

## 음성인식 기술동향



구명완 실장  
(한국통신 멀티미디어연구소  
음성언어연구실)

**최근** 미국의 시장조사회사인 가트너 그룹이 '99년에 주목할 만한 신기술 톱 10을 제안하였다. 이중 음성인식 기술은 인터넷폰, 디지털가입자회선과 케이블모뎀 등과 같이 중요한 기술로 인정되고 있다.

특히 사용자의 말만으로 컴퓨터를 조작하고 문서를 만들 수 있는 분야와 의학, 정보통신, 금융서비스 등에서 각광을 받을 것으로 예상된다고 한다.

음성인식 기술이란 음성 속에 내재되어 있는 언어정보를 자동으로 추출하는 과정이다. 음성에는 발화자의 감정, 신분, 성별 등의 정보가 내재되어 있다. 이중 발화자의 신분정도 즉, 누가 말을 하고 있는지에 대한 정보를 이용하는 기술이 화자인식 기술이며 메시지를 알아내는 기술이 음성인식 기술이다. 그러므로 화자인식 기술은 음성인식 기술의 하나의 분류의 포함될 수 있다.

이와 같이 음성인식 기술이란 사람을 “말”이란 미디어를 “글”이란 미디어로 바꾸어 주는 미디어 변환이라고 할 수 있으며 말과 글의 처리기술이므로 중요한 멀티미디어기술이라고 말할 수 있다.

음성인식에 대한 연구는 음성을 인식할 수 있는 로봇과 타이프라이터를 개발할 목적으로 수십년 전부터 수행되어 왔다.

음성인식에 관한 최초의 논문은 1952년에 발표된 미국 벨연구소의 숫자 음 인식기인 오드레이(audrey)에 관한 것이었다.

최초의 음성인식 기술을 상품화한 제품은 1980년 초에 출현하였으나 시각 장애인이나 초보적인 음성명령을 위한 것이었으며 본격적인 실용화가 되기 시작한 것은 1990년부터 시작되었다. 그러나 아직까지도 인간이 인식할 수 있는 수준의 음성인식 기술은 개발되어 있지 않다.

현재의 기술을 고려할 때 음성인식이 갖는 장점은 다음과 같다.

- 자연스러움 : 음성은 인간의 자연스러운 통신수단이므로 음성으로 기계에다 명령을 입력하는 것은 매우 쉽다. 타이핑을 하기 위해서는 어느 정도의 전문기술이 필요하나 음성을 입력하는 방법에는 전문기술이 필요 없다.
- 신속성 : 음성은 글로 쓰는 것보다 8~10배 정도 빠르며, 타이프라이터보다 3~4배정도 빠르게 명령을 입력할 수 있다.
- 동시성 : 사람이 눈, 귀, 손, 다리를 사용하여 다른 행동을 하고 있을 경우에도 음성으로 명령을 입력할 수 있다.
- 경제성 : 먼 지역에서도 마이크로나 전화를 통하여 음성으로 명령을 내릴 수 있으므로 명령을 입력하기 위하여 특별한 비용이 들지 않는다. 한편 현재의 음성인식 시스

템이 갖고 있는 문제점은 다음과 같다.

- 연음과 축약(reduction)현상 : 단어 혹은 문장 내에서의 음소는 이웃하는 음소에 따라 발음현상이 매우 달라진다. 특히 고립단어보다 연속음성인 경우 음소의 발음이 사라지는 현상이 있다. 이러한 문제점을 완벽히 고려할 수 있는 시스템을 만들기는 매우 어렵다.

- 음소를 정확히 분할하기 힘들다 : 음성은 손으로 쓰여진 글자와 정확히 대조되어야 하나 음소 구간을 정확히 시간 축에 대해서 분할하기는 어렵다. 무성자음(unvoiced consonant)은 어느 정도 분할이 가능하나 모음 등은 상당히 어렵다.

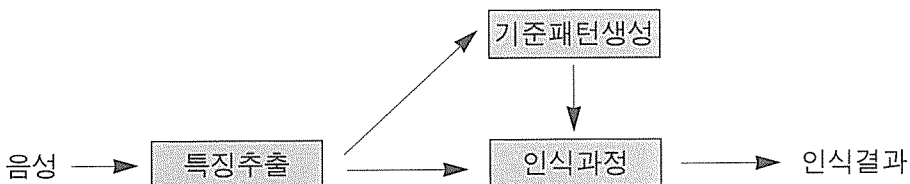
- 개인성 : 다른 사람이 동일한 단어를 발음하더라도 발성속도 및 발성기관에 따라 음성특징은 달라진다. 더구

나 어떤 음소는 다른 사람이 다른 음소를 발음했을 경우와 음성 특징이 동일할 수 있다.

- 언어정보의 부재 : 일반적으로 사람은 음성을 이해하는데 언어정보를 사용한다. 예를 들면 문법적으로 이해하기 어려운 음성을 발음하였다면 사람은 그 음성을 문법에 맞는 적당한 단어로 해석하는 경향이 있다. 이러한 해석을 구문(syntactic)해석이라고 한다. 유사한 방식으로 의미(semantic)해석이 있다. 이것은 문장의 의미가 파악되도록 음성을 이해하는 것을 의미한다. 그런데 현재의 음성인식 시스템은 위의 두가지를 정확히 구현하여 줄 수 없다.

현재의 음성인식 기본 구조는 (그림 1)과 같이 기본적으로 음성으로부터 음성 패턴(단

(그림 1) 음성인식 시스템 개념도



〈표 1〉 음성인식 시스템의 분류

	화자종속 여부	
	화자종속	화자독립
고립단어 인식시스템	상용화	상용화
연속음성 인식시스템 대화체 인식시스템	일부상용화	일부상용화

어, 음소 등)의 특징을 추출하여 기준 패턴을 만드는 훈련과정과 미지의 음성이 입력되면 저장된 기준패턴과 비교하여 가장 유사한 기준 패턴을 찾아내는 인식과정으로 나눌 수 있다.

이러한 알고리즘을 일반적으로 패턴 매칭(Pattern matching) 알고리즘이라고 부른다.

음성인식 시스템은 〈표 1〉과 같이 단어를 인식하는 고립단어 인식시스템과 연속적으로 발음된 문장을 인식하는 연속음성 인식시스템으로 크게 나누어질 수 있다.

연속음성 인식시스템은 연결단어(connected word) 인식시스템과 대화체음성(conversational speech) 인식시스템으로 세분화될 수 있다. 연결단어 인식시스템은 상대적으로 작은 단어를 매 단어마다 또박또박 발음하는 단어를 인식하는 시스템인 반면 대

화체음성 인식시스템은 상대적으로 대용량 단어를 매 단어마다 인식하는 것이 아니라 문장의 의미를 파악하는 것이다.

이와 같은 대화체 음성인식 시스템은 음성이해 시스템이라고도 불리며, 언어지식을 사용하는 것이 중요하다.

대화체 음성인식 시스템은 상당히 어렵기 때문에 문장에

서 필요한 단어만 선별하여 인식할 수 있는 단어선별(word spotting)인식시스템의 연구가 최근 활발히 진행되고 있다.

또한 음성 인식시스템은 화자종속 인식시스템과 화자독립 인식시스템으로 나눌 수 있다. 화자종속 인식시스템은 훈련이 된 특정화자만을 인식할 수 있는 시스템이고 화자독립 인식시스템은 어떠한 사람의 음소도 인식할 수 있는 시스템이다. 일반적으로 화자독립 인식시스템이 화자종속 인식시스템보다 훨씬 어려우나 이용 범위가 넓기 때문에 많이 연구되어지고 있다.

한편, 기준 패턴의 단위를 무엇으로 사용하느냐에 따라서 음성 인식시스템의 특징이 구분될 수 있다. 단어를 기준패

〈표 2〉 음성인식기술의 시장분야

분 야	응용사례
음성 명령, 통제 (command & control)	군사, CAD, 전자출판
음성인식 단말기	셀룰러전화, 장난감, 음성다이얼링 전화기, 가전제품
데이터 입력	물품분류, 감시
음성 타이프라이터	보고서 작성
화자 확인	보안, 출입 통제
전화음성인식	음성인식, ARS, 인터넷 음성인식

턴으로 사용하면 단위내의 연음(coarticulation) 현상을 고려할 필요가 없기 때문에 인식이 높은 반면 인식대상 단어가 많아질수록 메모리와 계산량이 증가한다는 단점이 있다.

또한 연속음성에 내재되어 있는 단어사이의 연음현상을 표현할 수 없다는 단점이 있다. 반면 기준패턴으로써 음소를 사용하게 되면 대상 단어가

늘 수가 있다.

음성명령 통제 시스템은 사용자가 정해진 단어 혹은 구문을 말하면 인식하는 시스템이며 컴퓨터의 입력으로 사용되거나 다른 전자장치를 작동하기 위해서 사용된다.

이러한 시스템은 컴퓨터를 이용한 설계시스템 혹은 전자출판기에서 키보드와 마우스 대신 음성을 통한 명령 입력을

에 따라 사용자는 지능적인 제어능력이 단말기에 부가되기를 요구할 것이다.

이미 대부분의 사람은 VCR을 프로그램하는 방식이 너무 복잡하다는 것을 인지하고 있다. 그런데 음성인식 기술을 이용하면 자연스러운 언어로 VCR 혹은 마이크로 오븐 등을 쉽게 조작할 수 있다.

이러한 지능적인 응용이 이

기술동향 · 음성인식 기술동향 · 기술동향 · 음성인식 기술동향 · 기술동향 · 음성인식 기술동향

늘어나더라도 계산량 및 메모리 사용량이 많이 증가되지 않으며 훈련과정 역시 간단하다.

또한 단어간의 연음현상을 쉽게 표현할 수 있다. 그러나 음소의 발음규칙이 명확히 해결되지 않았기 때문에 인식이 떨어지는 단점이 있다.

최근 이러한 현상을 극복할 수 있는 문맥종속음소(context-dependent phoneme)를 기준 패턴으로 사용하기도 한다.

이와 같이 음소를 기준패턴으로 사용하게 되면 훈련과정에 전혀 훈련이 되지 않는 어휘를 인식할 수 있다는 장점이 있다.

이러한 음성 인식 시스템을 단어독립(vocabulary-independent)인식시스템이라고 규정하고 있다.

음성인식 기술의 응용분야는 <표 2>와 같이 크게 6개로 나

가능하게 해준다.

최근에 일본 클라지온사는 “라디오를 켜라” 등의 운전자 명령을 수행하고 운전자에게 이야기를 걸 수도 있는 자동차용 음성명령 통제시스템을 개발하여 1,299달러에 미국에서 판매하고 있다고 한다. 특히 전화를 걸때도 음성명령이 가능하며 e-mail을 음성으로 알려주는 기능도 갖고 있다고 한다.

또한 벤츠의 BMW 뉴 7 시리즈에도 음성인식명령 기능을 채택하였다고 한다.

음성인식 단말기 분야는 전화기 등과 같은 단말기에 음성인식 기능을 추가하여 음성으로 발음한 전화번호 혹은 사람 이름 등을 인식하여 전화번호로 자동으로 다이얼링 해주게 하는 음성다이얼 전화기 등이 있다. 단말기가 더욱 복잡해

루어지기 위해서는 일상적인 환경에서 적절한 단어를 화자 독립적으로 인식할 수 있는 기술이 개발되어야 하며 싼 가격으로 IC칩화 되어야 할 것이다.

음성데이터 입력 장치는 음성인식 기술의 중요한 응용분야로써 산업체 등에서 눈과 손으로 제품검사를 하면서 화자가 있을 경우 음성명령을 가능하게 해주는 기능을 한다.

예를 들면 자동차 조립공장에서 페인트칠의 상태를 감시시키기 위하여 사용되는 보코렉트(Vocollect)와 같은 데이터 수집용 인식시스템 등이 대표적인 응용사례이다.

또한 우체국에서 소포, 편지 등을 자동으로 분류하기 위하여 음성명령을 이용할 수 있는 시스템이 개발되어 있다.

이런 분야의 인식시스템은

적용분야에 따라 비용이 달라지나 기본적인 시스템인 경우 매 인식 터미널당 2~4천 달러 정도이다.

현재 판매업자들은 좀더 낮은 가격, 짧은 훈련시간 및 설치경험 등의 향상을 위해서 노력하고 있다.

음성 타이프라이터는 대용량 단어를 인식하여 음성으로 문서를 작성할 수 있게 해주는

시스템은 전화가입자의 단순한 명령음과 숫자음을 인식하여 적절한 정보를 제공하여 주는 시스템이다.

이 시스템은 전화기의 푸쉬버튼을 사용하여만 명령입력이 가능한 기존의 자동안내 시스템(ARS:Auto Response System)을 대체할 수 있다. 이 분야는 음성인식 시스템의 시장 중 가장 빠르게 성장하고 있는 분야

“은행” 등의 이름을 말하면 자동으로 전화가 걸리는 기업체 음성다이얼 서비스가 서울지역 내에서 시범서비스가 진행되고 있다.

최근에는 인터넷의 활동이 일반화됨에 따라 외출중에도 인터넷의 정보를 알수 있도록 전화음성 인식 기술을 이용하여 다양한 정보를 검색할 수 있는 인터넷 음성인식 기술이

**기술동향 · 음성인식 기술동향 · 기술동향 · 음성인식 기술동향 · 기술동향 · 음성인식 기술동향**

소프트웨어 중심으로 개발되고 있다. 미국 IBM, MS 및 L&H 등의 회사가 음성워드 프로세서를 제품화하여 판매하고 있는데 100달러 이상으로 판매하던 소프트웨어를 대량판매하기 위해 최근에는 30달러로 할인 판매하고 있다고 한다.

화자확인 시스템은 미리 등록되어 있는 화자의 음성을 검색하여 새로이 입력된 음성이 등록된 화자의 목소리에 해당되는지의 여부를 알아내는 시스템이다.

이 시스템은 전화를 통하여 신용카드 소지자인지를 알아낼 수 있으며, 기존의 출입통제를 위해 사용되는 마그네틱 카드 대신에 사용될 수 있다.

전화망을 통한 음성인식 시

중의 하나이다.

현재 미국에서는 많은 지역 수신자 부담 전화(collect call)가 안내양의 도움 없이 자동으로 인식시스템에 의하여 이루어진다. 이러한 시스템은 “예”, “아니오” 등의 간단한 단어를 인식할 수 있다.

이와 같이 전화망을 통한 음성인식 기능을 첨가하는데 필요한 부가적인 비용이 전화 회선당 수백 달러 이하로 떨어지고 있다. 국내에서도 전화망을 이용한 음성인식 서비스가 선을 보이고 있다.

최근에는 음성인식 증권정보 서비스를 한국통신을 비롯한 회사에서 사용화 하였으며 특히 “1577”을 걸고 음성으로 “구청”, “영화관”, “대학병원”,

개발되고 있다.

현재 AT&T, Lucent Technology 및 모토롤러 등 3사가 VXML(Voice Extensible Markup Language)을 제안하여 월드와이드웹 컨소시엄(W3C)에 제안할 계획을 갖고 있다고 한다.

이 방식을 따르면 노트북, PC 등 인터넷 접속장치 없이 전화를 통해 인터넷의 전자우편, 날씨정보, 교통정보 내용을 검색할 수 있게된다.

현재의 음성인식 기술의 개발속도를 고려한다면 영화속에서 나오는 음성인식 제품이 꿈이 아니라 현실로 다가올 날이 그리 멀지 않을 것이다.