

# 음성정보처리산업 현황 및 전망

음성정보처리산업협의회

음성인식 기술은 앞으로 수년 내에 대용량 연속 음성 인식 기술이 보편화되고 안정적인 성능을 갖게 되어 이를 이용한 상용서비스가 급증할 것으로 보인다. 또한, 자연스러운 대화체를 인식할 수 있는 기술에 대한 연구도 상당히 진전하여 가상 현실, 대화할 수 있는 컴퓨터 등을 위한 인터페이스로 자리잡게 될 것이다.

60년대 기계번역의 실패로 시작된 언어처리 기술은 인간언어의 가장 큰 난점인 애매성을 해소하기 위해 모든 기술적 노력을 추구할 것이다. 70~80년대 기호/구조적 방식에 90년대 들어 각광을 받기 시작한 통계적 방법을 결합하는 혼합 방식의 시스템 연구가 주류를 이루었다.

90년대 들어 각광을 받기 시작한(사실은 60~70년대 방식의 재발견) 통계적 방식은 점점 대용량 코퍼스의 축적과 기계가독형 언어자원(사전, 시소러스 등) 증가에 힘입어 계속 대용량 통계처리를 통한 애매성 해소에 주력할 것이다. 여기에 통계시스템의 단점인 데이터 부족 문제 해결을 위하여 각종 인공지능 기계학습 이론의 활발한 도입이 예측된다.

응용분야는 과거 파일롯 시스템을 벗어나서 실질적으로 사용될 수 있는 데이터를 중심으로 점점 확대되어 고전적인 응용(ex.기계번역, 대화 이해)을 벗어나 웹과 결합된 새로운 응용분야(ex.질의응답, 자동 요약, 자동 분류)가 본격적으로 상업화 될 것이다. 또한 다른 분야와 접목되어 새로운 분야의 자연어 처리(ex.음성 번역, 음성 정보검색, 비디오 토픽 추출 등)가 본격적으로 시도될 것이며 다국어어를 위한 각종 방법론도 활성화되어 세계 메이저 7대 언어말고도 저밀도언어(low-density language)를 위한 각종 연구가 활발해질 것이다.

## 음성인식 관련 산업 연 95% 성장 예상

각국 주요 기관의 연구 동향을 참고할 때 향후 수년 내에 다음의 분야에서 연구 개발 성과를 보일 것으로 전망된다.

다양한 환경과 적용 범위에서 신뢰적인 성능을 보이는 '적응 기술', 이동형 컴퓨터/단말기 등에 적용할 수 있는 '음성 인식 기술', 대용량 연속 음성 인식을 위한 '음향 모델링 기술' 및 '고속 탐색 기술', 대화체 음성의 인식을 위한 '인식 기법'과 '언어모델 기술', 멀티 모달(multi-modal)에 기반한 음성 인식 기술 등이 그것이다.

현재 음성 합성 기술 수준은 낭독체 문장의 경우 상용화에 적당할 정도의 품질을 나타내고 있으나 아직도 인간 음성에 비해 부자연스러운 부분이 많은 실정이다.

현재까지도 합성음의 품질을 개선시키려는 노력이 계속되고 있으며 향후 3~5년 안에 대화체의 음성에 대하여 고품질의 합성음 생성이 가능하리라 전망된다. 이러한 전망을 이루려면 다음과 같은 연구가 더욱 필요하다.

대용량 코퍼스를 기반으로 한 합성방식이 품질 면에서 경쟁력을 갖고 있으므로 코퍼스 최적 설계기법, 합성단위 탐색 알고리즘 등에 대한 연구가 필요하며 더불어 신뢰도 있는 문장 전처리를 위하여 형태소 분석, 구문 분석, 문자-음소변환 등의 기술 발전이 필요하다. 또한 대화체의 자연스러운 음성을 합성하기 위하여 감성형 음성합성



기 개발이 필요하며 이를 위해 운용 모델링 및 운용 구현 방법에 대한 연구가 필요하다. 여러 가지 음색을 효과적으로 구현하기 위하여 음색 변환 기술에 대한 연구도 필요하며 다국어 지원을 위한 자연어 처리 분야의 체계적 연구도 음성 합성 분야 발전에 반드시 필요한 분야이다.

화자인식 방법 중에는 문맥종속형, 문맥독립형, 문맥요구형의 세 가지 방법이 있다.

그 중에서 문맥종속형 화자인식이 성능은 우수하지만, 비밀단어를 기억해야만 하고 사칭자가 녹음을 해서 사용할 수 있다는 단점이 있다. 그리고 문맥독립형 화자인식은 사용자가 임의의 문장을 발음하므로 편리한 점이 있는 반면에 사용자가 많은 데이터를 입력해야만 그 성능을 유지할 수 있다는 단점이 있다. 따라서 이 두 가지 방법을 보완하고 성능향상을 위해 문맥요구형(Text-Prompt) 화자인식 방식을 도입하고 있다.

각국 주요 기관의 연구 동향을 참고할 때 향후 수년 내에 다음의 분야에서 연구 개발 성과를 보일 것으로 전망된다.

음성으로부터 정확한 화자정보만 추출할 수 있는 기술 개발, 녹음과 도용으로부터의 보안 기술 개발(문맥요구형 화자인식 방법 개발), 주변 잡음에 강인한 화자 인식 시스템 개발, 시간의 경과에 따른 시스템의 성능 저하를 위한 화자적응 기술 연구, 유무선상에서 발생하는 채널 왜곡과 패킷 손실을 줄일 수 있는 기술, 시스템의 성능향상을 위한 다양한 방법의 개발, 다른 생체인식 기술과의 결합(지문, 얼굴, 홍채 등) 등이다.

## 2002년 월드컵 음성 번역 서비스 실시

자동음성번역 분야의 국제공동과제인 C-STAR III 에서는 여러 자동 음성번역 서비스 분야에서 전화망과 인터넷을 이용한 다국간 번역 시스템을 구현하려 한다. 이 자동번역 시스템은 전화 혹은 휴대폰을 이용하여 접속 가능한 서버기반 시스템으로 자동번역 시스템과 전화망 인터페이스를 갖추고 인터넷을 이용하여 다른 나라의 자동번역 시스템과 연계함으로써 자동 음성번역 서비스가 가능하도록 현재 연구 중에 있다.

〈표1〉 세계 음성정보처리산업 발전전망

(단위 : 백만불)

연 도		2000	2001	2002	2003	2004	연평균 성장율
음성 정보 산업	음성인식	4,899	11,623	21,496	36,822	71,802	95.0%
	음성합성	165	284	461	747	1,616	65.4%
	화자인식	292	306	385	501	602	19.7%
	계	5,357	12,214	22,343	37,071	74,020	90%

※ 출처 : TMA Report '99



자동 음성번역 분야의 상용화에 있어서는 미국의 VIA사와 일본의 NEC사에서 여행 안내를 위한 시스템을 개발하였다. 미국의 VIA사의 경우 영어권 사람들을 대상으로 한 휴대형 통역기를 개발하였다. NEC의 경우에는 노트북이나 휴대형 PC를 대상으로 한 자동번역 시스템을 개발하였으나 자연스러운 대화 환경에서 대화를 인식하고 번역하는데 많은 어려움이 있었던 것으로 알려져 있다. 또한 히다치에서는 삼성종합기술원과 공동 연구를 통하여 한일 자동통역시스템을 개발한 바 있다. 이 시스템은 일반전화나 이동전화로 별도의 통역 장치 없이 실시간 한일 자동통역이 가능하고 1500개의 문장을 인식할 수 있다. 이 시스템은 2002년 월드컵 기간에 본격적인 서비스를 위하여 인식 대상 어휘를 확장하고 인식률을 높이기 위해 현재 공동 연구를 수행 중에 있다.

최근 응용기술 분야는 미국을 중심으로 구체적인 응용분야가 개척되어 오고 있고 멀티모드/멀티미디어 환경 속에서의 다른 미디어와 통합에 관한 연구가 진행되고 있다.

향후 이러한 멀티미디어와 결합되는 연구가 더욱 활발하게 진행될 것으로 기대된다.

이러한 멀티모드/멀티미디어 기술의 활용분야로는 각종 멀티미디어 정보기기의 입출력 인터페이스, 카 네비게이션 시스템 개발, 시각 장애자를 위한 서비스 시스템, 대화형 자판기, 대화형 Robot, 3차 컴퓨터 시스템 개발, 제품의 검사, 멀티모드 의료 서비스, 각종 멀티모달 데이터 베이스 검색 멀티모드형 인터넷 검색기, 홈쇼핑, 자동 예약/문의 시스템, 음성 입출력 PC, 전자 메일 시스템 개발, 멀티모드형 자동항법 장치 개발, 키오스크(KIOSK) 개발 등 그 분야는 이루 헤아릴 수 없다. 이와 같은 응용연구와 더불어 자연어 처리기술을 적극적으로 이용하는 자연발화 대화체 연속음성 인식에 관한 연구가 더욱 활발하게 진행될 것으로 생각된다. 이와 더불어 각국간의 자동통역전화에 관한 연구도 가속화될 것으로 보인다.

음성인식 전반적으로는 현재의 통계적 방법을 기반으로 실제 대량의 음성 데이터에 기초를 둔 일상 언어의 언어모델을 구축하는 것, 다수화자의 음성데이터에 기저하여 개인차의 모델을 추출하여 이에 의한 다수 화자의 음성에의 적응화 알고리즘에 개발하는 것, 여러 종류의 잡음, 왜곡에 자동적으로 적응되는 방법을 확립하는 것 등이 중요한 기술적 과제로 될 것이다.

## 국내 기술 세계 음성정보처리 시장 5% 점유 목표

한편 UMS 메시징 서비스는 Ovum 자료에 따르면 1998년 말 현재 1만 7천개 정도에서 2006년에는 1억 7천만개 정도로 늘어날 것으로 보고 있다. 특히 미국과 유럽에 비해 상대적으로 늦게 서비스를 시작한 아시아가 전체의 27% 정도를 차지하게 될 것으로 예측하고 있다. <표2>에서는 2006년도의 전세계 UMS 메시징 서비스 예측을 나타낸 것이다.

〈표2〉 전세계 UMS 메시징 서비스 예측

연 도	America	Eruope	Asia-Pacific	Middle East & Africa	총 메일사서함
1998	62%	32%	6%	0%	17,000
2006	40%	28%	27%	5%	170,000,000

### 2004년까지 세계시장 5% 점유 목표

국내의 경우 음성정보기술개발, 음성정보 DB구축 등 음성정보산업 육성을 통해 2004년 3.5조원의 음성정보 처리 시장 창출 및 세계 음성정보처리 시장의 5% 점유를 목표로 하고 있다.

그 기본방향으로는 대화체 음성언어 기술개발 등 장기적이고 핵심적인 음성정보처리 기술개발은 선도기술개발 과제로 도출하고, 산업적 위험이 강하고 성공 시 파급효과가 큰 단기적인 기술개발 분야는 산업기반기술과제로 도출하여 중소·벤처기업을 지원하는 한편 산업기술개발분야는 음성정보처리산업협회가 주체가 되어 기획연구를 통해 도출하고, 산업체를 대상으로 지정공모를 통해 추진한다는 방침이다.

핵심기술은 대화형 음성언어기술, 음성번역 기술개발 과제 등을 선도기반기술을 통해 추진할 예정이며, 산업기술은 음성정보처리산업협회 주관으로 산학연 전문가가 모여서 산업기술개발 분야를 도출하고 지정공모 과제를 발굴할 것이다.

- 음성정보처리산업협회에서는 2001년도 산업기술개발과제수행 사업자로 구성된 음성정보처리기술협력위원회를 동 협의회 산하기구로 설립, 운영 계획(2월중 창립)
  - ▶ 목적 : 정부의 산업기술개발과제의 효율적 수행 및 국내 음성정보처리산업의 지속적인 발전 도모
  - ▶ 구성 : 정부의 산업기술개발과제 수행사업자(업체/지정, 자유공모 포함)
    - 위원장 : 김순협 사장(스피치사운드넷)
    - 부위원장 : 문석용 사장(인피니티텔레콤)
  - ▶ 추진일정 :
    - 2002년 1월 11일 : 음성정보처리기술협력위원회 발족 준비회의 개최
    - 2002년 1월중 : 구성 운영(안) 마련 및 가입 안내
    - 2002년 2월 중 : 음성정보처리기술협력위원회 창립회의
    - 2002년 3월 - : 사업추진