韓國 컴퓨터情報學會 論文誌
第11卷 第1號, 2006. 3.

2006-11-1-4-4

다양한 음성코퍼스의 통합 관리시스템 구축

유 경 택*, 정 창 원**, 김 도 관***, 이 용 주****

Construction of Integration Management System of Various Speech Corpora

Kyeong-Taek Rhyu *, Chang-Won Jeong **, Do-Goan Kim ***, Young-Ju Lee ****

요 약

본 논문에서는 다양한 음성코퍼스의 통합 관리 시스템을 설계하고 구현하기 위한 여러 고려 사항들을 검토 하고자 한다. 본 논문의 목적은 음성 연구에 필요한 다양한 음성 데이터베이스의 종류 그리고 서로 다른 데이터 형태로 구축된 음성코퍼스를 통합적으로 관리하는데 있다. 또한, 부가적으로 사용자가 요청하는 다양한 조건에 맞는 음성 데이터들을 효과적으로 검색 가능하고 새로 구성된 음성코퍼스를 손쉽게 추가 할 수 있도록 고려하였다. 이를 위해 기존의 구축된 음성코퍼스의 수정 없이 새로운 정보를 통합 관리하기 위한 전역 스키마(global schema)를 설계하고, 이를 기반으로 시·공간의 제약 없이 액세스 할 수 있는 웹 기반의 통합 관리 시스템을 구축하였다. 끝으로 서비스에 포함된 수행 결과인 웹기반 인터페이스를 기술하고, 통합 관리 시스템을 구현하기위해 인덱스 뷰를 사용한 효과성을 보인다.

Abstract

In this paper, we propose relevant to design and implementation of an integrated management system for various speech corpora. The purpose of this paper is to manage an integrated management system for various kinds of speech corpora necessary for speech research and speech corpora constructed in different data formats. In addition, ways are considered to allow users to search with effect for speech corpora that meet various conditions which they want, and to allow them to add with ease corpora that are constructed newly. In order to achieve this goal, we design a global schema for an integrated management of new additional information without changing old speech corpora, and construct a web-based integrated management system based on the scheme that can be accessed without any temporal and spatial restrictions. Finally, we describe the web based interface which are the executed results involved in the service and show the efficiency of using the index view for implementation of integrated management system.

▸ Keyword : 음성코퍼스(speech corpora), 전역 스키마(global schema), 웹기반 관리 시스템 (web-based management system)

• 제1저자 : 유경택
• 접수일 : 2005.12.23, 심사완료일 : 2006.02.06
* 극동정보대학 컴퓨터정보과 교수,** 전북대학교 차세대 LBS응용 연구센터 연구 교수
*** 원광대학교 경상대학 정보전자상거래학부 IT 초빙교수, **** 원광대학교 전기전자 및 정보공학부 교수

Ⅰ. 서 론

본 논문에서는 기존 음성 DB나 신생 음성 DB(이하 음성코퍼스로 표기)의 다양성과 이질성을 통합하여 정보의 일관성 있는 수용 및 확장이 가능한 통합 관리 시스템에 대해 기술한다. 이에 관련된 연구는 제한된 어휘의 인식 및 합성 이러한 기술 중 시스템은 실용화에 접어들고 있고 무제한 어휘를 대상으로 하는 시스템들의 연구 개발이 진행되고 있다[1,2]. 또한 무제한 어휘의 인식시스템 개발에는 음소(phoneme) 단위 인식기술 개발이 필수적이지만, 연속 음성 중의 음소는 발성하는 사람의 개인차뿐만 아니라 전후에 발성되는 음소의 영향을 받은 조음 결합에 따라 그 특성이 크게 변화한다. 이러한 개인차 및 조음결합의 현상을 분석하고 시스템의 객관적인 평가를 위해서는 많은 사람이 발성한 다양한 음성 데이터가 필요하다. 음성 데이터는 다종다양(성별, 연령, 방언, 발성자 수 등)한 정보가 필요 하지만, 현재까지 구축된 데이터는 각 연구자가 필요에 따라 적당한 규모의 음성데이터를 만들어 보관하고 이용해 왔다. 그러나 연구의 진보에 따라 처리 속도 및 어휘의 수가 개선되었으므로 측정을 위한 데이터의 양도 지속적으로 증가되고 있다.

최근에 사용되는 음성 인식 모델의 경우 은닉 마르코프 모델 (Hidden Markov Model : HMM) 이나 Bigram 또는 Trigram으로 대표되는 통계적 기법을 사용한다[3]. 이러한 모델은 인식 시스템 학습을 위해 대량의 음성 데이터를 필요로 한다.

이에 음성 연구가 발전됨에 따라 기존의 음성코퍼스가 대량의 데이터 형태와 다양한 포맷으로 구축되고, 발성자 수와 음성 데이터의 종류 등이 증가되면서 규모가 확장 되고 있다. 이러한 추세에도 불구하고 현재 기존 음성 DB나 신생 음성 DB(이하 음성코퍼스로 표기)는 체계적으로 표준화된 디렉터리 구조 및 파일 명칭 없이 음성코퍼스의 후처리 과정을 거치면서 파일 단위로 각각 저장된다.

따라서 본 논문에서는 서로 다르게 구축된 음성코퍼스를 인터넷을 이용한 단일화된 사용자 인터페이스를 통해 다양하고 이질적인 종류의 음성 데이터, 파일 등을 효율적으로 취합 및 관리가 용이한 통합 관리 시스템을 구현하여 일관성을 유지시키고자 한다.

이러한 통합 관리 시스템은 음성코퍼스 정보를 분야별로 세분화하여 시간과 장소에 구애받지 않게 체계적으로 열람 및 검색이 가능한데, 이는 음성코퍼스 정보를 체계적으로 현장에서 응용 및 적용한다는 측면에서 중요하다.

논문의 구성은 다음과 같다. 2장에서 음성코퍼스 구축과정과 보관 방법에 대한 내용과 문제점을 기술하며, 3장에서는 통합 관리 시스템의 구성과 구조에 대해 설명한다. 그리고 4장에서는 구축된 통합관리 시스템의 구현 환경과 실행 화면을 제시어 및 질의어를 통하여 검토한다. 마지막 5장에서는 결론 및 향후 연구 내용에 대해 기술한다.

Ⅱ. 음성코퍼스 구축 과정

음성코퍼스는 음성인식, 합성 등 음성신호 처리에 있어서 매우 중요한 요소이다. 그러나 음성코퍼스의 경우 일반적으로 구축된 음성 데이터의 디렉터리 구조 및 파일 명칭 등은 거의 없는 실정이다. 최근 들어서 이에 대한 표준화를 위한 연구가 진행되고 있다[4].

사용자가 음성코퍼스의 다양한 정보를 쉽게 이용하기 위해서는 이질성을 가지는 음성코퍼스의 정보를 하나의 파일 구조처럼 볼 수 있도록 하는 정보의 통합기능과 분산된 정보에 대한 통합 질의 기능이 요구된다. 기존에 제작된 음성코퍼스나 새롭게 제작되는 음성코퍼스는 공통의 정보 영역을 가지고 있다. 하지만 공통 정보 영역은 응용 환경에 따라 이질적인 정보로 나누어진다. 예를 들어 전화망을 통한 음성인식 코퍼스와 PC 환경에서의 음성인식을 위한 음성코퍼스의 공통 정보 영역은 각각 응용 환경에 맞는 규격과 정보를 갖고 있으며 (그림 1)과 같이 서로 다른 정보 영역을 가진다.

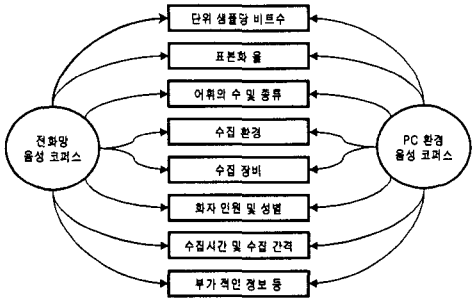

그림 1. 수집 환경에 의해 이질성을 가지는 공통 정보영역
Fig 1. The Common information area with heterogeneity by collection environments

또한 공통 정보 영역은 서로 다른 음성코퍼스의 응용에 따라 다양한 값을 가지면서 정보 수집이 된다. 위와 마찬가지로 해당 음성코퍼스 용도에 맞는 다양한 발성목록을 추출하여 사용한다. 발성목록은 모든 화자가 동일한 방법으로 발성하지 않고, 각 음성코퍼스 응용에 맞는 발성 조건으로 음성을 녹취한다. 발성목록 따른 발성 조건은 다음〈표 1〉과 같다.

표 1. 발성목록에 따른 발성 조건
Table 1. speaking condition by speaking lists

구 분		내 용
길이에 따른 분류	단어음성 (isolated words)	단어음성은 주로 음절, 단어, 어절 등을 대상
	연속음성 (continuous speech)	문장을 대상 - 자연스럽게 발성
상황에 따른 분류	낭독음성 (read speech)	주어진 발성목록을 그대로 발성
	대화음성 (dialog speech)	상황에 따라 주어진 목적을 달성하기 위해 주고받는 형태
자유도에 따른 분류	정형음성 (prompted speech)	정형음성은 발성할 내용을 미리 준비하여 발성 - 낭독체 및 대화체 포함
	자유발화음성 (spontaneous speech)	자유발화음성은 준비 없이 즉시 발성
적용 용도에 따른 분류	인식용 (for recognition)	발성하고자 하는 내용이 한 사람이 발성하기에는 너무 많아 부분적으로 나누어 자연스럽게 발성 - 발성자의 수를 증가
	합성용 (for synthesis)	합성용의 경우는 동일한 발성자의 음성 데이터가 필요

위와 같이 음성 데이터 수집은 발성목록에 따른 발성 조건을 갖는다. 또한 말뭉치에서 발성목록을 다양하게 설계할 수 있다. 일반적으로 이에 따라 음성코퍼스의 발성목록은 다음 〈표 2〉와 같이 다양하게 구축된다.

표 2. 발성목록별 응용 분야
Table 2. Application Area of speaking lists

용 도	발성목록 내용
음성인식용 단어	지명, 인명, 제품명 등
음성인식용 숫자	2~10연 숫자, 번호독식, 단위(십, 백, 천, 만 등) 등
음성인식용 문장	신문기사, 방송뉴스 문장 등
음성 합성용	발성 가능한 모든 음운환경을 포함한 문장 등
화자 인식용	2연 및 4연 숫자 등

발성목록 설계 부분이 이루어지면 화자를 선별하게 된다. 일반적으로 화자 선별 기준은 성별, 나이, 지역 등으로 이루어진다. 선별된 화자에 의해 음성 수집이 이루어지고 해당 정보에 따른 명칭을 붙여 디렉터리 및 파일 명칭을 정한다.

화자정보는 음성코퍼스에 있어 매우 중요하므로 대체적으로 〈표 3〉과 같은 정보를 수집하여 디렉터리 구조 및 파일 명칭 등의 수집된 정보를 기록한다.

표 3. 음성코퍼스 화자정보
Table 3. speaker information of speech corpora

NO	항목	NO	항목	NO	항목
1	이니셜	5	12세 이전 성장지	9	모출생지
2	성별	6	현 거주지	10	방언정도
3	나이	7	직업	11	발성목록
4	출생지	8	부 출생지	12	기타

수집 환경에 따라서도 다양한 음성코퍼스 형태로 구축된다. 수집 환경별 영역은 일반적으로 다음 〈표 4〉와 같다.

표 4. 음성코퍼스 수집 환경
Table 4. collection environments of speech corpora

NO	구 분	내 용
1	가 시 성	노출 or 비밀
2	환 경	녹음실 or 실내(외)
3	제 어	임의녹음 or 낭독음성
4	모니터링	온라인 or 오프라인
5	데 이 터	단일채널 or 다채널
6	장 비	마이크 or 전화

위와 같이 크게 음성코퍼스는 발성목록, 화자정보, 환경별 정보를 기준으로 하여 후처리 과정에서 표준화 되지 않은 디렉터리 구조 및 파일 명칭을 갖는다.

예를 들면 음성 데이터가 저장되는 디렉터리 구조는 일반적으로 다음과 같은 순으로 이루어진다.

2.1 디렉터리 구조

디렉터리 구조는 /〈corpus〉/〈sex〉/〈set#〉/〈speaker_id〉 와 같은 순으로 이루어지고 이에 대한 설명은 다음 〈표 5〉와 같다.

표 5. 음성코퍼스 디렉터리 구조
Table 5. directory structure of speech corpora

디렉터리 명칭	내 용	비 고
corpus	음성코퍼스 명칭	pbw/ TelNUM 등
sex	성별 (남 : m/ 여 : f)	남(male)/ 여(female)
set#	발성목록(전체 또는 sub 발성목록)	setpbw/ set001~set100
speaker_id	화자정보 예) f ss 3 hkh 00 s001 f: 성별(f : 여/ m : 남) ss : 지역 코드 정보 3 : 연령대 코드 정보 hkh : 화자 이니셜 00 : 화자 이니셜 식별번호 s001 : 발성목록	지역 코드 정보 연령대 코드 정보 〈표 6〉

화자정보는 대체적으로 발성목록, 성별, 지역, 나이, 이니셜 등으로 기록되어 진다. 특히 지역을 구분하기 위해 다양한 정보를 기록하고 있다. 다음 〈표 6〉은 일반적으로 사용하는 지역 코드 정보와 연령대 코드 정보이다.

표 6. 지역 및 연령대 코드 정보
Table 6. code information of region and age

NO	지 역	코드 정보
1	서울/경기	ss
2	충청북도	cb
3	충청남도	cn
4	전라북도	jb
5	전라남도	jn
6	경상북도	kb
7	경상남도	kn
8	강원도	kw
9	제주도	jj

NO	나 이	코드 정보
1	10대	1
2	20대	2
3	30대	3
4	40대	4
5	50대	5
6	60대	6

2.2 파일 명칭

파일 명칭은 보편적으로 〈file_id〉.〈file_type〉 과 같은 순으로 이루어지고 파일 명칭은 다음 〈표 7〉과 같다.

표 7. 음성코퍼스 파일 구조
Table 7. File structure of speech corpora

파일 명칭	내 용	비 고
file_id	전체발성목록 및 부분발성목록 번호	prw001~3813 : 전체 발성목록 set001 001 - set001 : 발성목록 set - 001 : 발성목록 set내에서 번호
file_type	파일종류	.wav : 음성 waveform 파일 .txt : 발성목록 set 내에서 번호 .trn : 발성단어의 orthographic transcription

보편적으로 위와 같은 디렉터리 구조 및 파일 명칭으로 음성코퍼스가 구축이 된다.

위와 같은 디렉터리 및 파일 정보를 바탕으로 전체 음성코퍼스의 정보 구조는 다음(그림 2)과 같다.

```
〈CORPUS〉/〈TYPE〉/〈SPEAKER_ID〉/〈FILE_ID〉.〈FILE_TYPE〉
        〈CORPUS〉 ::= 음성코퍼스 명칭
        〈TYPE〉 ::= 발성목록
         // 단독숫자, 4연 숫자, 단문, 장문 등
        〈SPEAKER_ID〉 ::= 〈SEX〉〈INITIAL〉〈DIGIT〉
          〈SEX〉 ::= M | F   // 남(MALE)/ 여(FEMALE)
            〈INITIAL〉 ::= 화자 이니셜, 영문 3 문자
            〈DIGIT〉 ::= 0~9, 동일 이니셜 구분
        FILE_ID ::= 〈FILE_TYPE〉
        〈FILE_TYPE〉 ::= 발성목록
       // 단독숫자, 4연 숫자, 단문, 장문 등
        FILE_TYPE ::= .WAV      // 음성 파일
```

그림 2. 음성코퍼스 정보 구조
Fig 2. information structure of speech corpora

음성코퍼스의 일반적인 전체 수집 과정(그림 3)은 다음과 같다고 할 수 있고, 후처리 과정에서 디렉터리 구조 및 파일 명칭에서 기존 음성코퍼스나 신생 음성코퍼스는 다양하고 이질적인 정보를 가지는 구조로 만들어지고, CD-ROM 및 디스크 등 저장 매체에 저장한다.

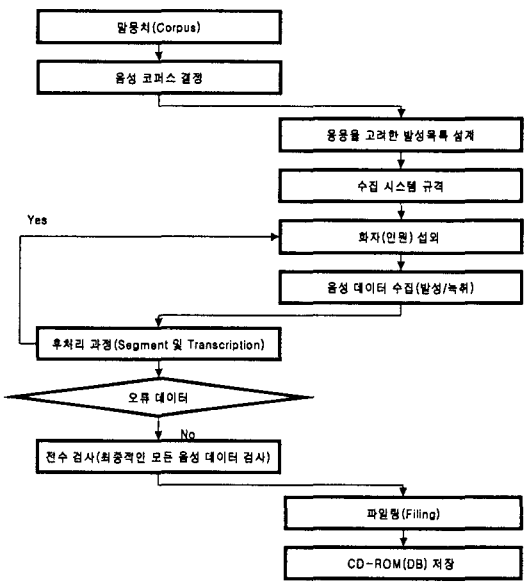

그림 3. 음성코퍼스 구축 과정
Fig 3. construction procedure of speech corpora

지금까지 음성코퍼스를 구축 하고 있는 국내의 경우를 보면 음성인식 및 합성연구에 필요한 음성코퍼스를 각 연구자가 각각 부분적으로 제작하여 자체 개발용으로 사용하고 있으며, 한국전자통신연구원(ETRI, Elect ronics Telecommunications Research Institu te), 음성정보산업기술지원센터(SiTEC, Speech Infor -mation Technology & Industry Promotion Center), 한국통신, 국어공학센터, 한국과학기술원(KAIST, Korea Advanced Institute of Science and Technology), LG 전자, 삼성종합기술원 등에서 개별적인 음성코퍼스를 보유하고 있다[5, 6, 7, 8, 9, 10].

국외를 살펴보면 LDC(Linguistic Data Consortium)에서는 음성 및 텍스트에 관한 코퍼스, 사전, 그 밖의 공유 가능한 자원에 대해서 데이터의 수집, 작성, 배포를 담당하고 있으며, DARPA(Defence Advanced Research Project Agency)의 중심으로 TI(Texas Instruments), MIT(Massachusetts Institute of Technology), CMU (Carnegie Mellon University), NIST(National Institute of Standards and Technology), AT&T(American Telephone and Telegraph Company)의 Bell 연구소, IBM(International Business Machines Corporation) 등이 음성코퍼스를 구축하고, NIST에서 이를 CD-ROM으로 제작하여 국립기술정보서비스를 통해 관련 연구기관에 배포하고 있다[11, 12, 13].

유럽의 경우 ELRA (European Language Resource Association)에서는 음성 및 텍스트에 관한 코퍼스 자원들을 모두 다루고 있다. 일본의 경우 음성코퍼스에 관한 연구는 주로 전자기술종합연구소(ETL, Electro technical Laboratory), 동북대학, 일본 전자공업 진흥협회(JEIDA, Japan Electronic Industry Development Association), ATR(Advanced Telecommunications Research Institute)의 자동번역전화 연구소, 오오사카 대학 등에서 수행하고 있다. 국내 및 여러 기관에서 음성코퍼스를 제작하여 CD-ROM 등에 수록하여 보관하고 있다.

표 8. 국내에 기 개발된 음성코퍼스
Table 8. existing developed speech corpora in internal

구분	음성코퍼스 명	수집 장비	저장 매체
ETRI	휴대폰 음성인식용 단어DB	휴대폰	CD-ROM
	유선전화 음성인식 단어 DB	유선전화	
	PC 음성인식용 단어 DB	VoIP	
		PC마이크/저가	
		PC마이크/중가	
		PC 헤드셋	
	휴대폰 음성인식용 숫자 DB	휴대폰	
	유선전화 음성인식용 숫자 DB	유선전화	
	PC 음성인식용 숫자 DB	VoIP	
		PC마이크/저가	
		PC마이크/중가	
		PC 헤드셋	
	PC 낭독체 음성인식용 문장 DB	VoIP	
		PC마이크/저가	
		PC마이크/중가	
		PC 헤드셋	
	전화망 대화체 음성인식용 문장 DB	휴대폰, 유선전화	
	정보전달용 낭독체 음성합성문장 DB	고성능 마이크	
	휴대폰 화자인식용 DB	휴대폰	
	유선전화 화자인식용 DB	유선전화	
	PC 화자인식용 DB	VoIP	
		PC마이크/저가	
		PC마이크/중가	
		PC 헤드셋	

구분	음성코퍼스 명	수집장비	저장 매체
SITEC	자동차 환경 단어 음성 DB	핸즈프리 마이크 콘덴서 마이크 헤드셋 마이크	CD-ROM
	자동차 노이즈 DB	핸즈프리 마이크 콘덴서 마이크 헤드셋 마이크	
	PC 환경 중국어 음성 DB	PC 마이크	
	PC 환경 낭독 문장 음성 DB	PC 마이크	
	PC 환경 아동용 음성 DB	PC 마이크	
	마이크 성능 시험용 음성 DB		
	클린 스피치 PRW 음성 DB	콘덴서 마이크	
	클린 스피치 낭독 문장 음성 DB	콘덴서 마이크	
	자동차 환경 단어 음성 DB	핸즈프리 마이크 콘덴서 마이크 헤드셋 마이크	
	영어 음성 DB	PC 마이크	
	스페인어 음성 DB	PC 마이크	
	운율 합성용 낭독 문장 음성 DB	콘덴서 마이크	
	기기내용형 음성 인식용 DB	콘덴서 마이크	
한국통신	부서 안내용 음성 DB	유선전화	CD-ROM
	증권 안내용 음성 DB	유선전화	
	음성 다이얼링 서비스 시스템용 음성 DB	유선전화	
	호텔 예약용 음성 DB	유선전화	

국내에서는 ETRI와 SiTEC이 비교적 다양한 음성코퍼스를 구축하고 있다. 그밖에 국어공학센터, KAIST, LG전자, 삼성종합기술원 등에서 다양한 음성코퍼스를 가지고 있다. 국외의 경우는 다음〈표 9〉와 같다.

음성 연구가 진보되어감에 따라 다양한 포맷과 대량의 음성코퍼스가 제작되어 사용되고 있지만 대부분의 경우 공통적인 표준화 없이 구축되고 있다. 일반적인 보관 방법은 보조기억 장치인 CD-ROM과 디스크 등에 파일 단위로 저장하여 관리하고 있다. 이러한 보관 방법으로 인하여 특정 음성코퍼스, 음성 파일, 지역별 특정 화자 등의 정보를 찾으려면 각각의 저장 매체에 대한 빈번한 조사 과정에 따르는 시간적, 공간적, 비용적인 문제가 발생한다[14].

표 9. 국외에 기 개발된 음성코퍼스
Table 9. existing developed speech corpora in external

구분	음성코퍼스 명	주관기관	비 고	저장 매체
미국	TI Digit	TI	음성 시스템 설계 및 평가용	CD-ROM
	TIMIT	DARPA	음성 시스템 설계 및 평가용	
	NTIMIT	NYNEX	음성 시스템 설계 및 평가용	
	DARPA 자원관리DB	DARPA	음성 시스템 설계 및 평가용	
	ATISO	DARPA	자유발화음성, 낭독음성	
	Switchboard (Credit card)	DARPA	자유발화음성	
	TI - 46	TI	음성 시스템 설계 및 평가용	
	RoadRally 대화형 음성 corpora	DOD	자유발화음성, 낭독음성	
	Switch board(Complete)	DARPA	자유발화음성	
	ATC	DARPA	자유발화음성	
	MapTask	HCRC	자유발화음성	
	MARSEC	ESRC	자유발화음성	
	ATIS2	DARPA	자유발화음성	
	WSJ-CSR	DARPA	음성 시스템 설계 및 평가용	
유럽	EUROM. 0 (ESPRIT Project 1541, SAM)	ESPRIT	음성 입력 평가용	CD-ROM
일본	일본문부성연속음성 DB	문부성	음성 시스템 설계 및 평가용	CD-ROM
	일본문부성방언음성 DB	문부성	음성 시스템 설계 및 평가용	
	동북대 음성 DB	문부성	음성 시스템 설계 및 평가용	

이러한 문제점을 해결 하고자 본 논문에서는 음성코퍼스의 다양하고 이질적인 정보를 통합 스키마로 연결하여 관리자 및 사용자가 일관된 명령으로 모든 음성코퍼스의 정보를 알 수 있게 하였다. 이를 위해 공통의 스키마를 설계해서 다양한 음성코퍼스의 일관성 유지 및 각각의 음성코퍼스 수집 정보를 단일화된 데이터베이스 화 하였다. 본 논문에서는 이질적인 음성코퍼스를 상호 참조하기 위해 이를 하나의 통합된 시스템으로 구성하고 상호 연동을 통하여 관리자 및 사용자 입장에서는 다양한 음성코퍼스의 모음을 마치 하나의 데이터베이스처럼 투명하게 사용할 수 있도록 하였다. 이를 위해 서로 다른 디렉터리 구조 및 파일 이름을 갖는 음성 데이터들에 대한 전역 스키마를 설계하고, 구축된 데이터베이스에 대한 분석을 통해 사용자 인터페이스를 설계 및 구현하였다.

III. 통합 관리 시스템

본 논문에서 제안하는 통합 관리 시스템은 앞서 언급한 문제점 해결하기 위해 주요 기능은 다양하게 구축된 음성코퍼스에 대한 데이터의 독립성을 제공한다. 그리고 필요한 음성코퍼스를 검색하기 위해 기존의 디렉터리 구조에 대한 명세서를 참고하여 찾던 방식에서 텍스트 기입형 검색 방법으로 정보를 실시간으로 검색할 수 있도록 한다. 다음 (그림 4)는 통합 관리 시스템의 구조를 나타낸다.

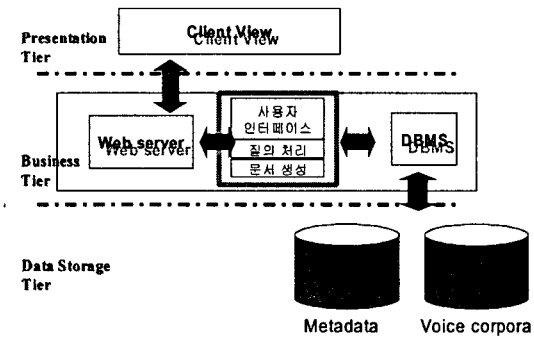

그림 4. 제안된 통합 관리 시스템 구조
Fig 4. proposed integrated management system architecture

또한 일반 사용자가 입력한 텍스트 검색 방법으로 음성코퍼스의 다양한 정보(발성자, 발성목록, 음성 파일)를 데이터베이스로부터 추출하여 이를 링크시켜 통합된 하이퍼미디어 문서를 생성할 수 있도록 한다. 이를 위한 통합 관리 시스템은 클라이언트/서버 구조로서 인터넷 기반에서 수행된다. 서버 측은 다양한 음성코퍼스를 제공하기 위한 데이터베이스 시스템과 웹 검색 및 관리를 위한 인터페이스와 연결되어 일관성 있는 하이퍼미디어 문서생성을 지원하는 인

터페이스를 제공한다. 클라이언트 측은 웹 브라우저를 통하여 서버 측과 연동하고 하이퍼 링크된 문서 생성과 검색 서비스를 요청한다. 앞서 언급한 바와 같이 통합 관리 시스템 구축 기반 모델은 3계층 구조의 클라이언트/서버 환경을 채택하였다.

시스템을 이루는 각 구성 요소의 주요 기능은 다음과 같다.

▸**사용자 인터페이스**
- 관리자는 음성코퍼스에 대한 관리부분으로 새로운 음성코퍼스에 대한 삽입, 수정, 삭제 및 각 트랜잭션에 의한 결과를 확인할 수 있도록 한다.
- 일반 사용자는 정보 검색으로 음성코퍼스에 대한 키워드 검색을 할 수 있도록 한다.

▸**질의 처리**
- 사용자 인터페이스를 경유하여 음성코퍼스 정보에 대한 검색, 입력, 삭제, 수정 처리에 해당하는 트랜잭션 처리를 한다.

▸**문서 생성**
- 사용자의 키워드 검색 요청에 대한 결과 문서 생성한다.

▸**데이터베이스**
- 다양한 음성코퍼스 정보 관리(텍스트, 전사, 음성 파일 등)를 구축한 파일 디렉터리 구조와 매핑 시킬 수 있도록 한다.

3.1 공통 음성코퍼스 정보 구축을 위한 스키마 설계

본 절에서는 기존의 구축된 음성코퍼스와 새로운 음성코퍼스 정보를 통합 관리하기 위한 스키마 구조에 대해 기술한다. 기존의 다양하게 구축된 음성코퍼스를 관리하기 위해 요구 분석 결과, 구축된 음성 데이터베이스에 대한 이질적인 구조 정보의 각 음성코퍼스에서 공통의 개체들을 추출하였으며, 이들 간의 관계 〈표 10〉을 설정하였다.

표 10. 음성 DB를 위한 개체 및 관계 정보
Table 10. entity and relationship information for speech DB

구 분	내 용	비 고
DB_information	- 전체 음성코퍼스의 정보의 기준으로 설정 - 각각의 음성코퍼스 정보 제공 - 화자(speaker)구분 정보 제공 - 전체발성목록(text_list) 정보 제공	데이터베이스의 세부적인 정보
speaker	- DB_information에서 음성코퍼스 정보 상속 speak_set 개체에 해당	발성화자에 대한 세부적인 정보
	음성코퍼스 화자 정보 부여	
text_list	- 서브 셋 목록(sub_text_list)에 각각의 전체 발성목록 정보 부여	발성 목록에 대한 세부적인 정보
sub_text_list	- 전체 발성목록 개체의 서브 셋 목록	음성파일로 저장하기 위한 전사 목록 리스트
set_voice	- 실제 음성 데이터 저장 - 서브 셋 목록에서 발성목록 상속 - 음성 데이터와 1 : 1 매칭	발성 화자별 음성 파일 리스트

이에 대한 내용은 관계 설정은 데이터베이스 모델링 툴인 ERWIN 이용하여 설계 하였다. 스키마에 대한 설계는 다음(그림 5)와 같다.

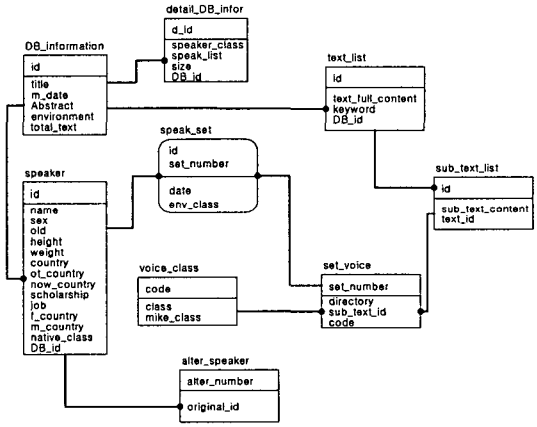

그림 5. 개체-관계 다이어그램
Fig 5. Entity Relationship Diagram

(그림 5)에서 나타난 바와 같이 스키마의 ERD는 DB_information 개체를 기준으로 화자(speaker)와 발성목록(text_list), 세부 데이터베이스 정보인(detail_DB_infor) 개체와 각각 1 : N 관계를 맺고 있으며, 화자 개체는 대체화자와 1 : N 관계를 맺고 있으며, 발성목록(text_list) 개체와 이에 대한 서브 셋 목록(sub_text_list) 개체는 1 : N 관계를 갖고 있다. 그리고 실제 음성 데이터가 저장된 개체(set_voice)는 음성 신호정보(voice_class)개체와 N : 1 관계를 갖으며, 각각의 화자 개체와 N : M 관계를 갖고 있다.

Ⅳ. 통합 관리 시스템의 구현 및 검토

이 장에서는 통합 관리 시스템의 구축환경과 사용자 인터페이스의 구현 및 수행 과정에 대해 기술한다. 그리고 사용자 측면에서 효과적인 검색을 위해 현재 구축된 음성코퍼스의 활용에 대한 분석을 통해 뷰 검색 기법을 적용하여 검색 시간을 줄이도록 하였다.

4.1 구축 환경

앞서 언급한 요구사항과 데이터베이스 스키마를 기반으로 통합 관리 시스템의 구축 환경은 다음 〈표 11〉과 같다.

표 11. 통합 관리 시스템 구축 환경
Table 11. construction environment for integrated management system

구 분	사용 환경	비 고
운영체제	Window 2000 Server	통합 시스템 플랫폼
DB엔진	MS SQL SERVER 2000	데이터베이스 관리 시스템
웹 서버	IIS 5.0	Explorer 및 Netscape
개발도구	ERWIN 4.0	데이터베이스 설계 모델링
개발언어	VB 및 Java script CGI(Common Gateway Interface)	인터페이스 및 구성요소

구축 기반은 확장성을 고려하여 3계층 구조의 클라이언트/서버 모델 채택하여 웹 기반의 시스템을 구축하였다. 관리자 또는 사용자 모두 웹 브라우저상의 인터페이스를 통하여 정보 관리 및 검색을 할 수 있도록 하였다.

4.2 구현

본 절에서는 구현된 전체 환경과 전역 스키마 구조를 기반으로 구축한 통합 관리 시스템의 데이터베이스 구조를 살펴보고 관리자와 일반 사용자의 사용자 인터페이스에 대해 설명한다.

4.2.1 데이터베이스 구축

관계형 데이터베이스에서는 단순한 자료를 처리하는 데는 효율적이지만 웹페이지, 음성파일 등과 같은 대용량의 복잡한 자료를 처리하는 데에는 제한점이 많다. 이러한 요구를 만족하기 위하여 단일프로세서와 멀티프로세서의 구조를 포함하여 하드웨어의 구조에 따른 성능을 극대화하기 위해 데이터베이스의 적용이 요구되었다. 이에 따라 다양한 음성 데이터베이스를 효과적으로 다룰 수 있는 Window 2000 Server와 MS SQL Server 2000을 사용하여 안정적이고 다양한 요구에 부합하도록 시스템 환경을 구축하였다. 다음 (그림 6)은 구축된 SQL SERVER 2000 데이터베이스의 화면을 보인다.

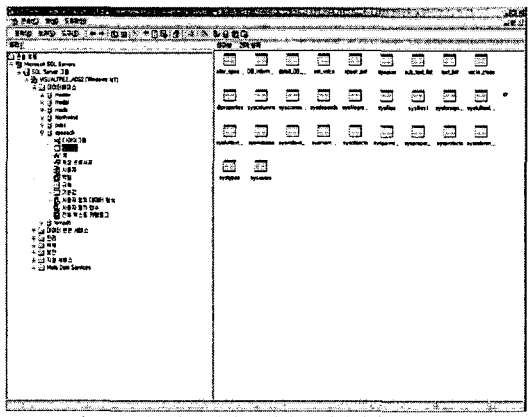

그림 6. 음성코퍼스 관리를 위한 데이터베이스 구축 화면
Fig 6. Screen of database for manage of speech corpora

(그림 6)에서 나타난 바와 같이 각각의 관련 테이블에 음성코퍼스 정보가 입력되어 있음을 보인다. 이를 기반으로 다음 (그림 7)은 사용자 인터페이스를 구현하여 검색된 음성코퍼스에 대한 다양한 결과 화면을 나타낸다.

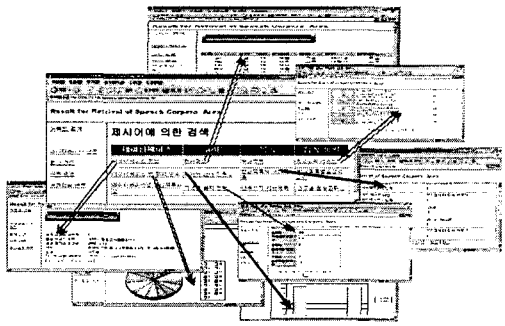

그림 7. Screenshots of the user interface
Fig 7. 사용자 인터페이스 화면

음성코퍼스 정보에는 크게 정형 및 비정형 데이터가 있

는데 텍스트 위주의 일정 형식을 갖춘 정형 데이터가 있고, 음성파일 및 문서 파일 등이 포함된 비정형 데이터가 있다. DBMS(DataBase Management System)를 통한 정형데이터의 검색에는 별다른 무리가 없으나 대량의 비정형데이터를 검색하는 데에는 두 가지 문제점이 있다.

첫째는 사용자가 질의어를 통하여 검색을 수행하였을 때 나타나는 시스템의 성능에 대한 문제이다. 텍스트 데이터의 양이 많은 경우 원하는 문서를 찾아내는데 걸리는 시간은 문서의 양과 사용자가 제시한 질의어의 수에 비례하므로 문서의 양이 일정량 이상 증가되면 점차 응답 속도가 저하된다. 둘째는 시스템이 사용자의 질의를 얼마나 근접한 순으로 문서를 랭킹하여 제공할 수 있는가 하는 문제이다.

따라서 제안된 검색 엔진을 적용하여 특수문자 처리, 사용자 정의 사전 등을 통하여 키워드를 효과적으로 추출하는 검색 시스템을 구현 하였다.

특히, 이전에 요청되었던 전역 질의에 대한 정보를 인덱스 뷰를 통해서 관리함으로써, 같은 전역 질의가 발생했을 때 처음부터 검색하는 시간적 지연을 최소화 하였다.

표 12. 일반 사용자 키워드 질의 인터페이스
Table 12. query interface of keyword for general user

NO	구 분	설 명
1	데이터베이스 관련	해당 질의어에 맞는 음성코퍼스 정보 출력 ▲ 환경별 질의 예) 자동차, 사무실, 가정집, 녹음실, 길거리 등
2	화자 관련	여러 음성코퍼스 종류에 속해 있는 모든 화자들의 특정 정보 출력 ▲ 화자정보 영역 - 전체 정보 및 세부 정보 - 여러 종류의 음성코퍼스에 속해 있는 공통 정보 통합 검색 .지역, 나이, 직업 등 .분산되어 있는 정보를 일관성 있게 출력
3	목록 관련	여러 음성코퍼스에 사용된 발성목록을 확인할 수 있다. ▲ 발성목록 - 중복된 발성목록 검출
4	음성파일 관련	발성목록에 의해 수집된 음성 파일 확인

4.3 검토

본 논문에서 구축한 통합 관리 시스템은 기존에 구축된 음성코퍼스의 저장 매체에 따르는 이질성과 데이터 포맷을 표준화하기 위하여 이를 수용할 수 있는 데이터베이스 구축과 검색을 위한 인터페이스에 중점을 두었다. 그러나 시스템의 성능문제와 직면하게 된다.

따라서 기존의 구축된 데이터베이스에서는 일반적인 질의에 대한 액세스 성능을 향상시킬 수 있는 인덱스가 있는 뷰 기법을 적용한다. 이로 인하여 최종으로 구축된 데이터베이스의 데이터 액세스 성능 향상과 작업 부하 기준으로 성능 향상을 높였다.

4.3.1 데이터 액세스 검증하기 위한 요소

구축되어 있는 데이터베이스에 데이터 액세스 비용을 비교하기 위해 뷰가 없는 경우의 일반적인 검색과 인덱스 뷰를 적용시킨 경우를 다음 〈표 13〉에 나타난 바와 같이 음성파일 중에 가장 많이 보유하고 있는 파일의 서브텍스트를 적용하였다.

표 13. 음성 파일에 해당하는 서브 텍스트 리스트
Table 13. sub text lists under speech file

NO	서브 텍스트 리스트
1	공
2	구
3	다시
4	육
5	사
6	삼
7	이
8	일
9	칠
10	하나

위 〈표 13〉 에 나타난 리스트를 검색하기 위한 질의는 다음과 같이 찾고자 하는 특정 음성 파일에 대한 서브 텍스트(txtname)를 기입하여 검색 비용을 산출하고자 한다. 다음 (그림 8)은 이에 대한 검색문을 나타내고 있다.

```
SELECT DB_information.id, text_list.keyword, sub_text
_list.sub_text_content, set_voice.directory
FROM DB_information, text_list, sub_text_list, set_voice
WHERE DB_information.id = text_list.DB_id
        and text_list.id = sub_text_list. text_id
        and  sub_text_list.id = set_voice.sub_text_id
        and sub_text_list.sub_text_content =
                        '"& txtname &"' " )
```

그림 8. 검색 수행을 위한 질의
Fig 8. query for retrieval execution

(그림 8)에서 나타난 바와 같이 검색할 서브 텍스트의 내용은 실제 구축된 데이터베이스 내에 입력된 값들 중에 가장 빈도가 높은 텍스트로 300개 이상인 음성 파일을 위주로 일반 검색과 인덱스 뷰 기법을 이용한 검색 비용을 비교하였다.

인덱스 뷰는 SQL Server 2000에서 제공하는 기능을 이용하여 다음 (그림 9)와 같이 Create문을 사용하여 view_test라는 뷰를 생성하였다.

```
CREATE VIEW view_test WITH SCHEMABINDING
AS
SELECT DB_information.id, text_list.keyword,
                    sub_text_list.sub_text_content,
                set_voice.directory
FROM DB_information, text_list, sub_text_list, set_voice
WHERE DB_information.id = text_list.DB_id and text_list.id
                = sub_text_list. text_id
          and  sub_text_list.id = set_voice.sub_text_id
               and  sub_text_list.sub_text_content
                Go
CREATE UNIQUE CLUSTERED INDEX view_testnd
          ON view_test(DB_information.id)
```

그림 9. 인덱스 뷰
Fig 9. index view

4.3.2 검색 방법에 따르는 검색 비용 비교

앞서 언급한 입력 키워드별 검색하였을 경우, 일반적인 검색과 인덱스 뷰를 이용한 검색 비용에 대한 차이를 (그림 10)을 통해 보이고 있다.

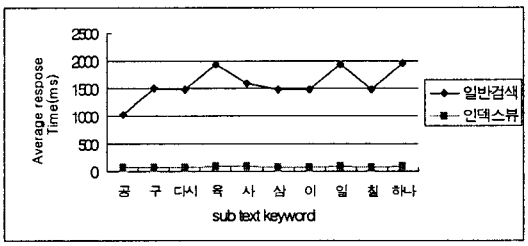

그림 10. 일반적인 검색과 인덱스 뷰 검색 방법의 비교
Fig 10. comparison between general retrieval and index view retrieval method

(그림 10)에서 나타난 바와 같이 일반적인 검색 방법에 의한 질의 결과는 평균 응답 시간으로 1584.5msec이며, 인덱스 뷰 검색 방법에 의한 질의 결과에 드는 비용은 76.4 msec로 1508.1msec의 차이를 보이고 있다.

따라서, 통합 관리 시스템의 검색 시간을 줄이기 위해 인덱스 뷰 검색방법을 채택하여 데이터의 액세스 비용을 줄였다.

V. 결론

본 논문에서는 기존 음성코퍼스 및 신생 음성코퍼스의 다양성과 이질성부분을 효과적으로 통합하기 위한 방법들을 제안하였다. 본 논문의 목적은 다양한 음성코퍼스의 이질적인 구조를 수용 및 확장할 수 있는 통합 관리 시스템을 구현하여 일관성을 유지시키는데 있다.

또한 인터넷을 이용한 단일화된 인터페이스를 통해 분산되어 있는 음성코퍼스의 정보를 통합 관리 시스템 내에서 효과적인 검색을 통하여 사용자가 원하는 정보를 제공하는데 있다.

따라서, 본 논문에서는 서로 다르게 구축된 음성코퍼스를 통합 관리하기 위한 시스템을 제안하였다. 이는 기존의 구축된 음성코퍼스뿐만 아니라 신생 음성코퍼스를 관리하는데 있어 효과적인 관리와 사용자를 고려하여 웹 기반의 통합 관리 시스템 형태로 구축하였다.

이를 위해 먼저 기존 음성코퍼스의 재사용과 새로운 음성코퍼스에 대한 파일 구조를 분석하여 공통요소의 스키마를 설계하고, 이를 데이터베이스로 구축하였다. 그리고 전체 시스템의 환경은 웹 기반의 3계층 클라이언트/서버 구조를 채택하여 확장성을 고려하였으며, 사용자 측면에서는 웹 브라우저를 경유하여 음성코퍼스에 대한 정보 관리와 검색이 가능하도록 하였다. 그리고 통합된 검색 인터페이스를 통하여 접근 시간을 고려하여 일반적인 검색 방법과 인덱스 뷰 검색 기법을 적용하여 필드 테스트를 거쳐 시스템에 대한 성능평가를 하였다.

향후 연구 내용으로는 현재까지 구축된 음성코퍼스 정보에 대해 데이터웨어 하우스 기술을 접목하기 위한 연구가 필요하다. 또한 응용으로 정보가전 분야에 음성 정보 안내 및 음성인식 시스템에 적용하고자 한다.

참고문헌

[1] 지진구, 윤성일, "음성을 이용한 화자 검증기 설계 및 구현", 한국컴퓨터정보학회 논문지, 제 5권, 3호, pp.91-98, 2000.9.

[2] 이기희, "선형 변환망을 이용한 화자적응 음성인식", 한국컴퓨터정보학회 논문지, 제 5권, 2호, pp.90-97, 2000.6.

[3] 오종훈, 최기선 , "은닉 마르코프 모델을 이용한 과학기술문서에서의 외래어 추출 모델", 제 11회 한글 및 한국어 정보처리 학술대회, pp.137-141.

[4] G.Doddington, "The next generation DARPA speech recognition/natural language DataBase," Proc. Of the ESCA Workshop, Speech I/O Assessment and Speech DataBase, 1989.

[5] ETRI, 음성/언어 정보연구부, http://voice.etri.re.kr

[6] 원광대학교 음성정보기술산업지원센터, http://speechnet.or.kr/

[7] 이영직 외, "ETRI의 음성 데이터베이스 구축현황," 한국 음향학회 제 12 회 음성통신 및 신호처리 워크샵 논문집, 1995. 6.

[8] 최인정 외, "KAIST 통신 연구실의 음성 데이터베이스 구축 현황," 한국 음향학회 제 12 회 음성통신 및 신호처리 워크샵 논문집, 1995.6.

[9] 김락용 외, "LG 전자 음성DB 구축 현황," 한국 음향학회 제 12 회 음성통신 및 신호처리 워크샵 논문집, 1995. 5.

[10] 김상룡 외, "삼성종합기술원의 음성 DB 구축 현황," 한국 음향학회 제 12 회 음성통신 및 신호처리 워크샵 논문집, 1995. 6.

[11] NIST : Speech Corpora Produced on CD-ROM Media by The National Institute of Standards Technology(NIST), April, 1991.

[12] F.Guyote, et al, "A Speech DataBase at the United States Air Force Academy," Proc., ICASSP-86

[13] Ronald Cole, et al, "Towards Automatic Collection of the U.S Census," 1994 IEEE ICASSP.

[14] 황경훈 외 3명, "다양한 음성코퍼스의 통합관리시스템의 설계 및 구현에 관한 검토", 2003 대한음성학회 가을 학술대회 발표논문집, pp.69-72, 2003.10.17-18.

저 자 소 개

유 경 택

1988년 원광대학교 전자계산공학과 졸업

1990년 광운대학교 전자계산기공학과 졸업 (공학석사)

2006년 원광대학교 컴퓨터공학과 졸업 (공학박사)

1995년 ~ 현재 : 극동정보대학 컴퓨터정보과 부교수

〈관심분야〉 XML, 멀티미디어데이터베이스, 분산컴퓨팅, 시스템통합

정 창 원

1993년 원광대학교 컴퓨터공학과 졸업

1998년 원광대학교 컴퓨터공학과 졸업 (석사)

2003년 원광대학교 컴퓨터공학과 졸업 (공학박사)

2004년 ~ 현재 : 전북대학교 차세대 LBS 응용 연구센터 연구교수

〈관심분야〉 분산객체 컴퓨팅, 멀티미디어 데이터베이스, LBS, 텔레매틱스

김 도 관

1993년 원광대학교 정보관리학과 졸업

1999년 Texas Tech Mass Com. Telecommunication 전공 석사 졸업 (MA)

2006년 전남대학교 경영학과 MIS 박사과정 수료

2004 ~ 현재 : 원광대학교 경상대학 정보전자상거래학부 IT 초빙교수 (프로그래밍 전문강사)

〈관심분야〉 멀티미디어, 전자상거래 응용 기술, 기업 정보화, 시스템통합

이 용 주

1976년 고려대학교 전자공학과 졸업

1986년 고려대학교 전자공학과 졸업 (석사)

1992년 고려대학교 전자공학과 (공학박사)

1980년 ~ 1994년 한국전자 통신연구소

1994년~현재:전기전자 및 정보공학부 교수

2001년 ~ 현재 : 원광대학교 음성정보산업지원센터 센터장

〈관심분야〉 HCI, 신호처리, 멀티미디어