한국어 공통 음성 DB 구축 및 오류 검증*

이수종(ETRI), 김상훈(ETRI), 이영직(ETRI)

<차 례>

<Abstract>

Common Speech Database Collection and Validation for Communications

Soo-jong Lee, Sanghun Kim, Youngjik Lee

In this paper, we'd like to briefly introduce Korean common speech database, which project has been started to construct a large scaled speech database since 2002. The project aims at supporting the R&D environment of the speech technology for industries. It encourages domestic speech industries and activates speech technology domestic market. In the first year, the resulting common speech database consists of 25 kinds of databases considering various recording conditions such as telephone, PC, VoIP etc. The speech database will be widely used for speech recognition, speech synthesis, and speaker identification. On the other hand, although the database was originally corrected by manual, still it retains unknown errors and human errors. So, in order to minimize the errors in the database, we tried to find the errors based on the recognition errors and classify several kinds of errors. To be more effective than typical recognition technique, we will develop the automatic error detection method. In the future, we will try to construct new databases reflecting the needs of companies and universities.

* Keywords: speech DB, Validation

* 본 연구는 정보통신부 출연 "음성정보처리기반기술" 과제의 일환으로 수행되었습니다.

1. 서 론

음성에 의한 정보기기의 제어를 가능하게 하려는 노력은 20세기 중반 컴퓨터가 개발된 이래 선진국을 중심으로 꾸준히 시도되어 왔다. 음성은 사람들 간의 의사소통을 위해 익숙한 수단으로서 이제 컴퓨터의 활용이 일반화되고 휴먼 인터페이스 개념과 접목되면서 21세기를 선도할 10대 유망 기술로 부상되어 있다. 특히, 각종 정보기기의 소형화와 이동성 추구에 따라 음성 인터페이스의 도입이 필연적인 요건으로 되어 있어서, 음성 기술의 적용은 피할 수 없는 과제가 되어 있다. 음성 처리 기술을 개발하기 위해서는 음성/텍스트 DB가 기반 요소로 요구된다. 이미 미국, 유럽 등의 선진국에서는 산/학/연 공동으로 연구기관을 설립하여 공통 음성 DB를 구축 보급하고 있다[1]. 미국에서는 ARPA와 DARPA를 통하여 수집한 DB를 기반으로 1990년대 초에 LDC를 구성하였고, 유럽에서는 1995년에 비영리기관으로 ELRA를 만들었다. 국내에서도 최근 100여개의 음성 전문업체가 설립되어 다양한 분야에 음성 정보 기술을 응용하고 있다. 그 동안 대다수의 음성 유관 기관과 음성 처리 업체들은 개별적으로 소규모의 DB를 구축하여 연구와 자체 엔진 개발에 활용하여 왔으나, DB 구축에 많은 시간과 비용이 소요되어 외국업체에 비해 음성 기술 경쟁력이 떨어진다는 지적을 받아왔다. 특히 국내 업체간 중복 DB 구축으로 인해 국가적으로 자원이 비효율적으로 활용되고 있어 음성 정보처리 관련 사업자들의 공동 이익을 도모할 수 있는 공통 음성 DB 구축이 시급하였다. 자동차 산업 등 전통 산업 분야에 특화된 대규모 음성 DB의 경우에는 음성정보기술산업지원센터(SiTEC, 원광대)에서 구축하고 있으나, 수요 기반이 확장일로에 있는 통신망 환경에서의 DB 구축은 음성 처리 업계의 요구 사항을 충분히 반영하지 못하고 있었다[2].

국내 음성 업계의 기술 개발과 애로 기술 지원을 목적으로 2001년도 후반에 한국전자통신연구원에 음성/언어정보연구센터를 설립하였으며, "언어 정보 처리 기술 개발(2003.1-2005.12)" 사업을 통해 다양한 통신망 환경에서 대규모 음성 DB를 구축 배포하고 있다[3, 4]. 2002년도부터 통신망에서의 한국어 공통 음성 DB를 대규모로 구축하는 프로젝트가 본격 시작되었다. 이는 한국어 음성처리를 위한 기반 기술로 활용되는 것을 목표로 하고 있다.

본 고에서는 국내 보급을 위해 수행한 한국어 공통 음성 DB의 구축 내용에 대하여 전반적으로 살펴보고, 음성 DB의 무결성을 목표로 오류 검증을 수행한 예를 소개하고자 한다. 서론에 이어 2장에서는 1차년도에 구축한 DB 구축 내용을 집약한 결과를 보여준다. 3장에서는 PC 환경 하에서 중가 마이크를 통하여 수집한 훈련용 단어 음성 DB에 대한 오류 검출과 오류 유형 분류 결과를 서술한다. 4장에서는 향후의 DB 구축 계획을 소개하고, 5장에서 결론을 맺는다.

2. 1차년도 DB 구축

1차년도에 구축한 음성 DB는 유선, 무선 및 PC 등 환경별로 25개 분야에 걸쳐서 수집되었다. 음성 DB의 구축은 음성인식, 합성, 인증 등 그 용도에 따라 화자의 선택, 시차 적용, 환경 및 규모를 고려하여 수집하게 되며, 마이크 또는 기타 장치를 통하여 입력받고 분류되어 저장 및 가공하여 활용하게 된다. 음성인식률의 제고를 위해서는 서비스 환경에 따라 관련 영역에 적합한 음성 DB가 충분히 확보되어야 하므로 보다 고품질의 음성 DB 확보에 많은 시간과 노력을 기울이고 있다.

1차년도에는 그 동안 국내의 음성학계 및 업계에서 누적되어 온 DB 수요를 긴급히 보급할 목적으로 통신망 환경에서 수집할 수 있는 전반적인 분야의 DB를 다양하게 수집하였다. 통신망 환경은 유/무선 전화망, VoIP, PC 환경으로 대별된다. PC 환경은 다시 세분하여 마이크의 가격을 기준으로 중가 마이크, 저가 마이크, 헤드셋으로 구분하였다. DB의 용도는 음성인식용 단어, 숫자, 문장으로 나눠진다. 문장에는 낭독체 문장, 대화체 문장, 언어 모델링용 문장, 음성합성용 문장, 화자인식용 단어/문장으로 다양하다.

2.1. DB 구성

음성인식용 단어는 10,000 단어를 대상으로 1,000명의 화자가 발성한 결과 파일들로서 모두 100,000 발성 어휘의 분량이며, 100개의 세트로 구분되어 있다. 각 세트는 100 단어를 대상으로 10명의 화자가 1회씩 발성한 어휘들의 묶음이다. 따라서 동일한 단어에 대해 10회씩 발성된 셈이다. 단어의 발성목록은 상장회사명, 지명, 인명, 상호명, 제품명, PC 명령어, PDA 명령어, 그 외에 일반명사로 구성되었다. 음성인식용 숫자는 전화번호, 주민등록번호, 계좌번호 등으로 구성되며, 1-16연속 숫자음으로서 번호 독식(예, 이삼오칠)과 봉독식(예, 둘셋다섯일곱)으로 구분하여 수집하였다. 한국어 숫자음 인식 기술은 실제 음성 서비스 응용 분야가 많아 그 수요가 가장 많은 분야로서, 음성업계에서 인식률 제고를 시급히 요청하고 있는 최대의 애로 기술로 부각되어 있는 실정이다. 구성 화자는 성별, 연령별, 지역에 따라 적정비율로 분포되었다.

한편, 구축된 DB는 80%를 훈련용으로, 20%를 평가용으로 구분하여 패키징하였다. 국내 공통 음성 DB의 취지를 살리고 또한, 다양한 인식률 제시의 표준DB로 활용되기를 기대하고 있다.

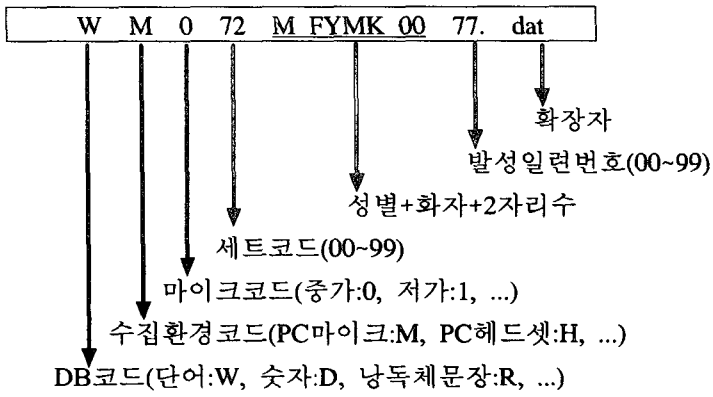
<표 1> 1차년도 한국어 공통 음성 DB 구축 내용 요약[5]

순번	목 적	통신망 환경	화자수 / 발화수	발성내용 / 발성조건	비 고
1	음성인식용 단어	휴대폰	1,000명/10만 발화	-주식상장회사명, 지명, 인명, 상호명, 제품명, PC 명령어, 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 50:50 으로 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 구성비율을 "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/준 낭독체 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	언어모델링 용 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			

2.2. DB 구조

음성 DB에는 음성데이터, 음성데이터의 내용을 텍스트로 표기한 철자 전사, 음성의 시작과 끝을 나타내는 음성 구간 정보 등이 각각 하나의 단위를 이루는 묶음들로 구성된다. 철자 전사의 각 어휘는 음소로 분리되어 어휘사전을 이루고, 음성데이터에 포함된 음향학적 정보로부터 해당 음운 및 언어적 정보와의 모델링 과정을 거쳐 통계적으로 서로 연결됨으로써, 음성과 어휘간에 서로 가장 유사한 데이터를 찾아서 반응하게 된다.

음성데이터는 16kHz 샘플링, 16bit linear PCM으로 저장되어 있으며, 음성구간의 앞, 뒤에 200 msec의 묵음구간이 포함되었다. 음성 DB의 각 파일명은 13개의 digit로 구성되어 코드화되어 있어서 이를 통하여 음성 DB의 종류, 수집 환경, 세트의 구분, 발성 순서 등의 식별이 가능하다. 또한, 한 단위의 음성 DB 묶음은 3종류의 확장자로 나누어져 있으므로 각각 필요한 정보를 확인할 수 있도록 되어 있다.



<그림 1> 음성데이터 파일명 코드

다음의 <표 2>는 한 단위의 음성 DB 묶음에 포함된 3종류의 파일을 보여주고 있다. 구축된 모든 음성 DB들은 모두 아래와 같은 묶음들로 이루어져 있다.

<표 2> 음성 DB 구성 파일 (예)

구 분	파 일 명	내 용
음성 데이터	WM072FYMK0077.dat	음성 파형
철자 전사	WM072FYMK0077.TEXT	계명정보(예)
마킹 정보	WM072FYMK0077.MARK	음성구간정보(시간정보) 0.200063 -1 SBEG 1.302063 -1 SEND

2.3. DB 구축 결과

본 DB는 다양한 통신망 환경, 성별, 연령별, 지역별 화자 분포가 고려된 국내 최대의 한국어 공통 음성 DB로서 다양한 영역(숫자, 단어, 문장)의 발성 음성으로 구성되었으며, HCI (Human Computer Interface), CTI (Computer Telephony Interface), 텔레매틱스(Telematics), 생체 정보 인식, 시각장애인용 음성 응용, 자동 통역 등 각종 음성인식 및 합성엔진 개발에 활용될 것이다. 아래는 1차년도 대량의 DB를 구축하는 과정에서 겪은 경험 및 애로사항을 요약한 것이며, DB 구축시 유의해야 할 사항으로 제시하였다.

(1) 준낭독체 녹음: 녹음 작업 진행 중 가장 화자들이 어려워하는 부분임. 녹음 진행 시간도 상당히 소요됨. 전사 작업의 난이도가 높았으며, 전사 관련 지출 비용 중 상당 부분을 차지한다.
(2) 화자 섭외: 기존 1,000명 외에 새로운 1,000명에 대한 화자 녹음이 어렵다.
(3) 마이크 숫자: 동시에 녹음할 수 있는 마이크 수량은 2-3개 적당하다.
(4) 발화 개수: 1회에 480토큰 발성(80토큰*6회 반복 발성)으로 화자와 모니터 요원이 공통으로 지루해 함. 토큰의 개수를 줄이는 방향으로 검토 요망된다.
(5) 녹음 장비: PC 3대와 믹서기 2대 등 각종 장비가 비대하고, 파일의 갯수가 너무 많아 관리가 어려움. PC 한대로 다채널 동시 녹음하는 툴이 필요하다.
(6) 발성목록: 예를 들어 "당신의 초등학교 이름을 말씀해주세요."와 같은 질문을 텍스트로 보고 답변에 대한 음성을 DB화 할 경우, 녹음에 소요되는 시간이 3배가 걸리게 된다.

한편, 공통 음성 DB 규격이 현실에 맞지 않는 경우가 발견되어 이를 재검토하였으며, 2차년도에 보완하여 적용하였다. 다음은 2차년도 DB 수집을 진행하면서 보완하여 추진한 내용을 항목별로 정리한 것이다.

(1) 화자 및 환경 분포
- 환경별로 모으는 것을 현실적으로 수정하자. 지하철 같은 곳에서는 모을 수 없다.
- 10대는 중/고등학생인데 카드번호, 계좌번호, 주식 등 현실적으로 맞지 않는 것을 수집했기 때문에 어색하다.

(2) DB 수집 방법
- 마이크를 현실화하자. 업체에서 실질적으로 사용하는 저가 마이크(1만원대

이하)를 사용하자.
- PDA와 PC 모니터에 부착한 마이크도 사용하자.
- 유선전화의 경우, 철저한 품질 관리가 필요하다.
- 유선전화는 수집해서 사후 관리하는데 드는 비용이 너무 크다.
- 수집 즉시 검토할 수 있게 하자. 수집하면서 모니터링 할 수 있게 하자.

(3) 발성 목록
- 단어, 낭독체 문장에 철자나 띄어쓰기 오류가 많다.
- 발음 전사는 예외 발음만 괄호를 사용하여 처리하자.
- 숫자는 철저한 발음 전사가 필요하다.
- 여러 개의 음성 파일과 하나의 전사 파일로 구성하는 것은 문제가 많음. 하나의 음성에 하나의 전사 파일로 하자.
- 음성 구간 마킹, 잡음 마킹 등의 정보를 없애자.

2차년도 DB 구축은 1차년도에 비해 화자 섭외가 어려운 점이 있으나 1차년도 DB 구축시 발생한 문제점을 이미 파악하고 있고, 수동/자동 검증툴을 확보하고 있으므로 1차년도 보다 효율적으로 고품질 DB를 수집할 수 있을 것으로 보인다.

2.4. DB 배포

1차년도에 구축된 DB는 전수확인 단계를 거쳐 기술 이전의 형태로 국내 음성학계, 연구기관 및 산업계에 광범위하게 배포하였다. 가능한한 많은 보급이 이루어질 수 있도록 유관기관 및 산업계의 의견을 반영하여 절차를 간소화하고 부담을 최소화하였다. 음성협회 회원/비회원, 대기업/중소기업, 영리/비영리기관 등으로 세세히 분류하여 적정한 규모의 DB를 선택할 수 있도록 하고, 학계의 경우에는 연구에 치중하는 점을 감안하여 20%의 물량으로 활용할 수 있도록 하였다.

한편, 최근 숫자음 인식률 제고에 대한 연구역량을 광범위하게 결집하고 음성산업에 피드백시키기 위하여, 일부 음성학회 학술 행사 기회에 한국어 유선전화 숫자DB (화자: 40명, CD 1장)를 무상으로 배포하였다. 향후에도 숫자음 인식률 향상에 도움이 될 수 있다고 판단되는 경우에는 학술 행사 등을 통하여 계속 배포하고 의견을 수렴할 예정이다.

<표 3> 한국어 유선전화 숫자DB 무상배포 (1차)

	학 계	연구 기관	기 업	계
User별	62	6	3	71
소속별	30	4	2	36

3. 오류 검증

음성데이터와 철자 전사간에 괴리가 있게 되면 음성 처리에 심각한 문제가 발생하는 원인이 된다. 그러므로 음성 DB를 신뢰성 있게 구축하기 위해서는 요구사항이 구체적이고 명확히 제시되어야 하고, 자연성 있는 시나리오에 의해 제작되어야 함은 물론 제작 이후에도 지속적으로 확인, 검증하여 오류를 추출 및 제거하여 그 적합성을 확보해 가야 한다. 음성 DB에 대한 검증은 여러 단계에 걸쳐 이루어지고 있는데, 전체 데이터를 대상으로 하여 규격과의 비교 검증 및 기초 데이터의 충실성 여부를 확인하는 기본 확인 단계, 인식 오류의 유발 가능성이 있는 데이터만을 추출하여 일정 범위 내에서 확인하는 집중 검증 단계 및 활용 과정에서의 지속적인 오류 보고에 의한 재확인 검증 단계로 나누어 볼 수 있다. 여기에서는 집중 검증 단계의 일환으로 1차년도에 구축한 DB중에서 가장 많이 활용될 것으로 보이는 분야를 대상으로 오류 검증을 수행하였으며, 지속적인 오류 검출 방법이 모색되어야 하겠다.

3.1. 오류 검증 대상

분석 대상 음성 DB는 PC 환경에서 중가 마이크를 활용하여 수집한 단어 음성 DB로서, 모두 100세트이나 10세트씩을 단위로 하여 스크립트 프로그램 수행, 인식 실험, 오류 확인 및 결과 집계가 이루어졌다. 인식 실험은 오류 확인 대상 데이터의 범위를 좁혀 오류 가능성이 높은 자료만을 대상으로 확인하기 위하여 수행되었으며, 전체 음성 DB를 대상으로 close-test를 수행하여 오류 가능성이 높은 음성 DB를 분리하여 실제 오류 확인을 실시하였다.

<표 4> 분석 대상 음성 DB 개요

종 류	단어 음성 DB, 중가 마이크, PC 환경
규 모	100,000 발성어휘(10,000 단어, 1,000 화자)
세 트	100 set (1set: 100 단어 x 10명)

3.2. 인식 실험

인식 실험에는 HMM 기반의 가변 어휘 인식 엔진을 활용하였다. 훈련 및 인식 수행 형상으로서, 특징 추출 파라미터는 MFCC 13차와 delta 13차를 합하여 26차로 하였고, 음소 단위별 state 변화, 특징 추출에 따른 HMM 모델 mixture, 학습 횟수, 인식 score 비교에 의한 어휘 인식은 3차까지 출력될 수 있도록 조건을 부여하였다.

3-best 조건 하에 수행된 인식 실험 결과는 음성데이터와 철자 전사와의 일치 여부를 3차까지 비교해 볼 수 있도록 하였으나, 1차에서 인식 성공한 경우 외에는 모두 확인 대상으로 하였다. 인식 실험 결과에는 음성데이터, 철자 전사, 3단계의 인식 결과 외에도 인식 Score 값을 출력해 볼 수 있다.

오류 가능 범위로 분류된 음성 파일에 대하여는 파일명 코드를 활용하여 원본 파일의 저장 경로를 추적한 후 녹음 내용을 직접 청취하여 녹음 상태를 확인하고 철자 전사와 비교하였다. 음성데이터의 청취 결과를 토대로 녹음 상태의 적정성과 발음의 명료성 여부를 확인함과 동시에 철자 전사와 비교하여 그 일치성 여부를 검증하였다.

3.3. 오류 유형 분류

음성 DB의 오류는 음성 오류와 음성표기 오류로 크게 나눌 수 있는데, 좀 더 세분하여 여섯 가지의 유형으로 다음과 같이 세분하였다.

(1) 녹음 오류: 발성 목록과 철자 전사는 되어 있으나 음성데이터가 없는 경우이다.

(2) 발음 오류: 모호하게 발음함으로써 철자 전사와 일치한다고 볼 수 없는 경우이다. 한꺼번에 발음하지 않고 더듬거린 경우에도 발음 오류에 포함시켰다.

(3) 철자 전사 오류: 발성음과 철자 내용이 다른 경우이다. 음성 처리 과정에서는 철자 전사 결과가 실제로 활용된다. 따라서 녹음 품질을 우선 확인한 다음에 철자 전사 결과와 비교하였다.

(4) 띄어쓰기 오류: 스크립트 작성 및 인식 테스트 과정에서 도구에 의해 자동으로 붙이도록 하기 때문에 경미한 오류로 분류될 수 있다.

(5) 맞춤법 오류: "청계휴게실", "물유본부"로 표기한 경우가 그 예이다.

(6) 파일명 오류: 여성의 발음을 남성의 파일명 코드로 또는 그 반대의 경우이다.

다음의 <표 5>는 오류 확인 대상 범위와 오류 유형별로 집계한 결과를 보여준다. 분석 대상 음성 DB의 수는 모두 99,867이며, 이들 중에서 인식 실험을 통하여 6,269종의 음성 DB를 추출하여 오류 확인을 수행하였다.

<표 5> 오류 유형 분류

구 분	오류 유형						계
	녹음오류	발음오류	철자오류	띄어쓰기	맞춤법오류	파일명오류	
오류갯수	2	17	77	50	7	77	230
%	0.87%	7.39%	33.48%	21.74%	3.04%	33.48%	100%

3.4. DB 보완

오류 확인 결과에 따른 음성 DB의 보완은 오류 유형에 따라 달리 처리되는데, 녹음 오류나 발음 오류로 최종 판정되는 경우에는 해당 음성 DB를 삭제하고, 음성데이터의 표기에 관련된 오류는 관련된 자료를 정정하게 된다. 음성 표기에 대한 오류는 사후 보완이 가능하나, 일단 구축된 음성데이터의 오류는 삭제 외에는 보완 방법이 없는 것으로 판단된다. 오류 확인 결과를 통하여 실제 원본 DB에 대한 최종 보완이 이루어지기까지는 관련 인원들에 의해 수차에 걸친 재확인 과정을 거쳐서 이루어졌다.

한편, 오류 가능 범위를 추출함에 있어서, HMM 기반의 인식엔진의 오인식을 감안하여 광범위하게 확장하게 되면서, 실제 오류 여부에 대한 확인 대상이 많아지게 되고 인적 오류의 소지도 그만큼 많아질 수 있었다. 따라서 지속적인 DB의 구축이 추진되고 검증 대상 DB가 점점 많아짐에 따라, 검증 전용 시스템의 구현을 추진하고 있다. 이를 통하여, 오류 확인 대상 데이터를 효과적으로 추출하고 자동화를 확대함으로써, DB의 신뢰성이 더욱 강화될 것으로 기대하고 있다.

4. 향후 계획

음성 적용 영역이 확대됨에 따라, 음성 정보 연구 기관 및 음성 산업계를 중심으로 다양한 음성 DB에 대한 요구가 계속되고 있다. ETRI에서는 2차년도에도 음성 업계의 요구를 수렴하여 DB 구축을 진행하고 있으며, 필요한 음성 DB 영역에 대한 DB 구축을 계속할 예정이다. 현재까지 수렴한 요구 사항은 다음과 같다.

(1) 원격 음성 DB 구축 필요
(2) 잡음 DB 필요
(3) 외국어 합성 DB 필요
(4) 외국어 인식용 단어 DB 필요
(5) 숫자음 seed 모델용 수동 음소 분할 필요
(6) 휴대폰/PDA/텔레메틱스 등 모발 환경 DB 구축 필요
(7) 공통 DB 보다 응용 서비스에 적합한 DB 수집 필요
(8) EU의 AURORA 표준 trace 필요
(9) 콘덴서 마이크 사용 필요
(10) 화자인식 DB의 경우, 시차별 DB는 매우 적절
(11) 화자인식 DB의 경우, 5-10음절 키워드가 적당
(12) 화자인식 DB의 경우 DB, 수집은 어렵고 DB 수요는 작음
(13) 합성 DB의 경우, 운율 레이블링 필요

(14) 한국인의 영어발음 DB 필요
(15) 어린이용 대화체 운율/어투 모델링용 합성 DB 필요
(16) 마이크 환경, 일상 대화 DB 필요

수렴 결과 중 일부는 2차년도 DB 구축에 반영하였다. 2차년도 DB 구축 계획은 다음 <표 6>와 같다.

<표 6> 2차년도 DB 구축 계획

순번	목 적	통신망 환경	화자수 / 발화수	발성 내용 / 발성 조건	비고
1	음성인식용 단어	휴대폰	1,000명/ 10만 발화	주식상장회사명, 지명, 인명, 제품명, PC명령어, PDA 명령어, 일반명사로 구성. 성별, 연령별, 지역별 분포 고려	
2		유선망			
3		VoIP			
4		마이크(증가)			
5		PDA			
6		헤드셋			
7	음성인식용 숫자	휴대폰	1,000명/ 10만 발화	1-16연숫자. 번호독식/봉독식 발성, 계좌번호, 단위, 전화번호, 주민등록번호로 구성. 성별, 연령별, 지역별 분포 고려	
8		유선망			
9		마이크(증가)			
10		PDA			
11		헤드셋			
12	음성인식용 낭독체문장	VoIP	1,000명/ 10만 발화	각 화자가 100 문장씩 발성한 낭독체 방송뉴스 문장	
13		마이크(증가)			
14		PDA			
15		헤드셋			
16	음성인식용 대화체 문장	유선/휴대폰 전화망	500명/ 5,000대화	가상 Call center에서 고객과 상담원 대화 녹취(시나리오 사용)	
17	언어모델링용 문장	텍스트	2,000만 어절 수동/ 4,000만 어절 자동	일간지 신문 3,000만 어절 수동 철자/띄어쓰기 수정. 9,000만 어절 자동 철자/띄어쓰기 수정	
18	음성합성용 문장	정보전달용 낭독체	남녀 1인/1만 문장	남녀 성우 20여명 후보에서 선호도 평가 후 2명 선정	
19		대화체 문장	남녀 2인/1만 문장	2인이 서로 대화하는 음성 녹음	

2차년도 DB 구축 일정은 다음과 같다.

(1) 2003년 2월: 공통 음성 DB 규격 보완 완료
(2) 2003년 3월-7월: DB 구축
(3) 2003년 8월-9월: DB 검증
(4) 2003년 10월: DB 배포

이와 같은 DB 구축과 고려되어야 할 분야로서는 DB 표준화가 있다. 즉, DB를 구축하고 활용하면서 애로를 겪는 분야 중 하나로서는 표준화가 미흡하다는 것이다. 이를 해소하기 위해 DB의 수집 환경, 파일 구성, 파일 구조, 레이블링 기호 등에 이르기까지 가능한 분야부터 표준화를 추진할 예정으로 유관기관과 협의를 진행하고 있다.

5. 결 론

이번에 구축한 대규모 한국어 공통 음성 DB는 음성인식/음성합성 엔진을 개발하기 위한 기본 자료로 매우 유용할 뿐만 아니라 한국어 공통 음성 DB 구축을 통해 국내 산업체간 음성/텍스트 DB 구축의 중복 투자를 방지하고 국내 음성 정보 산업 경쟁력 강화 및 음성서비스 시장을 활성화하는데 크게 기여할 것으로 판단된다. 본 고에서는 1차년도에 구축한 한국어 공통 음성 DB와 일부 단어 음성 DB을 대상으로 수행한 오류 검증 및 오류의 유형을 살펴보았다.

ETRI 음성/언어정보연구센터에서는 지속적으로 음성 기술의 발전 방향에 따라 요구되는 DB를 시의 적절하게 공급하여 국내업체의 경쟁력을 강화하고자 하며, 향후 각종 음성 언어 정보의 체계적인 표준화 작업을 수행하여 DB의 활용성을 높이는데 최선을 다할 것이다.

참 고 문 헌

[1] ETRI 음성/언어정보연구센터: http://voice.etri.re.kr

[2] 김봉완, 이용주, "음성정보기술산업지원센터의 음성 코퍼스 구축 현황 및 계획", *한국음향학회 하계학술대회논문집*, 21권, 1(s)호, pp.49-52, 2002.

[3] 김상훈, 오승신 et al., "공통 음성 DB 구축", *한국음향학회 하계학술대회논문집*, 21권, 1(s)호, pp.21-24, 2002.

[4] 오승신, "공통 음성 DB 구축을 위한 발성목록 설계", *한국음향학회 하계학술대회논문집*, 21권, 1(s)호, pp.29-32, 2002.

[5] 김상훈, 박문환, 김현숙, "통신망 환경 한국어 공통 음성 DB 구축", *대한음성학회 춘계학술발표대회*, pp.23-26, 2003.

[6] L. Rabiner, B-H. Juang, *Fundamentals of Speech Recognition*, New Jersey: Prentice Hall PTR, 1993.

[7] X. Huang, A. Acero, H-W. Hon, *Spoken Language Processing*, New Jersey: Prentice Hall PTR, 2001.

[8] 이건상, 양성일, 권영헌, *음성인식*, 한양대학교출판부, 2001.

[9] 김상훈, 오승신 et al., "공통 음성 DB 구축", *한국음향학회 하계학술대회논문집*, 21권, 1(s)호, pp.21-24, 2002.

접수일자: 2003년 05월 03일
게재결정: 2003년 06월 12일

▶ 이수종(Soo-jong Lee)
주소: 305-761 대전광역시 유성구 전민동 464-1 엑스포아파트 206-1606
소속: 한국전자통신연구원 컴퓨터소프트웨어연구소 음성/언어정보연구센터
음성 DB연구팀
전화: (042) 860-5584
FAX: (042) 860-6436
E-mail: sjleetri@etri.re.kr

▶ 김상훈(Sanghun Kim)
주소: 305-761 대전광역시 유성구 전민동 464-1 엑스포아파트 2405-907
소속: 한국전자통신연구원 컴퓨터소프트웨어연구소 음성/언어정보연구센터
음성 DB연구팀
전화: (042) 860-5141
FAX: (042) 860-6436
E-mail: ksh@etri.re.kr

▶ 이영직(Youngjik Lee)
주소: 305-755 대전광역시 유성구 어은동 99 한빛아파트 111동 601호
소속: 한국전자통신연구원 컴퓨터소프트웨어연구소 음성/언어정보연구센터
전화: (042) 860-6144
FAX: (042) 860-6436
E-mail: ylee@etri.re.kr