성 인식 기술

2.3.1 DTW(Dynamic Time Warping) 알고리즘

DTW 알고리즘은 음성의 발성을 변화에 의한 음성 패턴의 시간적 변동을 비선형적으로 정규화 시키는 패턴 정합 방식을 이용한 알고리즘이다. 인식과정에서 사용되는 알고리즘으로서 입력음성 패턴과 기준 음성 패턴간에 거리를 측정할때 동적 프로그래밍의

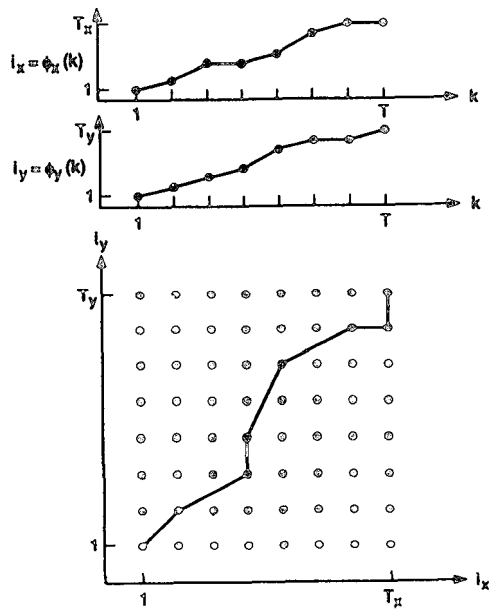

그림 2. 패턴 정합의 예
Fig. 2 Example of Pattern alignment

기법을 이용한다. 기준 음성 패턴과 입력음성 패턴의 발음시간의 차이가 있을 경우 두 패턴사이의 거리(distance)를 측정하기 위해서 우선 기준음성 패턴의 각 프레임과 그에 대응하는 입력음성 패턴의 프레임 번호 사이의 쌍(pair)을 찾고 이 대응쌍은 워핑함수에 의하여 구해지며 이때 동적 프로그래밍 기법이 이용된다. 이 예가 그림 2에 나타나 있다. 동적 프로그래밍 기법에 따르면 워핑 함수에 의해 구해진 경로는 모든 경로에 의한 거리중 최단의 경로로 정해 진다.

2.3.1.1 LB(Level Building) DTW 알고리즘

LB DTW 알고리즘은 연속음성을 인식 할때 DP 알고리즘의 단점을 극복하는 방법으로 super 기준 패턴을 격리 단어 인식과 같이 각각 한개의 기준 패턴으로 보고 시험 패턴과의 워핑을 각 level마다 연속으로 수행하는 알고리즘이다. 또한 이 알고리즘은 입력 단어수를 지정할 수 있고 계산량을 줄이기 위해 여러가지 DP range 감축을 시도할 수 있다. 한편 k =1, 2, ..., K로 구분된 기준 패턴의 time frame을 $j=1, \ldots J(k)$로 표시 할 때 (J(k)는 표준 패턴의 길이) 입력 패턴이 가장 잘 매칭되는 기준 패턴의 sequence q(1), q(2), ... q(R)을 결정하는 알고리즘으로 One Stage DP를 들 수 있다.

2.3.2 VQ(Vector Quantization) 이론

벡터의 계열을 이용하여 실제 데이타의 양을 압축시키는 방법으로 음성 인식할 때 음성 신호 데이타를 압축시키기 위해 기준 패턴을 생성하는데 VQ를 이용한다. 다음 그림은 VQ를 이용한 음성 인식시스템이다. 입력된 음성의 특징 벡터를 미리 저장해둔 특징 벡터들 중에서 가장 잘 매칭되는 하나의 벡터와 매칭시켜 준다.

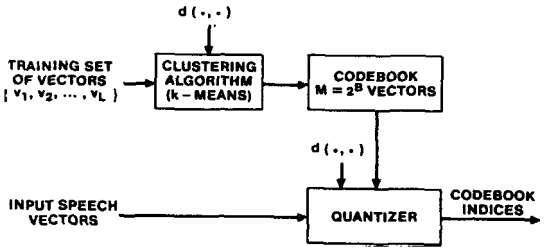

그림 3. VQ를 이용한 음성인식 시스템
Fig. 3 Speech recognition system using VQ

2.3.2.1 MS(Multi Section) VQ

기존의 VQ에는 음성 신호의 음향적인 특성만으로 VQ codebook이 생성되어 시간적인 정보가 포함되어 있지 않았다. 이에 한 단어를 발성 순서에 따라 몇개의 구간으로(Section) 나누어 구간 별로 독립된 codebook을 작성함으로 시간적 정보를 포함하는 알고리즘이다. 즉, VQ codebook의 계열로서 시간 변화 패턴을 고려하는 방법을 MSVQ codebook이라고 한다. 그림에서 단어를 동일 길이의 구간으로 나눠 각 구간마다 집단화 기법을 써서 MSVQ codebook을 작성하는 것을 보여주고 있다.

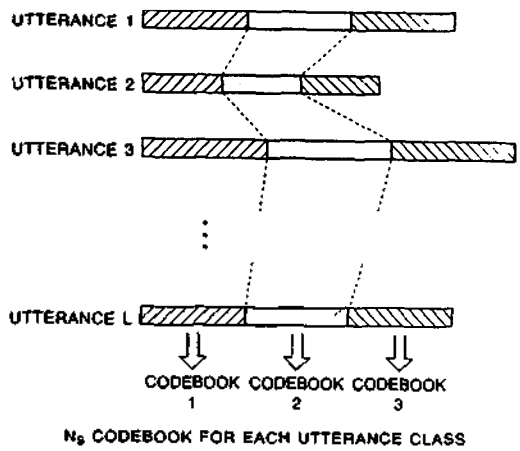

그림 4. MS VQ의 codebook 작성
Fig. 4 MS VQ codebook Generation

2.3.3 HMM(Hidden Markov Model)

HMM 알고리즘은 음성인식 시스템 개념도에서 훈련과정 및 인식과정을 수행하는 알고리즘으로서 1970년말부터 음성인식 알고리즘으로 많이 사용되었다. 최근에는 높은 인식률과 빠른 인식시간 때문에 대용량 음성인식 시스템이 많이 사용되고 있다. HMM 알고리즘의 기본적인 개념은 음성이 Markov모델로 모델링될 수 있다는 가정하에 훈련과정에서 Markov 모델의 파라미터를 얻어 기준 Markov모델을 만들고 인식과정에서는 입력음성과 가장 유사한 기준 Markov모델을 찾아냄으로써 인식한다. 앞에 Hidden을 붙이는데 음성패턴의 다양한 변화를 수용하기 위해서이며 state가 음성 패턴에 관계없이 모델 속에 숨어 있기에 붙여진 이름이다. HMM에서 state선정에

관한 stochastic process로 음성 패턴의 각 특징은 state의 선정 확률과 출력 확률등으로 표현할 수 있다. HMM model 표시 방법은 $\lambda = (A, B, \pi)$이며 실제 응용할때에는 다음 세가지 문제를 해결해야 한다.

1. 주어진 model에서 관측열의 발생 확률 $Pr(O|\lambda)$ 계산 문제
   ⇒forword-backward algorithm 이용
2. 최적 상태열 계산문제
   ⇒Viterbi algorithm 이용
3. $Pr(O|\lambda)$를 최대화 하기 위한 A, B, π의 조정문제
   ⇒Baum-Welch *reestimation algorithm* 이용

그림 5에서 HMM의 예를 보여주고 있다.

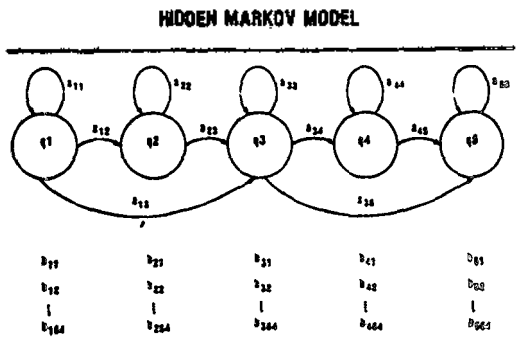

그림 5. HMM의 예
Fig. 5 HMM diagram

2.3.3.1 DHMM(Dynamic HMM)에 의한 인식

DHMM은 정적 스펙트럼의 특징 파라메타와 동적 스펙트럼의 특징 파라메타를 함께 모델링한 것이다. DHMM에서는 정적 특징과 동적 특징 파라메타를 모두 이용함으로

$Pr(O_t|S) = Pr(O_{It}, O_{Dt}|S)$ 라 할 수 있다.

그런데, 정적 특징(O_{It})과 동적 특징(O_{Dt})은 서로 상관 관계가 거의 없으므로 다음식과 같이 쓸 수 있다.

$$Pr(O_{It}, O_{Dt}|S) \cong Pr(O_{It}|S)\ Pr(O_{Dt}|S)$$
$$\cong Pr(O_{It}|M_I, M_D)\ Pr(O_{Dt}|M_I, M_D)$$

DHMM은 정적 특징과 동적 특징사이의 상관 관계가 매우 작다는 사실에 의해 파라메타의크기에 따른 그다지 계산량의 증가 없이 모델링 될 수 있으며 인식 알고리즘은 정적특징 및 동적특징이 조합된 것만 제외하면 HMM에 의한 인식과 동일하다.

2.3.4 Neural network

최근에 많이 부각되어 음성 인식에 쓰이고 있는 Neural network는 인간의 뇌가 하는 역활을 모델링하여 모델된 뇌세포들을 연결시켜줌으로써 인간의 뇌의 기능을 갖는 역할을 수행시켜 주는 알고리즘이다. 학습에 의한 방법에 따라 supervised learning neural network과 unsupervised learning neural network로 나눌 수 있는데 Supervised learning의 대표적인 알고리즘은 single layer perceptron과 MLP(multilayer perceptron)을 들 수 있다.

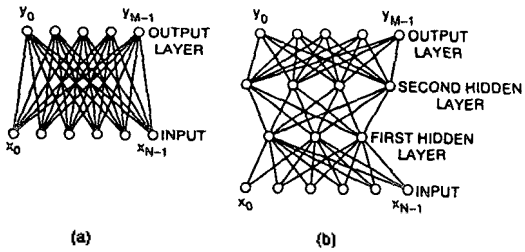

그림 6. 단층 및 다층 퍼셉트론
Fig. 6 Single layer and Multilayer percetron

single layer perceptron의 경우에 exclusive OR는 분류가 불가능함으로 실제에 있어서는 MLP가 음성인식 시스템에 이용된다. 한편 음성에 있어서 주파수에 의한 특징, 음성 발성 시간을 모두 포함할 수 있는 MLP의 변형으로서 TDNN(time delay neural network)이 개발되었다. 특별히 시간적인 특징이 포함되도록한 알고리즘으로 음소나 단어에 있어 높은 인식율을 보이고 있다. 그림 7에 TDNN 신경망 시스템을 보이고 있다. 또 다른 supervised learing 알고리즘으로서 Kohonen이 제안한 LVQ(learing vector quantization)를 들 수 있다.

Unsupervised learning알고리즘의 대표적인 것은 Kohonen의 feature map이다. Feature map알고리즘은 음성 데이타를 의미에 관계없이 훈련시켜 음소를 대

표할 수 있는 특징을 나타낸 알고리즘이다. 이 알고리즘은 실제 인간의 청각작용과 비슷하나 음성인식 시스템에 적용하였을 경우 supervised learing알고리즘 보다는 낮은 인식률을 보인다.

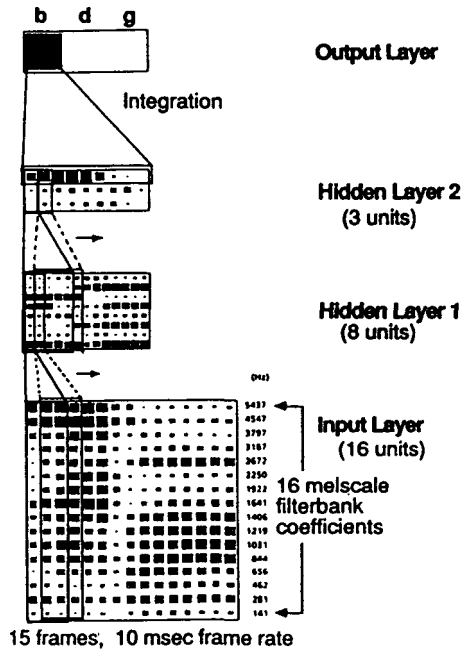

그림 7. TDNN 신경망 시스템
Fig. 7 TDNN neural network system

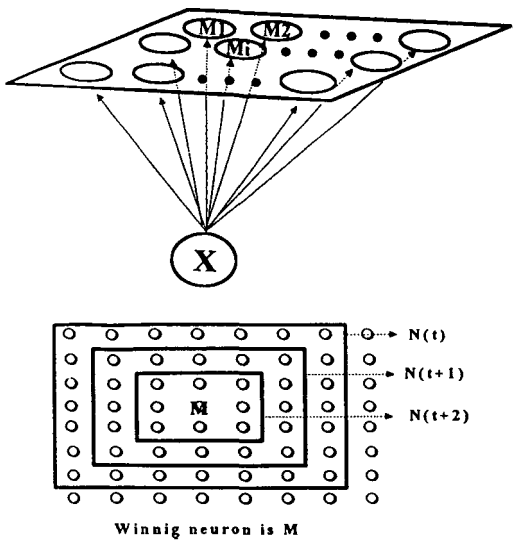

그림 8. Kohonen의 일반적인 신경망구조
Fig. 8 General neural network system of Kohonen's

2.3.4.1 DPNN에 의한 인식

DPNN(Dynamic Programming Neural Network 혹은 DNN)은 기존의 인식 방법중 DP에 Neural Nets를 가미한 새로운 인식 방법이다. 음성 인식을 수행함에 있어서, 음성은 특징 파라메타의 time sequence로 처리가 되는 데, 이러한 음성으로 인식을 함에 있어서 두가지의 난점이 있어왔다. 그중 하나는 시간축의 distortion이며, 다른 하나는 spectral pattern의 변화이다. 이러한 문제점을 DPNN에서는 DP를 사용하여 시간축의 distortion문제를 해결하고, Neural Nets의 패턴에 대한 학습능력을 이용하여, spectral pattern의 변화에 의한 인식율의 저하를 막는 방법이다. 이러한 Neural Nets의 기능은 화자독립의 인식에 적합한 방법으로 그림 9와 같이 표현된다.

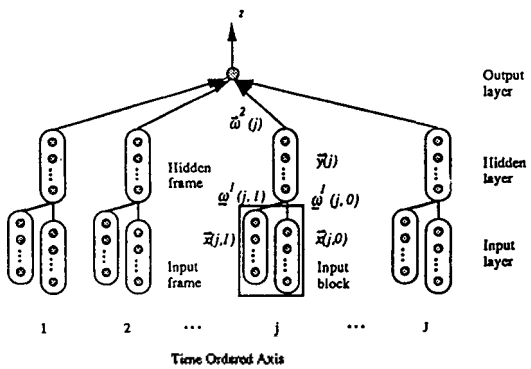

그림 9. DPNN 시스템의 예
Fig. 9 Example of DPNN system.

Ⅲ. 국내외 음성 인식 연구 동향

3.1 국 외

3.1.1 미 국

1971년에서 1976년까지 SUR(speech understanding research)이라는 음성이해연구 프로젝트가 수행되었으며 최근에는 1984년부터는 5년에서 10년 기간으로 음성 및 자연언어 처리에 관한 새로운 프로젝트가 수행되고 있다. 대용량 음성인식 시스템과 음성언어 이해에 관한 연구를 추축으로하여 특정 task 영역에서 자연스러운 음성을 실시간으로 인식하는 화자독립 혹은 화자적응 음성인식 시스템을 개발하는 것을 목표로 한다. 대용량 텍스트(text)처리에 필요한

기술을 개발하는 것도 목표로 하며 메시지 이해(message understanding), 자연언어 학습(natural language learning) 및 데이타베이스구축 등에 관한 연구를 한다. 또한 기계번역에 관한 연구도 포함한다. 프로젝트의 성공을 위하여 매년 음성 및 자연언어 워크샵을 개최하여 연구에 참여하고 있는 연구원들이 최신의 정보를 교류하고 앞으로의 연구방향을 모색하도록 하고 있다. 대용량 음성인식 시스템의 개발을 위해선 Wall Street Journal을 통해 얻어진 낭독체 문장을 이용하여 인식실험을 수행하고 있는데 초기단계로서 5,000단어로 한정하고 있다. 1992년 2월에 실험하였을 때 화자독립 인식율은 82.9%, 화자종속 인식율은 89.3%이었는데 1992년 12월에는 화자독립 인식율은 94.7%, 화자종속 인식율은 95.5%로 나타나 〈표 2〉의 1000단어 인식 시스템과 유사한 결과를 나타내었다.

표 2. DARP 음성인식 시스템의 성능 비교
Table 2. *Comparison of performance with DARP speech recognition system*

음성 데이타 베이스 종류	인식율	기관
RM 낭독체 1000 단어 연속 음성 인식 * 1991년 2월 실험	96.4%	CMU
	96.2%	BBN
	95.6%	MIT
	95.5%	AT&T
ATIS 회화체 연속 음성 인식 부문 * 1992년 2월 실험	94.2%	BBN
	92.7%	SRI
	89.6%	CMU
	87.5%	MIT
ATIS 회화체 연속 음성 인식 음성인식, 자연언어처리 부문 * 1991년 2월 실험	67.9%	SRI
	64.2%	BBN
	60.0%	MIT
	48.3%	CMU

3.1.2 일 본

일본에서의 음성인식기술은 1982년부터 추진한 제5세대 컴퓨터 프로젝트의 일부인 "음성과 자연언어를 통한 컴퓨터 입출력"이라는 제목으로 연구가 진행되었으나 연구결과의 대외발표는 거의 없었다. 최근에의 음성인식 관련 프로젝트는 ATR (Advanced Telecommunications Research institute) 산하 자동통역연구소에서 1986년부터 수행하고 있는 자동통역전화(automatic telephone interpretation) 프로젝트가 미국, 독일과 국제 공동 연구로 매우 발전적인 결과를 발표했다. 그리고 1987년부터 교육, 과학, 문화상의 자금지원을 받고 있는 "Advaced man-machine interface through spoken language"이라는 국가 프로젝트가 있다.

국가 프로젝트는 음성에 관한 기술을 분석, 특징추출, 인식, 합성, 지식처리, 잡음에서의 음성처리 및 평가기술 등 8가지의 핵심기술로 나누어서 약 185명의 연구자가 연구를 수행하고 있다.

3.1.3 유 럽

범유럽국가들이 수행하는 연구는 ESPRIT(European Strategic Program for Research and development in Information Technology)라는 정보통신에 관련된 유럽국가들의 공동 프로그램이 있다. 이 프로그램은 ESPRIT I(1984~1989), ESPRIT II (1988~1993), ESPRIT III(1992~1997)의 세단계로 나누어서 진행되는데 음성인식에 관한 연구는 매 단계마다 주요 추진 과제였다.

한편 영국에서는 국가주도의 Alvey program내에서 국가 연구소와 산업체 연구소가 협력하여 음성인식 관련연구를 수행하였으며 현재는 ITI(Information Technology Initiative)프로젝트가 시작되어 음성인식 및 데이타베이스 구축에 대한 연구가 진행되고 있다. 프랑스에서는 CNRS(national research agency)와 상공부에서 후원하는 "Human-machine communication"이라는 프로젝트에서 음성통신, 자연어처리에 관한 연구를 수행하고 있다. 독일에서는 SPICOS(Siemens-Philips-IPO Continuous Speech recognition)라는 대형 프로젝트에서 연속음성인식 기술에 관한 연구를 수행하였으며 최근 1991년 1월부터 ASL(Architecture for Speech and Language research)라고 불리우는 새로운 프로젝트가 4년 계획으로 시작되어 음성 및 텍스트 데이타베이스 구성 및 대용량 음성인식 알고리즘 개발에 역점을 두고 있다. 이 프로젝트의 연구결과는 실시간으로 음성의 자동통역을 실현하는 VERBMOBIL이라는 야심찬 프로젝트에 사용될 것이다. VERBMOBIL은 1991년부터 시작되어 20년간 지속될 대형 프로젝트이다. 다음은 연구 기관이나 대학에서 최근 개발된 음성 인식 시스템 및 현재 상용화 되고 있는 음성 인식 시스템을 표에 나타내고 있다.

표 3. 최근에 개발된 음성인식 시스템
Table 3. Recent developed speech recognition system

제조업체	국 명	시 스 템 명	특 징	단 어 수	인식률(%)
CMU	미 국	SPHINX	화자 독립 연속음성	1000	96. 4
SRI	미 국	DECIPHER	화자 독립 연속음성	1000	95. 2
BBN	미 국	BYBLOS	화자 종속 연속음성 화자 적응	1000	98. 7 94. 8
Lincoln Lab	미 국		화자 독립 연속음성	1000	87. 4
ATR	일 본		화자 종속 연구음성 화자 적응	1035	95. 3 89. 7
IBM	미 국	Tangora	화자 종속 고립단어	20,000	95
NEC	일 본		화자 종속 고립단어	1,800	97. 5

표 4. 상용 음성인식 시스템
Table 4. Commercial speech recognition system

제조업체	국 명	시 스 템 명	특 징	단 어 수	인식률(%)
Dragon System	미 국	Voice Scribe 1000 Dragon Dictate	화자 종속 고립단어 화자 적응 고립단어	1000 30,000	
IBM	미 국	Voice command	화자 종속 고립단어	64	95~98
Kurzweil Applied Intelligence	미 국		화자 종속 고립단어	1000	
NEC America	미 국	SAR-10 SR-10 DP-200	화자 종속 고립단어 화자 종속 고립단어 화자 종속 연결단어	250 128 150	98 98
Texas Instrument	미 국	Speech Command	화자 종속 연속음성	1000	
VOTAN	미 국	Voice-Card Voice-card Voice-Key	화자 독립 연결음성 화자 종속 연속음성 화자 종속 고립단어	13 640 64	 93 98
MARCONI	영 국	ASR-1000	화자 종속 연속음성	200	
VECSYS	프랑스	DATAVOX	고립단어 연결단어	5000 300	

3.2 국 내

3.2.1 국내의 기술동향

국내의 음성인식에 대한 연구는 1980년대 초부터 본격적으로 이루어 졌으며 일부 대학이나 연구소에서 진행되던 연구도 이제는 많은 대학에 분포되어 진행되고 있다. 음성인식 기술면에서도 초기의 DP 방법에서 부터 VQ 방법, HMM 방법 그리고 최근에는 NN 방법을 이용한 인식 실험결과가 발표되고 있어 10여년의 역사로 볼때 주목할 만한 발전을 이룩하였다고 볼 수 있다. 또, 인식 대상어휘가 소규모에서 중규모로, 인식대상도 특정화자에서 불특정화자로, 인식대상의 단독어에서 연결어 또는 연속어등으로 확장되어 연구되면서 국외의 음성인식 기술수준과 큰 차이없이 연구되고 있는 실정이다. 최근 3년간 국내 학회지 (한국음향학회지, 한국음향학회학술지, 대한전자공학회지, 대한전자공학회학술지, 정보 과학 회지)에 게재된 음성인식 관련 논문들의 연구동향을 살펴보면 표5와 같다. '80년대에는 VQ와 HMM, DP 방법을 이용한 음성인식 연구가 거의 비슷하게 진행되었다가 '90년도에 들어와서는 HMM방법을 이용한 연구가 과반수를 차지하고, 상대적으로 VQ에 의한 연구가 감소되었다. '91, '92년도에는 DP와 VQ 방법에 의한 연구가 현저히 감소되면서 '93년에 이르러서는 NN 방법을 이용한 연구가 HMM을 이용한 연구와 동등하게 진행되고 있어 국외의 연구동향과 비슷한 경향을 보이고 있음을 알 수 있다. 표 6에 '92, '93년도 국외 학술지에 게재된 음성 인식 연구 동향을 볼 수 있다. 한편 최근에는 HMM방식 및 Neural Network 방식을 접목한 Hybrid 방식의 연구가 계속 나오고 있으며 HMM에서도 파라메터의 개선에 관

표 5. 국내외 음성인식 연구동향

Table 5. The domestic and foreign tendency of speech recognition research

년도	기 관	인식단위	방 법	인 식 률
91년	경북大	0~9 숫자음	Chip 구현을 위한 IDMLP신경회로망	종속 100% 독립 96%
	관동大	한국어 7개단위	유클리드거리 판정법	91.01%
	명지大	148 지역명	VQ+histogram	종속 : 96.6%
	명지大	숫자음	Fuzzy VQ	86%
	건국大	10개도시명	DTW	84.6%
	명지大	연결단어	OSDP	85%
		숫자음		78.9%
	광운大	고립단어	HMM	99%
		146 DDD 지역명	DHMM	92.7%
	광운大	146개 DDD	퍼지를 이용한 HMM	90.6%
92년	아주大	음소 (자모음)	BP	84.3%
	명지大	단어 DDD 지역명 (28)	신경망 NPU	96.4%
	명지大	단어 DDD	HMM	93%
	경북大	숫자음 (단음절)	MLP	97.8%
	아주大	자음	MLP	87.5%
	명지大	28개 DDD 지역명	VQ	64.9%
			Fuzz VQ	76.1%
			MMF	95.4%
	연세大	10개 단독숫자음	DTW	93%
	명지大	28개도시명	뉴럴퍼지 패턴매칭	98. 8~100%
	부산大	7 모음	INNA모델 (Integrated NN)	종속 : 100% 독립 : 91.4%
	동아大	4연 숫자음 35개	1단 DP매칭	89.8% (조음결합처리 안했을때 84.1%)
	동아大	연속모음 6개 지명음	다층 신경회로망	종속 : 100% 독립 : 99.4% 음소분할결과 : 94.5% 정합률
93년	동아大	모음 (5개)	PNN+VQ	82.4%
	광운大	단어 (24단어)	HMM-LR	95%
	명지大	DDD지역명 28개	MSEqui-Segment Fuzzy	92%
	아주大	전체모음	신경망역단파BP	94.8%
	KAIST	음소	HMM	74.9%
		단어		93%
	광운大	단어	OSDP	종속 : 92.2%
				독립 : 86.2%
	관동大	음소	TSNN	음소 : 82.4%
		단어	DTW	단어 : 99%
	광운大	단모음 8개	VQ/HMM	89%
			VQ/MLP	92.68%
	KAIST	음소	LVQ2	60.4%
	서울大	연결숫자	TDNN, 구문분석	연결숫자 : 82%
				숫자 : 95.6%
	연세大	단독음(종속, 자동차 소음환경시속 100km)	변형 DTW	
	KAIST	단독숫자음	Hidden Control NN에 의한 비선형 예측, HMM의 segmentation기능접합	가중거리 HCNN : 97.6% 유클리드거리 : 95%

표 6. 국외 음성인식 연구동향
Table 6. The foreign tendency of speech recognition research

년도	기 관	인식단위	방 법	인 식 률
92년	CMU(U.S.A)	word	HMM	97%
	Tokyo Institute of Tech.(Japan)	113 word	HMM + LR Parsing	97.3%
	NTT(Japan)	Phoneme	K-means-VQ + NN	87.6%
	ART(Japan)	Speaker dependent isolated word(5240word)	LVQ-HMM-LR	91.6%
			TDNN-LR	92.6%
93년	NTT(JAPAN)	Phoneme	HMM	종속 : 74.5% 독립 : 67.5%
	Oregon Graduate Institute of Science Tech.	Phoneme Classification	DP	90%
	Matra Communication (France)	77 word	DIHMM	97%
	CMU(U.S.A)	400 Sentence 2000 Word	MS-TDNN	98.5%
	France + U.S.A		STNN(selectively Trained NN)	word : 93% noisy word : 85%

한 연구가 진행되고 있다.

Ⅳ. 음성인식 기술의 실용화 전망

실용화를 위하여 필요한 인식률이 98%라는 사실을 감안할 때, 또 현재의 음성 인식 기술 현황에 비추어 볼때 2~3년 내에 실용화가 가능한 분야는 고립단어 인식기술을 이용하는 단계로 생각된다. 숫자음과 수십개 정도의 고립단어를 포함하는 어휘를 가지고 서비스 가능한 분야를 생각할 수 있다. 연결단어를 인식하는 실용화된 시스템의 개발이 현재로서 아주 불가능 한것은 아니지만 만족할 만한 서비스를 제공할 확률이 상대적으로 상당히 낮은 것이 현실이다. 이런 간단한 기술을 활용할 수 있는 서비스로 먼저 700서비스를 생각할 수 있으며 이 경우 숫자음과 적은 어휘의 음성인식 기능의 첨가만으로도 push button의 기능을 대체할 수 있으므로 시스템의 부가가치뿐만 아니라 보다 다양한 서비스를 사용자에게 제공할 수 있다. 이외에도 은행의 잔고조회, 특정기관의 내선 전화번호 안내 시스템, 장애자용(car phone용) voice dialing 시스템, 음성작동 가전기기 등을 들 수 있을 것이다. 따라서 현재의 실용화 음성인식 기술은 고립단어, 소어휘 인식에 효과적인 DTW-based 기술이 바람직하다고 생각된다. 한편 실용화를 목적으로 하는 경우 key word spotting 기술에 대한 연구도 병행되어야 할 것이다.

4.1 음성 인식의 기술 전망

먼저 HMM의 성능향상을 들 수 있다. 90년대에 들어서 많은 HMM 알고리즘을 이용한 음성인식 연구가 진행되었으며 또한 매우 좋은 결과를 얻었지만 아직도 보완되어야할 부문이 많다고 본다. 그러므로 우선 HMM 파라메터에 대한 개선이나 neural network와 같이 사용하거나 HMM알고리즘의 변형인 HMM-Net(Hidden Markov network)등에 대한 연구가 깊이 진행 될 것으로 본다. 다음으로 음성언어처리(spoken language processing)연구이다. 음성신호 처리에서 얻은 지식뿐만 아니라 언어 처리에서 얻은 지식을 효율적으로 결합시키는 방법이 모색 되어질 것으로 본다. 또한 이들 간에 상호 보완을 위한 새로운 특징사용에 대한 방법등도 함께 연구 되어지리라 본다. 그리고 인식의 실시간 처리를 위한 알고리즘 개발에 대한 연구 등이 포함되며 회화체 음성에 관한 연구가 중심이 되어 Topic 위주의 시스템 개발에 더욱 박차를 가할 것으로 본다. 현재 1000 단어를 인식할 수 있는 시스템은 개발되어 있으나 수십만 단어를 인식할 수 있는 시스템은 아직 초보적인 연구단계에 머물고 있어 초 대용량 음성인식 기술에 대한 연구로 고속 검색 알고리즘 및 유사한 단어 사이의 변별력 향상을 위한 알고리즘에 대한 연구도 수행되어질 것으로 본다. 마지막으로 잡음 환경에서의 인식을 위한 연구가 병행하여 진행될 것이다. 이러한 음성인식 기술을 바탕으로 전화망을 통한 음성정보 검

색 시스템이 실용화될 것이며 지능망에서 음성인식 기술을 이용한 IP(intelligent peripheral)시스템 개발이 많아질 것이다. 현재의 조금씩 선보이고 있는 특정목적을 위한 음성인식 시스템의 개발이 더욱 증가될 것이고 자동통역 전화시스템 개발과 같은 장기적인 시스템 개발이 향후 계속해서 지속될 것이다.

V. 결 론

선진국에서는 국가가 주도하여 음성 인식 분야에 대한 연구를 활발히 하고 있다. 미국은 회화체 연속 음성인식에 주력하고 있으며 최근에는 초 대용량 음성인식 시스템 개발도 시작하고 있다. 중간 기술을 이용한 실용화에도 힘을 써 음성 다이얼링 서비스 및 요금부담 선택의 자동화 서비스 등을 개발하여 현재 사용중에 있다. 일본은 음성번역 통신연구소를 중심으로 자동통역 전화 시스템 개발에 주력하고 있으며 회사에서는 실용적인 시스템 개발을 수행하고 있다. 유럽은 범 국가 프로젝트를 중심으로 다국적 언어 인식이 가능한 실용적인 시스템 개발에 중점을 두고 있다. 이렇게 선진국의 예를 보더라도 앞으로의 음성인식 기술 분야는 응용 위주의 시스템 개발 즉 실용화가 가능한 Topic 위주의 인식 시스템개발이 활발히 진행되어 일상생활에 많은 영향을 줄 것으로 보며 단순히 음성 신호 처리의 측면을 벗어나 언어처리의 측면을 함께 고려한 회화체 중심의 인식 시스템이 주종을 이룰 것으로 전망된다. 현재의 인식 단계에 있는 시스템에서는 화자 독립성(speaker independence), 연속음성 인식, 대용량 단어 인식 시스템의 구성, 무제한 문법(unconstrained grammar)등등 많은 제약을 가지고 있는 실정이며 이 제약성을 극복한 시스템의 개발을 위해선 전자공학자, 언어학자, 전산공학자 및 심리학자 등의 공동 연구가 필수 불가결하다고 본다. 국내의 음성인식 기술분야는 기술력 뿐만 아니라 인력과 연구비가 외국의 경우에 비하여 볼 때 부족한 실정이지만 관련 업체나 국가 기관으로부터의 보다 많은 관심과 투자를 병행하여 이 연구 분야에 대한 연구를 지속적으로 전개 해야 할 것으로 본다.

참 고 문 헌

1. D.S. Pallett, "DARPA resource management and ATIS bench mark test poster session," Processings of the DARPA speech and Natural Language Workshop, pp. 49-58, Feb., 1991.
2. C. Delogu et al, "New directions in the evaluation of voice input/output systems," IEEE Journal on Selected Areas in Comm., vol. 9, pp.566-573, May 1991.
3. K.F. Lee, "Automatic speech recognition : the development of the SPHINX system," Kluwer Academic Publisher, 1989.
4. J. Takami and S. Sagayama, "A successive state splitting algorithm for efficient allophone modeling," Proceedings of Int. Conf. on Acoustics speech and Signal Processing, pp.573-576, Mar. 1992.
5. M. Ostendorf and S. Roukos, "A stochastic segment model for phoneme-based continuous speech recognition," IEEE Trans. on Acoust., Speech, Signal Processing, vol. 37, pp.1857-1869, Dec. 1989.
6. M. Lenning et al., "Flexible vocabulary recognition of speech," Proceedings of and Int. Conf. on Spoken Lang. Proceeding, pp.93-96, Oct. 1992.
7. G.G. Matison, "Emerging voice services in the NyNEX network," Proceeding of voice Systems Worldwide 1992 pp. 9-13, Feb. 1992.
8. S. Kuroiwa et al., "Architecture and algorithms of a real-time word recognizer for telephone input," Proceedings of and Int. Conf. on Spoken Lang. Processing, pp.1523-1526, Oct. 1992.
9. 구명완, "음성인식 기술의 현황과 전망," 정보과학회지 제11권 5호 pp.21-34, 1993.
10. 한민수, 정유현, 이항섭, "음성처리기술의 응용 현황 및 전망," 전자공학회지 제20권 5호 pp.542-547, 1993.
11. G. Elius et al., "Bellcore effects in applying speech technology to telephone network services," Int. Conf. on Speech Lang. Process., Kobe, pp.20.2.1-20.2.4, 1991.

▲김 순 협
제8권 5호 참조