특 집

대규모 공통 음성 DB 구축 현황

이 용 주*, 김 상 훈**

*원광대 전기전자 및 정보공학부, **ETRI 음성/언어정보연구센터

I. 서 론

음성인식이나 합성 그리고 음성번역 등으로 대표되는 음성정보처리기술의 개발에 가장 기본적인 연구자원이 음성 DB(또는 음성 코퍼스)이다. 자연어처리에서의 텍스트 코퍼스와 마찬가지로 음성의 경우에도 막강한 컴퓨팅 파워와 통계적 처리방법에 의존하는 소위 코퍼스 기반 접근방법이 주류를 이루고 있어서 음성 DB의 중요성 및 역할은 막중하다. 음성 DB의 양과 질이 개발하고자하는 시스템의 성능을 좌우한다고 할 정도이다. 본고에서는 음성 DB에 대하여 개략적으로 설명하고 국내외 동향을 간단히 소개한 후, 공동으로 이용하기 위해 구축하여 보급하고 있는 음성 DB의 현황을 소개하고자 한다.

II. 음성 DB의 개요

1. 정의 및 응용

음성 DB란 "음성관련 연구 개발자들이 언제든지 재사용이 가능하도록 부가적인 정보와 도큐먼트가 갖추어져 있으며, 컴퓨터로 읽을 수 있는 형태로 구성된 음성자료의 모음"을 말한다[1]. 그런 의미에서 보면 부가적인 자료가 불충분하고 컴퓨터로 읽기 어려운 아날로그 테이프 형태의 대량의 방송자료들은 이러한 정의에서 제외되나, 반면에 음성신호와 함께 수집된 발성에 관계되는 생리적인 신호(EMG, EGG 등) 등은 이 범주에 포함시켜 다루고 있다. 이러한 자료의 묶음을 음성 코퍼스(corpus) 또는 음성 DB라고 부른다.

이러한 음성 DB는 여러 가지 응용을 생각할 수 있는데, 크게는 연구용과 기술적인 응용(개발용)으로 나눌 수 있다. 연구용의 경우는 먼저 음성 그 자체의 생성, 전달, 지각 과정을 규명하고 그 언어적인 현상을 중심으로 한 음성학적인 연구, 음성언어를 통해 성별, 연령별, 지역별, 계층별 변화 및 방언 등에 관심을 둔 사회언어학적 연구, 언어의 심리적 현상을 다루는 심리언어학적 연구, 모국어나 제2외국어의 언어 습득 및 훈련에 관한 연구, 일반적인 언어학 연구, 청각학(audiology) 및 음성병리학적인 연구 등에 그 기본 연구 환경으로 쓰인다. 기술적 응용으로는 음성의 합성에 필요한 기본적인 합성단위의 추출 및 음운, 운율규칙을 위한 기본 자료로 쓰이며, 음성인식 및 화자인식의 경우에는 인식 알고리듬의 훈련 및 평가용으로 필수적인 자원이다.

2. 음성 DB의 구축을 위한 음성자료의 분류

음성 DB의 구축대상이 되는 음성을 편의상 다음과 같이 분류할 수 있다. 먼저 수집환경에 따라 분류하면 마이크로폰음성 및 전화음성으로 나눌 수 있다. 즉 주파수대역폭이 넓은 마이크로폰을 통해 입력하는 경우와 300 Hz~3400 Hz로 대역제한된 음성을 통신회선을 통해 전화기를 이용하여 수집하는 경우이다. 물론 마이크로폰을 이용하는 경우에도 수집 장소에 따라 실내소음이 있는 사무실음성과 소음을 제한한 방음실 음성(clean speech)으로 나눌 수 있다. 전화음성의

경우도 발성장소에 따라 사무실, 공중전화부스 등 다양한 소음환경이 있을 수 있고 통신회선을 통한 음성이므로 다양한 회선 잡음이 포함된다. 물론 사용하는 전화기세트에 따라서도 특성이 다른 음성이 얻어진다. 수집대상에 따라 분류하면,

1) 단어음성과 연속음성

주로 음절, 단어, 어절 등을 대상으로 한 단어음성과 문장을 대상으로 하는 연속음성이 있다.

2) 낭독음성(read speech)과 대화음성 (dialog speech)

주어진 목록을 그대로 읽는 낭독음성과 상황에 따라 주어진 목적을 달성하기 위해 주고받는 형태의 대화음성이 있다. 예를 들어 음성타자기나 구술시스템(dictation system)에 입력하기 위해 발성하는 경우가 낭독음성에 해당한다. 대화형의 음성명령에 의한 사람과 기계간의 인터페이스를 목적으로 한 대화시스템의 구현을 위해서는 사람간의 대화와 유사한 상황에서의 대화음성 수집이 필요하다. 기술수준의 제한으로 그 동안의 음성연구는 낭독음성이 주 대상이었으나 대화음성으로 발전하고 사람 간의 대화에 가까운 자유발화음성으로 발전하고 있다.

3) 정형음성과 자유발화음성

발성할 내용을 미리 준비하여 발성하는 이른바 정형화된 음성(낭독이건 대화건 간에)에 비해 발화 현장에서 준비 없이 즉시 발성하는 것을 자유발화음성(spontaneous speech)이라 한다. 음성언어정보처리의 궁극의 목표는 이러한 자유발화음성을 대상으로 한다.

4) 표준어음성과 방언음성

발성의 대상이 표준어인가 방언인가는 음성 DB의 활용처에 따라 다르겠으나 현재 공학적인 응용은 주로 표준어를 대상으로 한다. 그러나 표준어라 하더라도 글자형태에서만 표준어이고 발화된 상태에서는 개인차 및 지역차가 포함될 수밖에 없다. 불특정화자의 음성인식을 위해서는 이러한 개인차 및 지역차까지도 흡수할 수 있는 다양한 지역, 다양한 연령층의 음성이 수집되어야 한다. 아울러 방언음성의 수집 및 보존도 우리말의 연구 뿐 만아니라 문화유산의 계승 발전 측면에서 그 의의가 크다.

5) 인식용, 분석용, 합성용

분석용인가 인식용인가 또는 합성용인가를 구분하기는 쉽지 않고 어떤 면에서는 구별할 필요가 없는 경우도 있고 공통적으로 활용이 가능한 면도 많다. 다만 인식용의 경우, 발성하고자 하는 내용이 한사람이 발성하기에는 너무 많아 어려운 경우는 부분적으로 나누어 발성하되 참여하는 발성자의 수를 늘리는 경우도 있으나 합성단위의 추출을 위한 음성데이터는 동일한 발성자의 음성자료가 필요하다.

6) 우리말, 외국어, bilingual

최근의 음성연구 중 기계번역기술과의 조합을 통한 음성번역연구가 시도되고 있다. 이럴 경우 그 대상은 우리말만이 아니라 상대국의 음성도 그 대상이 된다. 실제로 각기 자국어의 음성으로 외국인과의 대화가 이루어지려면 그러한 상황을 모의하여 중간에 동시통역자를 중개로 한 음성대화가 수집되어야 하고 이를 바탕으로 bilingual 텍스트 코퍼스가 만들어진다.

bilingual 음성 DB가 초기에 직접적으로 필요하지는 않으나 음성번역 성능평가 등에 부분적으로 필요할 것이다.

3. 음성 DB의 구성

음성 DB는 단순히 음성을 기록하여 보존하는 것만이 아니라 어떤 음성이 어디에 보존되어 있는가 하는 색인정보도 가지고 있다. 따라서 지정한 단어 또는 문장을 바로 음성으로 들어볼 수도 있고 어떤 음소열이나 음운현상을 포함한 음성자료들 (예를 들면, "앞뒤에 유성음으로 둘러 쌓인 'ㄱ', 'ㄷ', 'ㅂ' 가 포함된 단어 또는 문장들을 모두 찾아라" 등)만을 임의로 검색해 볼 수도 있다. 또한 발성내용 이외에도 발성자에 관한 정보(성

별, 연령, 출신지 등)도 포함되어 있어 발성자에 따른 여러 음성현상들도 분석해 볼 수 있다. 이와 같은 검색이 가능하도록 하기 위해 음성언어학적인 여러 구분에 관한 부가정보를 부여하는 것을 레이블링(labelling)이라고 부른다. (*언어레벨의 경우는 태깅(tagging), 음성레벨의 경우는 레이블링이라고 부른다.) 레이블링의 단위로는 음소, 단어, 어절, 문장 등이 있다. 단어나 그 이상을 단위로 할 경우는 비교적 큰 문제는 없지만 음소 이하의 단위로 레이블링을 할 경우는 시간적으로 연속된 파형 상에서 그 구분(segmentation)을 정하는 것이 쉽지 않다. 따라서 연구자들 간에 공통적으로 사용할 수 있도록 일정한 기준을 마련해 두어야 한다. 또한 음운정보 만이 아니라 운율정보(F0의 변화 등 억양정보)를 부여한 음성 DB도 있다.

Ⅲ. 외국의 관련기관 및 활동

선진 각국에서는 자국어 음성 DB에 대한 체계적인 구축이 음성 및 언어처리기술 확보를 위한 가장 기본적인 연구 환경임을 깊이 인식하고 이에 대한 체계적인 확보가 공공연구기관을 중심으로 활발히 추진되었다. 특히 산학연이 협력하여 그 결과물들을 공유하기 위한 컨소시움 형태의 관련 조직들이 구성되어 보급이 원활하게 이루어지고 있다.

1. 미국의 LDC(Linguistic Data Consortium)

LDC는 대학, 기업, 정부연구기관의 컨소시움 형태로 구성된 조직으로, 음성언어에 관한 연구에 필요한 각종 코퍼스를 공유할 수 있도록 서비스하고 있는 조직이다. LDC에서는 음성 및 텍스트에 관한 코퍼스, 사전, 그 밖의 공유 가능한 자원에 대해서 데이터의 수집, 작성, 배포를 담당하고 있으며 미국의 DARPA의 지원으로 1992년에 설립되었다. 현재 사무국은 펜실베니아대학교에 위치하며 필요한 사용계약, 지적재산권 등에 관한 업무를 담당하고 있다. LDC는 회원제로서, 회원에 등록되면 필요한 음성 DB 및 텍스트 코퍼스를 제공받을 수 있다. (참고 : http://www.ldc.upenn.edu)

2. 유럽의 ELRA(European Language Resources Association)

ELRA는 유럽에서의 음성언어 관련 자원의 작성, 검증, 배포를 촉진할 목적으로 1995년에 설립되었으며 사무국은 룩셈부르크에 있다. ELRA는 비영리조직으로 회원제이며, 보급뿐만 아니라 음성언어자원의 개발을 지원하는 역할도 한다. ELRA에서는 음성 및 텍스트에 관한 코퍼스, 문법, 사전, 전문용어 등 공유 가능한 자원들을 모두 다루고 있다. 주로 EU(European Union)가 지원하는 연구개발 프로젝트에서 작성된 자원들을 모아서 배포하고 있다. (참고, http://www.icp.grenet.fr/elra/home.html)

3. 유럽의 EAGLES(Expert Advisery Group on Language Engineering Standards)

1993년 2월에 EU의 지원으로 구성된 전문가 그룹으로 음성 및 언어를 포함하여 언어공학적인 응용을 위한 각종 spec의 가이드라인을 제시하는 것을 목적으로 하며 산하에 text corpora, lexicon, formalisms, assessment, speech 등 5개의 각종 워킹그룹이 있다. 그리고 speech 워킹그룹 안에 음성 DB 관련 서브그룹이 활동한다. 이 그룹의 노력으로 음성언어시스템 관련의 표준 및 resource에 관한 핸드북이 발되었다[1].

4. COCOSDA(International Coordinating Committee on Speech Databases and Speech I/O Systems Assessment)

COCOSDA는 음성입출력시스템의 개발과 평가를 위한 음성언어데이터의 수집과 관련된 가이드라인과 방법의 연구를 주제로 정보교환과 논의를 위한 국제적인 협의체로서 음성언어와 관련한 여러 연구그룹과 각국의 조직, 국제적인 표준화 기구 간의 연락 역할을 한다. COCOSDA는 1991

년에 발족하여 CCC(Central Coordinating Committee)라는 중앙위원회와 데이터베이스(코퍼스), 음성합성 평가법 및 음성인식 평가법 등 3개의 워킹그룹이 있고, 미국, 유럽, 아시아의 3개 지역을 주체로 하여 구성되어 있다. 각 지역별로 지역 COCOSDA가 구성되어 있으며 1년에 1회꼴로 음성관련 국제회의 개최시에 워크샵을 통해 활동보고, 토의 및 정보교환이 이루어지고 있다. 아시아 지역은 우리나라, 일본, 중국, 대만, 홍콩 태국, 싱가폴, 인도 등이 참여하는 oriental COCOSDA가 결성되어 활동 중이며 1998년 5월 일본 쓰쿠바를 시작으로 타이페이(1999), 베이징(2000), 대전(2001), 태국 와힌(2002) 등을 돌며 매년 워크샵을 개최하였다 2003년에는 10월에 싱가폴에서 개최예정이다. (참고, http://www.itl.co.jp/cocosda)

5. 일본

일본의 경우는 그 동안 각 기관별로 또는 학회를 중심으로 음성 DB가 구축되어 왔으나 LDC나 ELRA와 같은 공유를 위한 컨소시움은 없었다. 최근에 "언어데이터공유계획"(약칭 LRSI)을 준비 중에 있으며 구체적인 활동은 아직 없다. 따라서 지금까지는 전자공업진흥회. 일본음향학회, ATR 및 각 대학교에서 만든 음성 DB를 주로 학회나 연구소를 중심으로 보급 활용하고 있다.

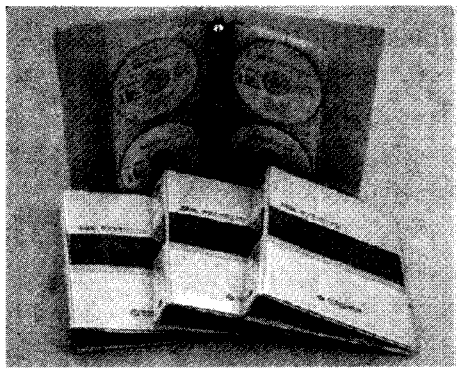
Ⅳ. 국내의 현황

선진 각국에서는 자국어 음성 DB에 대한 체계적인 구축이 음성 및 언어처리기술 확보를 위한 가장 기본적인 연구 환경임을 깊이 인식하고 이에 대한 체계적인 확보가 공공연구기관을 중심으로 활발히 추진되고 있다. 우리나라에서도 그 동안 학계, 연구소, 업계 등 기관별로 자체 연구를 목적으로 소규모로 개별적인 구축이 이루어져 왔으나 최근에는 공동으로 사용할 수 있는 우리말 음성 DB의 중요성을 인식하고 관련 산업의 육성 지원 차원에서 정부의 지원을 받아 체계적인 구축 및 공개가 이루어지고 있다. 소요되는 음성 DB의 방대한 양과 응용 분야에 따른 다양한 종류를 감안하여 현재는 두 기관이 응용영역에 따라 그 영역을 분담하여 상호 협조 하에 음성 DB를 구축하여 배포하고 있으며 그 내용은 다음과 같다.

1. ETRI 음성/언어정보연구센터 (http://voice.etri.re.kr)

1) 구축목적

음성처리기술을 개발하기 위해서는 대규모의 음성/텍스트 DB가 필수적으로 요구된다. 이에 대다수의 국내 음성처리업체들은 개별적으로 DB를 구축하여 자체 엔진개발에 활용하고 있으나, 일반적으로 DB 구축에 많은 시간과 비용이 소요되어 외국업체에 비해 기술개발 경쟁력이 떨어지고 있다. 특히 국내업체간 중복 DB 구축으로 인해 국가적으로 자원이 비효율적으로 활용되고 있어 음성정보처리 관련 사업자들의 공동 이익을 도모할 수 있는 공통음성 DB 구축이 시급하다. 이에 한국전자통신연구원(이하 ETRI) 음성/언어정보연구센터에서 추진하고 있는 정보통신부 출연 사업은 다양한 통신망환경에서 대규모 음성 DB를 구축 배포하여 국내 음성시장 활성화 및 국내 음성처리업체의 국제경쟁력을 강화할 수 있는 기반을 마련하는데 공통음성 DB 구축의 목적이 있다.



〈그림 1〉 배포용 공통음성 DB

〈표 1〉 1차년도 ETRI 공통음성 DB 구축 내용

순번	목적	통신망 환경	화자수/발화수	발성내용/발성조건	비 고
1	음성인식용 단어	휴대폰	1,000명/10만 발화	- 주식상장회사명, 지명, 인명, 상호명, 제품명, PC명령어, PDA명령어, 일반명사 - 전화망인 경우, 전화망 인터페이스 보드는 NMS 계열 및 Dialogic JCT 계열을 이용. "디지털보드 : 아날로그보드"="50 : 50" 비율로 수집. 유선전화기 사용을 유도하고, 무선전화기의 사용은 10% 미만이 되도록 함. 전화기 모델은 제한 두지 않음 - 남/녀 비율은 50 : 50으로 하며 최대 5%까지의 차이 허용. 연령별 구성은 "10대 : 20대 : 30대 : 40대 이상"의 구성비를 "20 : 30 : 30 : 20"으로 하고 오차는 각 5% 이하 허용. 지역별 구성은 "서울/경기 : 경상 : 충청 : 전라 : 제주 강원"의 구성 비율을 "40 : 20 : 15 : 15 : 10"으로 하고 최대 2%까지 차이 허용	
2		유선망			
3		VoIP			
4		마이크(중가)			
5		마이크(저가)			
6		헤드셋			
7	음성인식용 숫자	휴대폰	1,000명/10만 발화	- 번호독식 방식과 봉독식 방식에 대해 수집 - 전화번호, 주민등록번호, 계좌번호 등으로 구성된 1- 16연 숫자 - 전화번호, 계좌번호 중 '- '를 '에', '다시', '국'으로 발성 - 일부는 한자식과 우리말 숫자 혼합형으로 발성 - 봉독식 방식은 99,999까지의 무작위 숫자를 한자식으로 발성하고, 일부는 우리숫자로 발성	성별/연령/지역별 구성비율은 단어와 동일
8		유선망			
9		VoIP			
10		마이크(중가)			
11		마이크(저가)			
12		헤드셋			
13	음성인식용 낭독체/준낭독체 문장	VoIP	1,000명/10만 발화	- 낭독체 문장은 발성목록은 방송뉴스에서 추출 - 준낭독체 문장은 발성목록 없이 화자가 즉흥적으로 주어진 주제에 대해 발성(예 : 자기 소개하기, 가장 친한 친구 이야기 하기, 자신의 학교 소개하기 등)	"
14		마이크(중가)			
15		마이크(저가)			
16		헤드셋			
17	음성인식용 대화체 문장	유선/휴대폰 전화망	250명/2,500 대화	- 예약, 은행, 증권, 관광안내, 텔레쇼핑 등 최소 30개의 시나리오를 작성하여 그 중 10개를 선택 - 각 화자당 10개의 상황에 대한 대화음성을 실제 call center에서 수집 - 1대화당 평균 20문장 또는 5분 이상이 되도록 구성	"
18	언어모델링용 문장	텍스트	2,000만 어절 수동/4,000만 어절 자동	- 신문기사 대상 띄어쓰기 및 철자오류 검증 - 심볼, 영문 등을 한글로 변환	
19	음성합성용 문장	정보전달용 낭독체	남녀 1인/1만 문장	- 방송뉴스에서 추출한 낭독체 문장 발성 - 트라이폰 분포 고려	
20	화자인식용 단어/문장	휴대폰	250명/45만 발화	- 2연, 4연 숫자음 및 10개의 질문에 대한 단답형 대답과 10개의 단문을 수집 - 화자는 정해진 시차 간격에 따라 4차례 발성에 참가 - 시차 간격은 1주, 1달, 3달임. 1주 간격의 경우, 2일의 오차 허용. 1달 간격의 경우, 5일의 오차 허용. 3달 간격의 경우, 10일의 오차 허용 - 각 시차별 1명당 1차례 발성량은 2연 숫자 100개, 4연 숫자 250개, 10개의 단답형 대답 및 10개의 단문을 각 5회씩 한번 발성시 총 450개 발성	성별/연령/지역별 구성비율은 단어와 동일
21		유선망			
22		VoIP			
23		마이크(중가)			
24		마이크(저가)			
25		헤드셋			

2) 구축내용

ETRI 음성/언어정보연구센터에서는 2002년부터 2004년까지 3년 동안 음성인식, 음성합성, 화자인식 등 다양한 용도의 음성 DB를 수집할 예정이며, 1차년도인 2002년에는 〈표 1〉과 같이 총 25종의 음성 DB를 구축하였다.

공통 음성 DB는 다양한 통신망(마이크, 헤드셋, VoIP, 유무선 전화망), 지역, 성별, 발성환경(사무실, 지하철, 도로 등)을 고려하여 설계하였으며, 발성대상은 숫자, 단어, 문장이고, 발성방법은 자유발화, 대화체, 낭독체 등 다양한 스타일의 음성 DB로 구성되어 있다.

3) 구축계획

ETRI 음성/언어정보연구센터에서 추진하고 있는 2차년도 DB 구축계획은 음성인식용 단어/숫자/문장 DB, 콜센터 대화체 음성인식용 DB, 언어모델링용 텍스트 DB, 음성합성용 낭독체/대화체 DB를 확장 구축할 예정이며, 업체 및 관련 기관의 의견 수렴 결과를 반영하여 마이크의 종류나 잡음환경 등 수집환경을 좀더 다양화할 예정이다. 또한 이미 배포된 음성 DB에 대해 사용자의 오류 feedback과 자체 오류검증기를 통해 지속적으로 DB오류를 개선하여 고품질의 음성 DB를 배포할 수 있도록 하고 있다.

본 DB는 다양한 통신망 환경, 성별, 연령별, 지역별 화자 분포가 고려된 국내 최대의 한국어 공통음성 DB로써 다양한 영역(숫자, 단어, 문장)의 발성음성으로 구성되었으며, HCI(Human Computer Interface), CTI(Computer Telephony Interface), 텔레매틱스(Telematics), 생체정보인식, 시각장애인용 음성응용, 자동통역 등 각종 음성인식 및 합성엔진 개발에 활용될 수 있다. ETRI 음성/언어정보연구센터에서는 지속적으로 음성기술의 발전방향에 따라 요구되는 DB를 시기적절하게 공급하여 국내업체의 경쟁력을 강화하고자 하며, 향후 각종 음성언어정보의 체계적인 표준화작업을 수행하여 DB의 활용성을 높이는 노력도 병행하여 추진할 예정이다.

2. 원광대 음성정보기술산업지원센터 (SiTEC, www.sitec.or.kr)

원광대 음성정보기술산업지원센터(Speech Information TEchnology and industry promotion Center, 이하 SiTEC)는 전통 산업에의 음성 정보 기술의 응용을 위해, 우리말 음성언어 코퍼스의 중요성을 인식한 산업자원부의 정책적 지원에 의해 2001년 5월에 설립되었으며, 음성 DB의 체계적인 구축과 공개를 위해 노력하고 있다. 여기에서는 SiTEC에서 현재까지 구축 및 배포하고 있는 음성 DB에 대하여 기술한다.

1) 자동차 용용을 위한 코퍼스

가) 자동차 소음 및 음성 DB 프로토타입

최근 자동차 환경에서의 음성인식 응용에 대한 관심과 수요가 많아지고 있다. 자동차 환경에서의 소음 및 음성 DB의 경우 그 수집 절차, 환경 요인 등에 있어서 일반적인 경우와 달리 매우 많은 변수가 있기 때문에 1차년도(2001. 5. 1~2002. 4. 30)에는 이러한 수집 절차 및 환경 요인에 대한 연구와 분석을 위한 프로토타입 음성 DB를 구축하였다. 구축된 자동차 소음 DB는 자동차 요인, 도로 요인 등의 총 270종의 자동차 소음 환경을 정의하고, 각 환경에 대하여 동시에 8개의 채널을 통하여 소음 데이터를 수집하였다. 자동차 음성 DB 프로토타입의 경우 80 km의 주행상황으로 환경을 한정하고 100명의 화자에 대한 음성을 8개의 채널을 통하여 수집하였다.

나) 대규모 자동차 음성 DB

2차년도(2002. 5. 1~2003. 4. 30) 자동차 음성 DB 구축 계획은 300명 화자, 5채널 동시 수집, 1인당 100 토큰 규모였으나, 자동차 환경에서의 음성 인식에 대한 관심과 요구가 증대되면서 업체의 요구가 많아 그 규모를 400명 화자, 8채널 동시 수집, 1인당 200여 토큰 규모로 확대하였다.

2) 수출 지원을 위한 외국어 음성 DB

수출 지원을 외국어 음성 DB로 1차년도에는

중국어 음성 DB를 구축하였고, 2차년도에는 대상 언어를 확대하여 영어, 스페인어 음성 DB를 구축하였다.

가) 중국어 음성 DB

중국어 음성 DB의 발성 목록은 421개의 음절, PBW 단어, 사연숫자, 날짜 관련 단어를 포함하는 2,648개의 단어와 400개의 문장으로 구성되어 있다. 음성 DB 수집을 위해 연변대학 지역협력 사이트를 이용하여 북경어를 사용하는 화자를 중심으로 화자를 모집하였으며, 방언을 사용하는 화자는 가능한 배제하였다. 총 300명의 화자가 발성하였으며 성비는 2 : 3으로 구성되어 있다. 1인당 발성량은 110~123 단어와 20 문장으로 구성되어 있다.

나) 영어 음성 DB

영어 음성 DB의 발성목록은 단독숫자, 예/아니오, 알파벳, 화폐단위, 내장 명령어, 특정날짜 시간표현, 응용어, 요일, 비밀번호, 전화번호, 주 및 도시이름, 신용카드 번호를 포함하는 총 1,586개의 단어로 구성되어 있다. 미국 태생의 영어를 모국어로 사용하는 성인 남녀 총 400명의 화자를 대상으로 하여 미국 현지 녹음 수록하였으며 1인당 140~142 단어를 발성하였다.

다) 스페인어 음성 DB

스페인어의 경우 수출 시장을 감안하여 본토 스페인어를 대상으로 하지 않고, 미국 내에서 거주하는 히스페닉계 사람들이 사용하는 스페인어(히스페닉 스페니쉬)를 대상으로 하여 음성 DB를 구축하였다. 음성 DB 수집을 위해 스페인어를 구사하는 총 300명의 화자(남자 154명, 여자 146명)를 대상으로 미국 서남부 현지에서 녹음을 하였으며, 1인당 120~123 토큰을 발성하였다.

3) 산업 응용을 위한 기반 기술 연구용 DB

가) 다양한 산업 소음 환경에서의 Simulated 음성 DB 제작을 위한 기본 단어 음성 DB

한국어에서 발생할 수 있는 다양한 음운환경 및 음절을 고려한 PRW 4,178 어절을 선정하고 이를 발성목록으로 사용하였다. 센터의 지역협력 사이트를 활용하여 전국적으로 500명 화자의 음성 데이터를 방음실에서 수집하였으며 남녀 성비는 1 : 1이고, 1인당 417~418 단어를 발성하였다. 또한 수집된 음성 데이터 전량에 대하여 음운 레이블링 기준(센터 권고안)에 의해 지역 협력 사이트에서 레이블링 전문 인력에 의해 레이블링 검증 및 수정 작업을 진행 중이다.

나) 다양한 산업 소음 환경에서의 Simulated 음성 DB 제작을 위한 기본 문장 음성 DB

1차년도에 수집된 클린스피치를 단어에서 문장으로 확대 구축하기 위해 총 20,000여 문장을 200명의 화자를 대상으로 방음실 환경에서 수집하였다. 발성 목록의 구성은 21세기 세종계획 형태소분석 균형 말뭉치 1,000 만어절[3]을 분석하여 고빈도 형태소를 포함한 문장으로 구성하였다.

다) 산업 응용을 위한 운율 합성용 문장 음성 DB

남, 녀 전문 성우 각 1인이 4,392 문장을 방음실에서 발성하였다. 마이크는 Rode NT-2를 사용하였으며 EGG 신호도 동시에 수집되었다. 수집된 남, 녀 음성데이터는 모두 음운 및 K-ToBI (Korean-Tone and Break Index) 기준(Ver 3.1)을 적용하여 운율 레이블링을 실시하였고, 현재 레이블링 결과에 대한 검증을 진행하고 있다.

라) 가전 제어용 문장 음성 DB

1차년도에는 KAIST에서 구축된 4,300만 어절의 KAIST Corpus[4]를 사용하여 고빈도 어휘에 대한 분석을 수행하여 고빈도 어휘로 구성된 20,833 문장을 선정하여 음성 DB를 구축하였다. 남, 녀 각 200명의 화자에게 음성 데이터를 수집하였으며 1인당 발성량은 104~105 문장이다. 2차연도에는 다양한 산업 소음 환경에서의 Simulated 음성 DB 제작을 위한 기본 문장 음성 DB 발성 목록과 동일한 총 20,000여 문장을 발성목록으로 사용하여 총 400명의 화자를 추가하여 PC 환경 낭독 문장 음성 DB를 확장하였다.

마) 산업기기 제어용 숫자음 DB의 보완

기존에 공유된 500명분의 숫자음 DB를 보완하기 위하여 구축된 DB이다. 숫자음 DB의 확장을 위한 발성목록은 총 25,000종의 2~3 음절로 이루어진 단위 숫자로 구성되어 있으며, 총 500명의 화자로부터 데이터를 수집하였다.

4) 기타 음성 DB

가) 마이크 시험용 음성 DB

음성 인식 시스템의 성능에 영향을 미치는 다양한 요인 중 마이크의 음향적 특성, 마이크의 위치 및 마이크와 화자와의 거리도 매우 중요한 요인 중 하나이다. 따라서 센터에서는 이러한 다양한 변인에 따른 시험용 음성 DB의 구축을 위해 1차년도에는 마이크의 종류에 특성변화 시험용 음성 DB를 구축하였고, 2차년도에는 마이크로폰의 거리에 따른 영향을 분석하기 위한 음성 DB를 구축하였다.

나) 산업 응용을 위한 기기내장형 음성 DB

최근 다양한 음성정보기술이 실생활에 적용되기에 이르렀고, 차세대 사용자 인터페이스 수단으로 부각되면서 완구, 로봇, 홈오토메이션과 같은 다양한 임베디드용 음성인식 어플리케이션이 개발되고 있다. USB-DSP 임베디드용 음성 수집 툴킷을 이용하여 총 300명의 화자를 대상으로 기기 내장형 electret 콘덴서 마이크를 통해 수집하였으며 1인당 107~108 단어를 발성하였다.

다) 아동용 음성 DB

최근 완구, 교육용 S/W 등 아동용 응용에 대한 요구가 증가함에 따라 아동용 음성인식 응용을 위한 음성 DB를 구축하였다. 총 500명의 초등학교 학생을 대상으로 데이터를 수집하였으며 남녀 성비는 1 : 1이며, 1인당 발성량은 100~101 단어이다. 수집 환경은 사무실 또는 가정집에서 PC의 사운드카드와 Andrea ANC 750 마이크를 이용하여 수집하였다.

5) 3차년도 음성 DB 구축 계획

센터에서는 차기연도 음성 DB 구축을 위한 계획을 작성하기 위하여 음성정보기술 관련 전문가들을 대상으로 수요조사를 실시하였다. 차기연도의 음성 DB의 구축계획은 조사된 결과 및 R&D Roadmap 등을 참고하여 결정하였다. 3차년도에 구축 또는 보완할 음성 DB는 수출 지원을 위한 외국어 음성 DB, 자동차 음성 DB, 자동차 환경 화자 인증 음성 DB, 복지 응용을 위한 음성 DB, 잡음 환경에서의 음성 인식 성능 평가용 DB, 모의 환경 음성 DB 등이 있다.

V. 결 론

지금까지 음성 관련연구에 필수적인 음성 DB의 현황에 대하여 살펴보았다. 음성 DB의 구축에는 많은 시간과 예산 그리고 전문적 지식과 경험이 있는 인력의 참여가 필요하다. 따라서 공동노력으로 함께 제작하여 같이 이용하는 것이 효율적임은 두말할 나위도 없다. 그런 의미에서 이에 대한 깊은 인식을 바탕으로 정부의 적극적인 지원을 받아 영역별로 역할을 분담한 두 기관이 상호 협조하에 대규모의 음성 DB들을 구축하여 보급하게 된 것은 매우 의미있는 일이라 할 것이다. 그럼으로써 관련 업계로서는 다양한 종류의 대규모 음성 DB를 제품개발에 바로 응용할 수 있게 되어 개발 기간의 단축 및 예산 절감이 가능하게 되었다. 아울러 음성관련 학계에서도 그동안의 숙원이었던 음성 DB를 손쉽게 확보 할 수 있게 되어 연구의 활성화에도 기여하고 있다고 생각된다. 음성 DB는 앞으로도 새로운 수준의 음성정보기술의 개발을 위해 한걸음 먼저 확보되어야 할 필수 연구기반이므로 적기에 적량의 음성 DB가 계속 공급될 수 있도록 지속적인 관심과 지원이 필요할 것이다.

참 고 문 헌

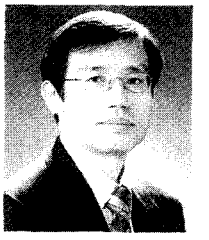
〔1〕 Dafydd Gibbon, Roger Moore, Rchard Winski, Handbook of Standards and Resources for Spoken Language Systems, Mouton de Gruyter 1997

〔2〕 이용주 "음성언어코퍼스" 한국정보과학회지 1998년 2월

〔3〕 문화관광부, 국립국어연구원(2002). 21세기 "세종계획 2001년도 국어 기초자료 구축 분과 연구 결과 보고서"

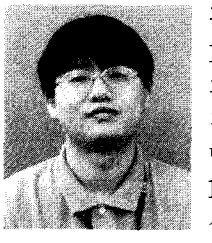
〔4〕 최기선, KAIST 언어자원 2001년도판, 과학기술부 핵심 소프트웨어 과제 결과물 1995-2000 (http://kibs.kaist.ac.kr)

저 자 소 개



이 용 주

1976년 2월 고려대학교 전자공학과 졸업(공학사), 1987년 8월 고려대학교 대학원 전자공학과 졸업(공학석사), 1992년 8월 고려대학교 대학원전자공학과 졸업(공학박사), 1995년 7월~1996년 2월 : 일본 토호쿠대학 응용정보학연구센터 연구생, 1980년 8월~1994년 2월 : 한국전자통신연구소 자동통역연구실 실장, 책임연구원, 1994년 3월~현재 : 원광대학교 전기전자 및 정보공학부 교수, 2001년 5월~현재 : 산자부 지정 음성정보기술산업지원센터 센터장, 〈주관심 분야 : 음성인식, 음성합성, 음성DB, 멀티미디어 응용〉



김 상 훈

1990년 2월 연세대학교 전기공학과 (학사), 1992년 2월 KAIST 전기 및 전자공학과 (석사), 2003년 4월 Univ. of Tokyo (박사), 1992년 3월~현재 : 한국전자통신연구원 음성 DB연구팀 선임연구원, 〈주관심 분야 : 음성합성, 음성인식, 자동통역, 멀티모달〉