휴먼-컴퓨터 인터페이스를 위한 음성솔류션 시스템 개발

김 진영

전남대학교 공과대학 전자공학과
전화 : 062-530-0472 FAX : 062-530-0472
E-mail : kimjin@dsp.chonnam.ac.kr
WWW : dsp.chonnam.ac.kr

1. 연구목표

가. 장기연구목표

라이브러리 형태의 MMI(Man-Machine Communication)을 위한 음성 솔루션 개발

나. 단기연구목표

① 음성인식/합성/분석용 음성처리용 분석프로그램 개발(window95용)
② 음성합성기의 고품질화를 위한 요소기술 개발
③ 억양구단위의 음성인식 실험 및 억양구단위 경계 추출 알고리듬 개발
④ 음성인식 및 합성용 음성DB의 구축

2. 현재 수행중인 연구

본 연구실에서는 음성을 사용한 Man-Machine Interface(MMI)에 대한 연구를 꾸준히 진행하고 있다.

가. 음성인식과 음성합성을 위한 한국어 운율정보처리

음성합성시스템에서 운율정보는 음성이 가지는 정보 중 억양과 길이 그리고 세기에 해당하는 것으로서 합성기의 자연성 향상에 매우 결정적인 영향을 미친다. 한편, 최근에 이르러서는 운율정보를 음성인식에 적용하여 인식기의 성능을 향상시키고자 하는 연구가 매우 활발하다. 운율정보를 음성인식에 적용하면, 서로 유사한 문장들을 구분하는 것이 가능하고(ambiguity문제) 문장의 탐색(search)에서도 인식시간을 대폭 줄일 수 있는 장점이 있는 것으로 알려져 있다. 예를 들어 독일의 VERBMOBIL과 같은 실시간 독-영 음성번역시스템은 운율정보를 효율적으로 사용한 예이다.

본 연구에서는 억양구현 방법으로서 global tone과 segmental tone으로 나누어 Fujisaki모델과 어절톤 사전을 적용하는 방법을 개발하였으며, 음운환경에 따른 세기제어 규칙을 개발하였다. 또한, 운율정보를 음성인식에 적용하기 위하여 주어진 text와 음성신호로부터 TOBI(Tone Break Index)를 예측하는 알고리듬을 개발하여 음성인식에 적용하기 위한 방안을 연구중에 있다.

나. 음성합성기의 자연도 향상을 위한 억양의 계산적 모델

지금까지 음성합성연구는 주로 공학을 전공한 연구자들의 연구대상이었으나, 합성음의 품질이 한계에 다다름에 따라 좀더 근본적인 연구가 필요하게 되었다. 따라서, 본 연구실에서는 음성학을 전공한 언어학자와의 공동연구를 수행하고 있다. 언어학 전공자의 추상적인 이론과 개념을 구체적인 알고리듬과 모델로 정립한다. 본 연구에서는 통계적인 계산모델을 도입하여 억양을 모델링하고 이를 음성합성기에 탑재하여 그 성능을 평가한다. 통계적인 계산모델로서는 Breiman이 제안한 CART(Classification and Regression Tree)를 사용하고자 한다.

다. 입술정보를 이용한 바이모달 음성인식 알고리듬 개발

음성인식 시스템을 실험환경에서는 훌륭한 성능을 보이지만, 실제 적용환경에서는 주변잡음과 채널왜곡으로 인하여 인식률의 저하가 매우 심하다. 이러한 왜곡의 문제점을 음성신호의 개선(Speech Enhancement)기법을 사용하여 개선하고자 하는 연구가 있어왔으며, 이를 견인한 음성인식(Robust Speech Recognition)이라고 한다. 물론 여러 가지의 신호처리 기법이 동원되고 있다. 한편, 농아들이 사람의 입술모양만을 가지고도 음성인식을 할 수 있다는 사실에 아이디어를 얻어, 최근에 활발한 연구가 수행되고 있는 분야가 Bi-modal 음성인식이다. 즉, 입술정보와 영상정보(입술정보)를 이용하여 잡음이 있는 경우에는 적절한 음성인식결과를 보일 수 있도록 하는 것이다. 본 연구에서는 입술정보를 어떻게 분석할 것인가와 입술정보를 음성인식시스템과 통합하는 구체적인 방법에 대하여 연구한다.

3. 연구 실적 ('94년 이후)

가. 연구논문

■ 국제학술지

[1] Jinyoung Kim and Keongmo Sung, "Speech Recognition using Linear Prediction Considering Aural Characteristics," ACOUSTICA, Vol. 80, No. 4, pp.419-423, 1994.

[2] Suckwon Han, Jinyoung Kim, Keongmo Sung, "Extended Generalized Total Least Squares Method for the Identification of Bilinear Systems," IEEE Trans. on SP, Vol.44, No.4, pp.1015-1018, 1996, 4.

[3] Yongjin Choi, Jinyoung Kim and Keongmo Sung, "A Robust Algorithm of Total Least Squares Method," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E80-A, No.7, pp.1336-1339, 1997.

■ 국내학술지

[1] 김재인, 김진영, 이태원, "발음속도에 따른 한국어의 휴지기 규칙 및 평균음절길이 조절규칙," 한국음향학회지 14권 1호, pp.56-64, 1995년.

[2] 김재인, 김진영, 이태원, "GTLS의 ARMA시스템 식별에의 적용 및 적응 GTLS알고림듬에 관한 연구," 한국음향학회지 제14권 제3호, pp.37-48, 1995, 6.

[3] 한석원, 김진영, 성굉모, "Extension and Application of Total Least Squares Method for the Identification of Bilinear System," 한국음향학회지 제15권 1E호, pp.59-64, 1996, 3.

[4] 김진영, 최승호, "견인한 완전최소자승법과 시스템식별에의 적용," 한국음향학회지 제15권 4호, , pp.93-97, 1996, 8.
[5] 김진영, 최승호, "인지 LPC cepstrum의 새로운 구현 및 음성인식에의 적용," 한국음향학회지, 제15권 5호, pp.61-64, 1996.10.
[6] 김진영, 엄기완, 최홍섭, "Wiener filtering을 이용한 잡음환경에서의 음성인식," 음성과학회지 제1권 1호, pp.277-284, 1997, 4.
[7] 김기중, 김진영, 진달복, "한국어 글로벌 톤의 Fujisaki 모델링," 한국음향학회지, 제16권 제7호, pp.103-107, 1997.
[8] 김 진영, 엄 기완, "음성합성 시스템을 위한 음색제어규칙 연구," 한국음성과학회지, 제2권, pp. 25-44, 1997년 11월.
[9] 김 기중, 김 진영, 진 달복, "음운환경을 고려한 피치 기반 세기 제어 모델," 한국음향학회지, 제17권 1호, pp.80-85, 1998.

나. 수행과제

연구과제명	연구기간	지원기관
고품질 음성합성 방식 비교 연구	~95.10	한국통신
견인한 완전최소자승 알고리즘	~96.02	전남대
자동차 통신용 음성인식 알고리즘 구조 연구	~96.03	전남대
한국어 운율구조 연구	~96.12	한국통신
음성합성을 위한 음색제어 연구	~97.07	학술진흥재단
음성인식과 음성합성을 위한 운율정보처리	~98.06	정통부
한국어 음성언어처리 모듈개발	~97.12	한국통신
한국어 문서를 읽어주는 음성합성기 개발	~98.02	(주)송우정보

4. 연구 기자재

H/W		S/W
기자재명	수량	
Workstation(hyperSpac & UltraSpac)	2대	Multi-Speech(음성분석용)
X-terminal	1대	Waves/ESPS
IBM-pc(pentium)	4대	HTK 2.1
Laser Printer	2대	MATLAB PC/WS용
InkJet Printer(color)	1대	Segmentor(음성분석용)
DSP보드(PC용)	1대	
DSP보드및 인터페이스	1대	
Audio 시스템(Amp 포함)	1set	
Multi-ch. DAT	2대	
SHURE 마이크로폰	1대	
스캐너	2대	

5. 향후 연구 계획

가. 단기연구내용

① 음성분석 소프트웨어 개발
- 음성의 입출력이 기능
- 스펙트럼분석과 시간영역 정보(피치, 세기, ZCR)등 분석 기능
- 음소 및 변이음 등의 tagging 기능
- 피치 에포크 분석 및 음성구간 검출 기능
- 유/무성음 및 묵음 판별 기능

② 음성합성기 고품질화 요소기술연구
- Decision tree를 길이/세기 및 휴지강도 정보의 분석

③ Sub-word단위의 인식기 개발
- 연속음의 억양구단위 추출기 개발
- Tri-phone을 이용한 억양구단위 음성인식 실험

④ 음성DB 구축
- 기존에 구축된 음성DB 분석
- 운율분석용 음성 DB 녹음 및 분석(약 100문장)
- 음성인식용 DB 녹음(40문장 10인)

나. 중·장기 연구내용

① 음성 솔루션을 위한 음성합성기 개발 및 음성인식 기초연구(2차년도)
- 경쟁력 있는 양질의 음성합성기 개발(자연성 확보)
- 음성합성기의 이해도 향상을 위한 합성단위 연구개발
- Key word spotting 기법연구
- 한국어 tri-phone정의를 위한 실험 및 정의
- 음성인식용 DB 확충

② 음성솔루션용 단어단위 음성인식기 개발 및 합성기 업그레이드(3차년도)
- 새로운 음성합성단위에 의한 합성기 업그레이드 및 최적화
- 음성합성기 상품화
- Sub-word 단위의 무제한 단어 단어인식 알고리듬 개발
- Key word spotting 기법과 연계된 음성단어명령 프로그램 개발