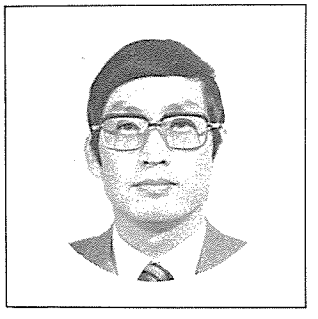
음성통신 및 신호처리 워크숍

音聲認識기술의 최근동향과 국내연구개발현황



殷 鍾 官
〈韓國科學技術院 교수〉

근래 컴퓨터, 반도체, 신호처리 기술등이 빠르게 발전함에 따라 인간과 기계가 말로써 대화를 할 수 있는 가능성이 이제 현실화 되게 되었다. 말은 통신의 가장 간편하면서도 빠르고 정확한 수단으로서 모든 기계를 말로 작동시킬 수 있다면 그 기술적인 파급효과는 물론, 경제적, 사회적인 면에서도 그 영향이 참으로 지대할 것이다. 이 기술을 음성인식기술이라고 하는데, 음성에 의한 man-machine interface 기술의 핵심이 된다.

음성인식기술은 1970년초부터 연구가 진행되어 많은 업적이 이루어졌지만 현재까지 인간과 비슷한 능력을 지닐 수 있는 인식기술은 아직 개발되어 있지 않다. 그러나 최근 10년간의 기술개발로 미국, 일본 등지에서는 격리단어를 인식할 수 있는 상업용 제품이 나와 있으며 연속단어를 인식할 수 있는 시스템들도 연구실에서 개발이 되고 있다. 특히 대용량 단어로 이루어진 연속단어 인식시스템은 음성학, 언어학, 컴퓨터기술학, 음성신호 처리기술 등이 집합이 되어야만 개발될 수 있는 첨단기술분야로서 실용화될 수 있다면 그 응용범위는 무궁무진하다고 할 수 있겠다.

본고에서는 이러한 음성인식기술변화의 추이를 살펴보고 현재 우리나라에서 개발되고 있는 음성인식기술을 재조명시켜 음성인식기술발전의 토대로 삼겠다. 본론에서는 최근 10년간에 이루어진 음성인식 기술, 특히 대용량 단어로 이루어진 연속 음성인식시스템의 핵심기술을 분야별로 나누어서 서술하였다. 이들 분야로서는 음성의

특징을 추출하는 음성분석기술, 분석된 특징으로부터 음성을 인식하는 인식기술, 기존의 음성인식시스템이 새로운 화자에게도 사용할 수 있게 하는 화자적응기술 및 인식대상이 되는 언어를 분석 처리하여 연속음성인식을 가능하게 하는 언어처리기술등이 있다. 또한 이러한 분야를 통합한 인식시스템개발 상황등을 기술하였으며 인식시스템 성능평가에 사용되는 데이타베이스(DB) 구성 현황에 대하여도 고찰하였다. 또한 국내의 음성인식 기술개발현황을 학계와 연구소로 나누어서 개발사례, 연구계획 등을 재조명하였으며 마지막으로 결론을 맺였다.

음성 인식 기술

음성의 분석

음성인식을 위한 음성의 특징추출 연구는 지난 수십년에 걸쳐 이루어져 왔으며 그 결과 현재 여러가지 음성특징 파라메터가 개발되어 음성인식기에 적용되고 있다. 음성신호로부터 적절한 음성특징을 추출해 내기 위한 연구는 크게 세가지 분야에서 이루어져 왔는데 이는 음성의 발성과정을 모델링하는 방법, 음성의 인지과정을 모델링하는 방법, 그리고 음성신호 자체를 주파수영역에서 해석하는 방법이다.

첫번째로 발성과정을 모델링하는 것은 사람이 음성을 발생시킬 때 각각의 발음이 성대로부터 성도를 거쳐 입술에까지 이르는 과정을 모델링하고 이로부터 특징 파라메터를 구하는 방법이다. 이러한 방법중 가장 대표적인 방법이 all-pole모델로 모델링하는 것으로써 이로부터 얻어지는 특징 파라메터가 잘 알려진 linear predictive cooling (LPC) 계수이다. 이 LPC 계수는 모음에 대해서는 비교적 정확히 모델링하지만 자음이나 비음등에 대해서는 성능이 저하되는 단점이 있다.

이러한 단점을 보완하기 위해 pole-zero 모델링을 적용하기도 하는데, 이는 모델에 zero를 추가함으로써 비음과 같은 경우를 보다 정확히 표현할 수 있는 반면 계산량이 증가하게 된다. LPC계수에 의한 특징벡터들 사이의 distance는 잘 알려진 측정방법으로 Itakura-Saito measure를 들 수 있는데, 이는 계산량이 많고 잡음에 민감하다. 이와 다른 방법으로 LPC계수로부터 cepstral 계수를 구하고 이에 적절한 weighting을 가하여 이를 특정벡터로 사용하면 Euclidean distance로 특징벡터 사이의 거리를 측정할 수 있고 또한 잡음에도 robust한 음성특징을 구할 수 있음이 밝혀졌다. 그외에도 LSF(Line Spectral-pair Frequency)로 음성특징을 표현하고 이를 음성인식에 이용하는 경우도 발표되었다.

두번째로 인지과정을 모델링하는 것은 발음된 음성에 대하여 사람의 귀가 어떻게 그 특징을 구별해 내어 이를 두뇌(brain)로 전달시키는 지를 연구하고 이를 토대로 음성인식에 유용한 특징을 추출하는 것이다. 이러한 방법중 가장 간단한 방법이 filter-bank 출력을 이용하는 것이다. 이때 이 filter-bank의 각각의 주파수 대역폭은 사람의 귀의 주파수 인지 특성을 실험하여 결정하며 실험 결과 비선형의 주파수대역으로 분할하는 것이 인식 시스템의 성능향상에 도움이 되는 것으로 밝혀졌다. 또한 청각기관의 인지 과정으로부터 얻은 지식을 이용하여 더욱 정밀한 음성특징을 추출하는 방법이 있는데, 이를 auditory model에 근거한 특징 추출방법이라 한다. 이 방식은 계산량이 증가하는 단점에 비해 인식율의 향상이 크지 않은데, 이것은 아직 청각기관의 인지과정에 대한 정보를 정확히 이해하지 못하기 때문이다.

세번째로 음성신호 자체를 주파수영역으로 변환하고 이의 시간적인 변화를 관찰하여 이로부터 유용한 음성특징을 추출해 내는 방법이 있다. 이 방식은 수많은 음성샘플의 spectrogram을 관찰하고 이를 실제 발음학상의 음운기호와 비교하여 적절한 음성 특징을 결정하게 된다. 이러한 음성특징으로는 에너지 정보, formant 정보, time du-

이 글은 지난 8월24일 한국음향학회가 주최한 「음성통신 및 신호처리 워크숍」에서 발표된 것으로 이황수 · 구명완 · 구준모 · 김회린 연구원들과의 공동연구를 요약한 것임. 〈편집자 註〉

ration 정보 등 여러가지가 있는데, 상당기간 연구되어 왔음에도 불구하고 좋은 인식성능을 보여주지 못하고 있다. 이는 앞에서와 마찬가지로 이해의 부족과 이러한 음성특징을 효율적으로 제어하기에 어려움이 있기 때문이다. 그러나 이러한 음성특징이 앞서의 모델에 근거한 음성특징이 지니지 못한 특성을 보완적으로 표현해 주기 때문에 이를 적절히 활용하면 보다 성능이 우수한 인식 시스템을 구현할 수 있음이 입증되었다.

지금까지 설명한 세가지 방향에서의 음성특징들은 각기 독립적으로는 음성신호에 포함되어 있는 특성을 충분히 표현해 주지 못하기 때문에 실제로 음성인식기에 적용될 때에는 이들을 상호보완적으로 결합시키는 것이 바람직하다. 이러한 음성특징들은 음성 신호의 어느 한 순간에서의 특성을 표현해 주므로 이를 instantaneous feature라고도 말한다. 그러나 사람이 음성을 인지할 때는 어떤 순간에서의 특징뿐만 아니라 음성특징의 시간에 따른 변화과정에도 큰 영향을 받기 때문에 이 정보를 표현해 주기 위하여 dynamic feature를 함께 사용하는 것이 바람직할 것이다. Dynamic feature를 사용했을 경우의 성능향상에 관한 연구보고가 그동안 많이 발표되어서 이를 많은 인식시스템이 활용하고 있다.

음성특징 파라메터를 인식시스템에 적용시킬 때에 continuous한 값을 그대로 사용할 수 있지만 이를 discrete한 symbol로 변환하여 처리할 수도 있다. 이와같이 discrete한 symbol을 사용하면 음성신호의 정보량을 감축시키게 되고 결국 전체 계산량을 크게 줄일 수 있는 이점이 있다. 이러한 변환 알고리즘중 대표적인 방법이 vector quantization (VQ) 방식으로서 음성신호의 부호화나 영상처리 분야에서 널리 이용되고 있다. 그동안 꾸준히 연구가 이루어져 total distance를 최소화시킬 수 있는 알고리즘들이 개발되었는데 이중 대표적인 것이 K-means 알고리즘이다. 최근의 연구결과 음성인식에 있어서 total distance를 최소화시키는 것이 가장 우수한 인식결과를 보여주지는 않고 대신 classification rate을 높일 수 있는 VQ 알고리즘이 더 낳은 인식성능을 보여준다는 연구보고가 있었다. 이 방식은 인간의 두뇌속에서의 정보처리 과정을 인공적으로 모델링한 neural network 이론을 적용한 것으로서, Kohonen의 self-organizing feature map 알고리즘을 Bayesian optimal classification rule에 접근하도록 수정한 LVQ(Learning VQ) 알고리즘이다.

마지막으로 음성특징 추출에 있어서 고려할 사항은 발음 환경에 관한 문제이다. 이를 위해서는 실제로 개발한 인식시스템을 어떤 주변환경 속에서 사용할 것이냐에 따라 그 환경에 영향을 받지 않는 특징 파라메터를 추출해 내어야 할 것이다. 이러한 잡음에는 white noise와 어떤 특정한 colored noise가 있을 수 있는데, 이러한 잡음의 영향을 받지 않기 위해 전처리 과정에서 미리 filtering 시키는 방법과 이러한 잡음에 robust한 음성특징을 찾아내는 방법이 있을 수 있다. 또한 인식될 음성이 전화선과 같은 특정 channel을 통할 경우에는 그 channel에서의 주파수 왜곡이나 기타 잡음이 음성특징에 영향을 끼칠 수 있으므로 이를 보상해 주기 위한 연구가 수행되어 왔다.

음성의 인식

앞에서 살펴본 것 처럼 음성을 분석하기 위한 기술이 발전함과 더불어 기계가 인간의 음성을 인식하는데 필수적인 음성인식방법도 매우 큰 진전을 이루었다. 즉, 적은 수의 어휘가 격리 단어의 형태로 한 사람의 화자에 의해서 발음되는 경우에만 효과적인 음성인식을 할 수 있었던 과거와는 달리 많은 수의 어휘가 연속음성의 형태로 화자의 구분없이 자연스럽게 발음되는 경우에도 음성인식이 가능하도록 하는 연구가 진행되고 있다.

이를 위하여 많은 기술들이 도입되어 음성인식에 효과적으로 사용되고 있다. 예를 들면, 많은 수의 어휘를 인식하기 위해서 벡터 양자화 기술, 새로운 인식 단위의 선정 방법, 시간감축 방법, 사전구성 방법등이 연구되었다. 자연스러운 연속음성을 인식하기 위해서 적은 수의 인식단위를 이용하여 문장단위의 음성을 모델링하는 방법, 사용하는 언어의 문법이나 구문등을 이용하는 방

법, 자연스러운 음성에서 나타나는 다양한 음가의 변화를 수용할 수 있는 인식단위의 선정방법 등이 연구되었다.

다양한 화자의 음성을 인식하기 위해서는 다수의 표준 패턴을 선택하는 방법과 새로운 화자에 대하여 인식시스템을 적응시키는 화자적응 방법, 화자 독립적인 음성특징의 추출방법등을 연구하였다. 이상에서 언급한 것 이외에도 여러가지 방법이 음성인식에 이용되었고 음성을 인식하기 위한 기본적인 접근방식도 많은 발전을 이룩하였으며 새로운 방법들이 제안되었다.

화자 적응

음성인식시스템은 인식 대상에 따라 특정 화자의 음성만을 인식하는 화자 종속시스템과 화자에 관계없이 비슷한 인식율을 얻을 수 있는 화자독립 시스템으로 나눌 수 있는데 현재의 기술로는 화자종속시스템의 인식율과 비슷한 화자독립시스템은 개발되어 있지 않다. 그러나 화자 종속시스템은 화자가 바뀔때마다 특정 화자에 맞게끔 새롭게 화자 종속시스템을 구성하여야 하기 때문에 많은 데이타와 시간이 필요하게 되고, 그러한 이유로 다양한 화자의 음성을 인식하여야 하는 분야에서는 별로 효용이 없다.

화자적응이란 화자종속시스템을 화자가 바뀔 때 마다 특정 화자에 대하여 새롭게 화자종속시스템을 재구성하는 것이 아니고 기존의 시스템정보를 충분히 이용하여 새로운 화자의 음성정보를 약간만 사용하여서 화자종속시스템의 성능과 비슷하게 하는 방식이다. 실제로 화자적응이란 완벽한 화자독립시스템을 만들 수 없으므로 화자종속시스템을 화자에 독립적으로 사용할 수 있도록 하는 방식이지만, 동일한 화자라도 화자의 건강상태, 기분, 또는 발음할때의 주변환경에 따라 음성이 다양하게 변하기 때문에 특정화자에 맞게 구성된 음성인식 시스템을 이러한 조건에 따라 시스템 내부의 정보를 바꾸어 줄 때에도 이용될 수 있다. 이러한 생각은 인간이 새로운 환경, 혹은 새로운 화자의 음성에 적응하면서 음성을 인식한다는 사실과 유사하다.

현재까지 화자 적응을 위하여 개발된 알고리즘은 인식시스템에 화자적응만을 위한 모우드의 존재 여부에 따라 정적 적응(static adaptation), 동적 적응(dynamic adaptation)으로 나눌 수 있으며 화자적응시 사용되는 음성데이타의 내용을 시스템이 미리 인지하고 있는지에 따라 지도적응(supervised adaptation), 독자적응(unsupervised adaptation)으로 구분된다.

정적 지도적응 알고리즘은 화자적응만을 위한 특정한 모우드에서 미리 지정된 어휘를 새로운 화자가 발음하면 시스템내에 저장된 정보를 새로운 화자의 특징에 맞게끔 화자 적응시키는 알고리즘으로서 가장 많이 연구되어 왔다. 이러한 이유는 정적 지도적응 알고리즘이 화자적응을 위한 알고리즘중 가장 기본적이며 이 알고리즘의 성능향상이 우선적으로 이루어져야 다른 알고리즘의 성능향상을 기대할 수 있기 때문이다.

언어 처리

음성인식기가 처리할 대상이 문장인 경우에는 언어의 문법 및 의미 정보를 이용함으로써 시스템성능을 향상시킬 수 있다. 일반적으로 문장 음성인식에 있어서는 음성학적 처리에 의한 인식오류를 수정하기 위해 상위 레벨정보인 운율 특징(prosodic feature)과 구문론(syntactics)에 의거해 문장구조를 결정하고 계속해서 의미론(semantics)과 실용론(progmatics)에 따라 최종 인식결과를 얻게 된다. 음성이 입력되면 먼저 음성학적 해석을 한 후에 문장론과 구문론 등에 입각하여 언어학적 처리가 수행되는 경우를 bottom-up approach라 하고 이와는 반대로 언어학적 처리의 결과를 토대로 필요한 음성학적 해석을 하는 방법을 top-down approach라고 하는데 일반적으로 두가지 방법을 병행해서 사용한다.

〈그림-1〉는 음성인식시스템중 상위레벨의 지식을 처리해 주는 부분을 보여주고 있다. 이 그림에서 구문분석은 특정한 순서로 배열된 단어가 문법적으로 맞는지를 검토하고 문장내에서 특정한 위치에 올 수 있는 단어의 종류도 추정하는데 이용된다. 여기서 문법이란 언어학, 자동이론 및

프로그래밍 언어 등에서 사용되는 올바른 언어를 표현해 주는 규칙의 집합이다.

이러한 구문규칙의 구현방법으로는 phase-structure rule을 사용하는 방식, finite-state 혹은 Markov model을 사용하는 방식, ATN(augmented transition network)을 이용하는 방식 및 production rule을 사용하는 방식 등이 있다.

다음으로 의미 분석은 구문론적으로 올바른 문장이 실제로 의미가 있는지를 결정해 주는 지식을 적용하는 부분으로 화자가 의도하는 의미를 논리적으로 분석한다. 실용분석은 대화를 통하여 문장을 이해할 수 있는 능력을 제공해 주는 부분으로 같은 문장이라도 서로 말하는 사람의 신분과 알고 있는 지식 및 이전에 대화하였던 내용에 따라서 의미가 달라질 수 있기 때문에 언어학적 처리부에서 이러한 정보를 사용하는 것이 바람직하다.

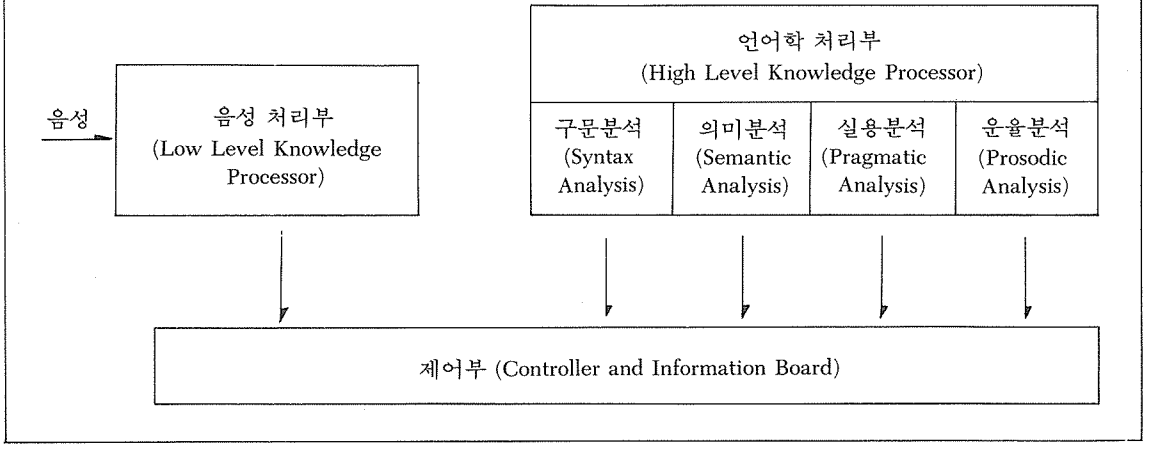
마지막으로 운율분석이란 강세(stress), 억양(intonation), 정지(pauses) 및 발음속도(timing structure) 등의 분석을 의미하며 이러한 운율정보를 사용하면 음성 이해력이 향상될 것이라는 가정은 언어처리 연구의 초창기부터 연구되어 왔지만 현재까지는 운율이 인식시스템에 성공적으로 구현되지는 못하였다.

언어학적 처리를 수행하는 방향에는 여러가지 방법이 있을 수 있다. 몇가지 대표적인 것을 보면, 문장의 왼쪽에서 오른쪽으로 순차적으로 해석하는 left-to-right method, island라 일컫는 문장의 keyword를 먼저 찾은 후 그 island로부터 오른쪽 혹은 왼쪽으로 해석하는 island-driven method, 그리고 위의 두가지 방법을 병행해서 사용하는 left hybrid method 등이 있다.

지금까지 기술한 일반화된 언어학적 처리부가 모두 필요한 인식대상영역은 natural task이다. 그러나 실제로 구현되어 사용되고 있는 많은 시스템은 언어학적 처리부의 극히 일부만을 사용하는 artificial task를 task domain으로 설정하고 있다. 이 artificial task로서 대표적인 것으로는 new Raleigh language, airline information and reservation language 등이 있다. 또한 natural task로서는 business letter의 test, patent applications 등이 있는데, 여기서는 언어학적 처리를 지식에 토대를 두기 보다는 확률적 모델링을 이용하여 수행하였다.

일반적으로 artificial task는 상대적으로 적은 양의 인식대상어휘를 가지는 시스템에 적당하며 미리 정해진 문법에 의해서 언어를 설계하므로 구문분석을 통하여 단어의 오인식을 대폭 줄일 수 있다. 그러나 연구자가 직접 언어를 설계해야 하고 인식대상어휘수가 늘어남에 따른 복잡성의

〈그림-1〉 음성 이해 시스템의 구조



증가로 확장성이 없기 때문에 수천 단어 이상의 어휘에 대해서는 natural task로서 언어학적 처리부를 구성해야 한다.

국내 연구 현황

학계 연구 활동

국내에서의 음성인식 연구는 그 역사가 매우 짧아 약 7년에 불과해서 선진 외국에 비해 10년 이상 늦게 연구가 시작되었지만 국내 여러 대학과 연구소에서 꾸준히 연구를 지속한 결과 음성인식의 기본기술이 구축되었다고 할 수 있다. 본 절에서는 국내에서의 한국어 음성인식기술 연구 중 학계의 연구활동을 살펴보고자 한다.

한국어 음성인식기술에 대한 본격적인 연구는 한국과학기술원 통신연구실에서 한국전기통신공사의 지원을 받아 1984년부터 시작되었다. 1차년도에는 잡음이 섞인 음성의 개선과 음성발생모델로 부터의 성도계수 검출에 대하여 연구하였고, 격리단어 인식을 위해 vecor quantization과 matrix quantization을 적용하였다. 2차년도에는 격리단어 인식을 위해 시간정보를 고려한 finitestate VQ 알고리즘을 적용하였고, 연결단어 인식을 위해 단어를 유사음소의 연결로 생각하여 이에 DTW 알고리즘을 적용한 결과를 보고하였다.

또한 DTW 알고리즘에서의 계산량 축소방법과 dynamic reference updating 방법에 의한 화자독립성 연구도 수행하였다. 3차년도에는 한국어 음소인식에 적응 알고리즘인 recursive least squares(RLS) 알고리즘을 적용하여 실험하였고, 격리단어 인식을 위해 finite-state VQ 알고리즘과 HMM 모델링을 결합시킨 방법을 연구하였다. 또한 화자인식을 위한 기초연구를 수행하였고, DTW 알고리즘을 이용한 격리단어 인식기술을 hardware로 구현한 음성인식 전화기를 개발하였다.

4차년도에는 음소분할 연구를 위해 formant tracking 방법을 적용하고, HMM 모델링을 이용한 음소인식 알고리즘에 대해 연구하였다. 또한 대상어휘수가 200단어인 연결단어 인식 알고리즘에 대해 연구를 수행하였는데, 이 알고리즘은 level-building DTW 알고리즘에 문법적인 제한을 가한 것으로서 자동전화번호 안내시스템으로 구현되었다.

5차년도에는 어휘수를 대폭 확장하여 1200단어 인식시스템을 개발하였는데, 이것은 음소의 HMM 모델링을 기본으로 하는 화자종속 문장인식 시스템이었다. 이 시스템에서는 인식시간의 단축을 위해 rule-based 알고리즘을 적용하여 후보단어 수를 줄여 주는 처리과정을 부가하였고, 단어 단위의 인식결과로부터 문장전체의 의미적 인식율을 높여주기 위한 언어학적 처리부도 포함되었다.

또한 이와는 별도로 화자 독립성을 부여하기 위하여 음소모델의 화자 적응기법에 대한 연구와 음소의 자동분할을 위한 음소분리 알고리즘도 연구하였다. 마지막으로 6차년도에는 전년도에 개발된 1200단어 인식시스템의 음성학적 처리부 성능향상을 위한 연구와 화자적응 알고리즘의 개선 및 시스템과의 통합, 그리고 인식시간 감축 알고리즘의 개선 연구가 수행되었다.

과학기술원에서 현재 연구중인 대용량 단어 및 문장인식시스템의 전체적인 흐름도가 〈그림-2〉에 나타나 있다. 먼저 인식할 대상어휘가 결정되면 이로부터 각 단어의 발음사전이 구성된다. 발음사전은 음소의 연결로서 표시되며 이때 음운론적인 규칙이 고려되어야 한다. 이와는 별도로 단어의 연결로 이루어진 문장의 문법으로부터 구문해석기를 설계한다. 각 음소의 기준모델을 결정하기 위해서는 수십개의 phonetic balanced words로 부터 음소들을 분류해 내고 이를 사용하여 각 음소의 HMM 모델을 훈련한다.

한편 입력음성과 비교할 후보단어의 수를 줄이면 인식시간을 감축할 수 있는데, 이를 위해 음소군 분류 및 단어군 분류 알고리즘을 개발하였다. 이와같이 하여 후보단어가 결정되면 이 단어의 발음사전을 참조하여 음소모델의 연결로 이루어진 단어모델을 구성하고 이를 입력음성과 Viterbi scoring을 통해 likelihood를 계산한다. 인식대상 화자가 바뀌면 그 새로운 화자에 대해 음소모델

파라메터를 바꾸어야 하는데, 이를 위해 화자적응 알고리즘을 적용하고 있다.

한편, 서울대 전자공학과에서는 음성신호의 분석을 위해 음성신호의 프레임당 평균 진폭의 분포를 이용한 새로운 끝점 검출 알고리즘을 개발하였고, 비강 및 방사 임피던스의 효과를 기존 성도모델에 포함시킨 일반화된 성도모델을 결정하고 이 모델로부터 pole-zero 선형예측 모델을 유도하였다. 음소분류 인식을 위해서는 간략한 에너지곡선을 이용한 새로운 비음구간 검출알고리즘과 종성 내파음 검출알고리즘을 개발하였다.

연속 숫자음 인식에서는 DTW의 일종인 UELM (Unconstrained Endpoint Local Minimum)을 적용하였고, 고립단어 인식연구에서는 자동으로 초기화되는 K-means clustering 알고리즘, 교정학습의 HMM, 시간 압축을 사용한 HMM, 그리고 고립단어의 새로운 모델로서 분할 확률모델을 제시하고 인식실험을 통하여 기존방법들과 비교하였다. 또한 음소단위 인식을 위해 음성학적 지식의 응용과 신경회로망의 응용등을 연구하고 있으며, 소단위 인식의 기초로서 분할(segmentation) 알고리즘의 연구도 수행되고 있다.

연세대학교의 음향, 음성 및 신호처리 연구실에서는 최근 자동차 소음환경에서 선형예측 방법에 기초한 네가지 스펙트럼 matching 방법, 즉 log likelihood ratio, LPC cepstrum, spectral slope distance measure, weighted cepstral distance measure 등을 적용하여 숫자음 인식을 수행하였다. 또한 Itakura-saito distance measure를 이용하여 한국어 음소의 분리 및 인식에 관한 연구도 발표하였다.

광운대학교에서는 146개의 전국 DDD 지역명을 인식하기 위해 12차 LPC cepstrum계수에 시간정보를 가지는 multi-section VQ 알고리즘을 적용하여 화자독립 인식시스템을 구현하였다. 성균관대학교 전자공학과에서는 패턴매칭을 이용한 격리단어 인식에 관한 연구를 수행하였고, 최근에는 연속음성 인식을 위한 음소단위의 인식에 관심을 집중시키고 있다. 이를 위해 LSP 파라메터를 이용한 음소의 분할 및 분류연구를 수행하고 이를 단어인식으로 확장하는 연구를 수행하고 있다. 전국대학교 전자공학과에서는 LSP 방식에 의한 음소분석과 연결단어 인식에 대해 연구하였고, OSDP 알고리즘을 이용한 한국어 연속 숫자음 인식도 수행하였다.

경희대학교 전자공학과에서는 최근 전화선을 통한 화자의 검증을 위해 전화시스템을 모델링하고 이로부터 원음을 복원하여 검증을 하는 방법을 연구하였다. 영남대학교 전자과에서는 formant 분석을 통한 한국어 단모음 및 파열음에 대해 연구하였고, 이를 토대로 음소를 인식 단위로 하는 소규모 특정화자 인식시스템개발을 수행하고 있다. 그외 한국과학기술원 전산과에서는 spectral-peak-weighted binary spectrum을 이용한 한국어 비음인식에 관한 연구를 수행하였다.

지금까지의 학계 연구동향을 살펴보면, 음소단위나 단어 단위의 인식연구가 여러가지 알고리즘으로 구현되어서 그 결과가 발표되고 있지만 그 성능은 아직 높은 수준에 이르지 못하고 있으며 표준적인 음성데이타베이스가 구축되어 있지 않아서 연구결과에 대한 비교 평가가 제대로 이루어지지 않고 있다. 앞으로는 공통적인 음성데이타베이스의 구축과 이를 이용한 보다 깊이 있는 한국어 음성특성에 대한 연구 및 인식 알고리즘 연구, 그리고 언어학적 처리연구에 학계가 공동으로 연구 노력을 집중해야 할 것이다.

연구소

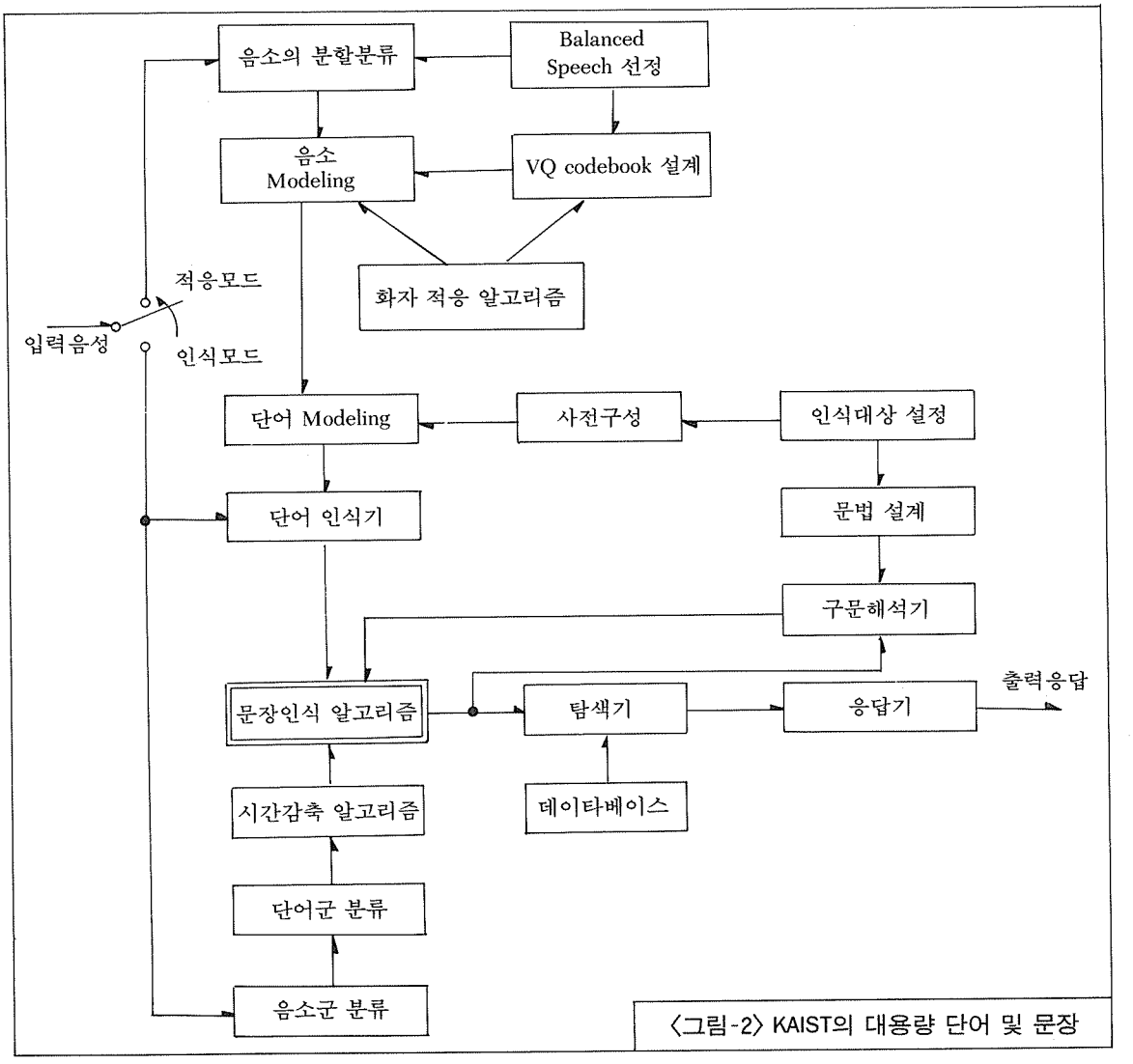
음성인식기술은 상품의 부가가치를 높일 수 있는 기술이므로 국가의 연구소 뿐만 아니라 기업에서도 연구소를 중심으로 하여 활발한 연구를 진행하고 있으나 아직 상품화단계에 이른 것은 매우 적은 형편이다. 우선 한국전자통신 연구소에서 진행되고 있는 음성인식에 관한 연구를 살펴보면, 그 동안 음소나 격리단어의 인식에 관한 연구를 수행하여온 전자통신 연구소는 최근에 대어휘 연속음성 인식을 위한 음소인식 기술을 개발하고 있다.

이 연구의 최종 목적은 phoneme-balanced word 음소단위 DB 구축과 음소단위에 의한 대어휘 음

성인식 기술개발에 있으며, 이를 위하여 단음절 음성 데이타베이스, 단음절 지속시간 분포조사, 어두 파열음의 분할 알고리즘 개발, 음소 특징추출을 위한 지각실험 시스템개발이 이루어졌다. 또한 음운 balanced 단어에 대한 DB가 구성중에 있으며 spectral의 회귀분석을 이용하거나 모음의 평균 패턴과의 상관을 이용한 단음절 segmentation 방법에 관한 연구가 수행되었다. 그리고 HMM을 이용한 연결단어인식시스템, Hopfield network을 이용하는 단모음 인식방법과 화자적응 방법에 관한 연구도 수행하는등 활발한 연구활동을 계속하고 있다.

한국전기통신공사의 연구개발단에서도 음성인식에 관한 연구를 수행하고 있는데, 이곳은 그동안 한국과학기술원에서 수행한 연구과제의 결과를 전수받고 이를 바탕으로 연구를 계속할 예정으로 있다. 전기통신공사는 음성인식기술을 이용하는 제품의 실용화보다는 당분간은 음성인식기술에 관한 기초연구를 수행할 것으로 알려져 있다.

기업의 연구소에서도 음성인식에 관한 연구가 진행되고 있으며, 그 중 대표적인 연구를 살펴보



〈그림-2〉 KAIST의 대용량 단어 및 문장

면 다음과 같다. 삼성종합기술원에서는 숫자와 지명을 포함하는 30여 단어를 화자 종속으로 인식할 수 있는 시스템을 개발하여 상품화를 추진하고 있다. 이 인식시스템의 특징은 잡음제거 알고리즘을 채택하고 있어서 잡음과 음성이 섞여있는 상황에서도 좋은 인식성능을 보여준다는 점이다.

금성사 중앙연구소에서는 지난 2년간 격리단어 인식시스템에 관한 연구를 수행하였다. 그 결과로 200단어를 인식할 수 있는 화자종속 음성인식시스템이 디지탈 신호처리 칩을 이용하여 PC plug-in board의 형태로 구현되었다. 현재는 이를 바탕으로 연결 숫자음을 인식할 수 있는 음성인식시스템을 개발하는 것을 목표로 하여 연구가 진행되고 있다.

마지막으로 주식회사 디지콤의 정보통신연구소에서는 화자 종속 또는 독립으로 최대 200단어를 인식할 수 있는 음성인식시스템을 개발하였다. 상공부에서 시행한 공업기반 기술개발사업의 일환으로 개발된 이 시스템은 특정한 화자가 최대 200단어까지의 임의의 어휘를 선정하여 인식을 수행할 수 있으며 복수의 화자나 임의의 화자도 사용할 수 있도록 지원하는 별도의 소프트웨어가 갖춰져 있다. 이 시스템은 여러가지 용도로 사용할 수 있으나 특히 컴퓨터를 키보드의 조작없이 음성으로 동작시킬 수 있는 음성 단말기로 사용될 수 있도록 설계되었다.

결 론

음성에 의한 man-machine interface 기술의 핵심이 되는 음성인식기술의 최근 동향을 대용량 단어음성인식시스템기술을 이루고 있는 분야별로 나누어서 기술하였다.

음성의 특징을 추출하는 음성분석기술의 최근 동향은 음성의 발생과정을 단지 모델링하여 음성의 특징을 추출하는 단계에서 벗어나 음성의 인지과정을 모델링하고 잡음이 섞인 음성으로 부터도 음성특징을 추출하여 높은 인식율을 얻는 방향으로 바뀌고 있으며, 추출된 특징으로 부터 음성을 인식하는 인식기술은 몇개의 단어를 인식할 때 적합한 DTW기술보다 대용량 단어인식에 적합한 HMM기술을 사용하는 추세로 바뀌어 가고 있다. 특히 최근에는 neural network를 이용한 음성인식기술이 개발되었으나 음소, 혹은 몇개의 단어를 인식할 수 있는 정도이며 기존의 HMM 기술과 결합시켜 대용량 단어인식에 사용할 수 있도록 연구가 진행되고 있다.

화자적응기술은 화자독립인식 시스템의 성능이 화자종속인식시스템에 비하여 떨어지기 때문에 그 대안으로 최근에 연구되는 기술인데 화자종속인식시스템을 기본으로 하여 새로운 화자가 발음한 짧은 음성을 이용하여 기존 시스템의 파라미터로 변경하는 방식을 채택하고 있다. 언어처리기술은 음성인식과정에 언어학적 지식을 이용한다면 인식의 개선을 이룰 수 있다는 전제하에 연구되어 왔는데 연속음성인식에 활용되고 있다. 이러한 기술등을 통합시킨 음성인식시스템은 여러 나라에서 개발되었는데 가장 높은 인식율을 나타내고 있는 음성인식시스템은 CMU의 SPHINX 시스템이다. 음성인식을 위한 DB의 기술은 국가의 주도로 표준화되어 개발되는 경향이 있는데, 개발된 DB는 인식시스템의 성능테스트에 이용된다.

한편, 국내의 연구동향을 살펴보면 학계에서는 KAIST가 대용량연결단어 화자종속시스템을 개발하였으며 그외는 대부분 단어 혹은 음소인식기술에 머무르고 있다. 연구소는 주로 상품화로 위한 연구개발과 기초 연구에 주력하고 있는 실정인데 주로 고립단어로 인식하는 수준이다. 그러나 최근 KAIST를 중심으로 음성정보연구센터가 운영되어 관련기술의 연구가 활성화 되고 있다.

이상에서 살펴본 것처럼 한국에서 한국인에 의해 개발되어야 한다는 특수한 환경아래 세계적인 인식기술에 비해 많이 뒤떨어진 한국어 음성인식기술이 부단히 발전하기 위해서는 관계 기관의 끊임없는 협조와 지원이 뒤따라야겠고 무엇보다도 관련기술을 갖고 있는 연구인들의 상호협조 및 공동연구 노력이 있어야 겠다.