음성 DB 부가 정보 기술방안 표준화를 위한 제안*

김상훈(ETRI), 이영직(ETRI), 한민수(ICU)

<차 례>

<Abstract>

Standardization for Annotation Information Description of Speech Database

Sanghun Kim, Youngjik Lee, Minsoo Hahn

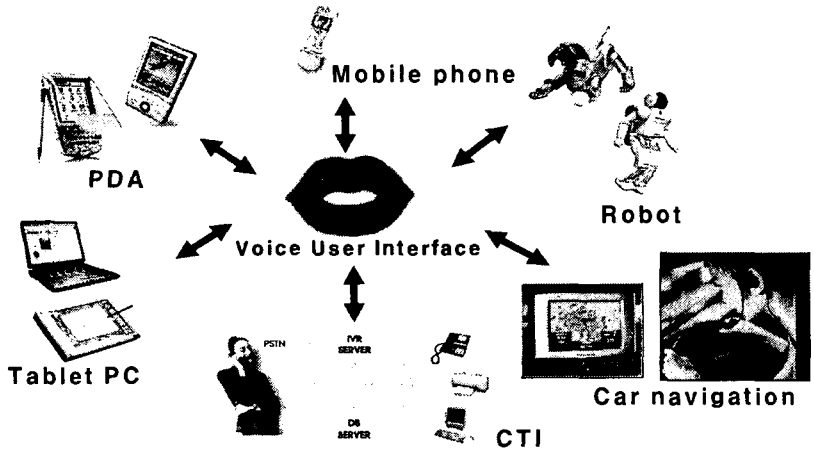
This paper presents about the activities of speech database standardization in ETRI. Recently, with the support of government, ETRI and SiTEC have been gathering the large speech corpus for the domestic speech related companies. First, due to the lack of sharing the knowledge of speech database specification, the distributed speech database has a different format. Hence it seems to be needed to have the same format as soon as possible. ETRI and SiTEC are trying to find the better representation format of speech database. Second, we introduce a new description method of the annotation information of speech database. As one of the structured description method, XML based description will be applied to represent the metadata of the speech database. It will be continuously revised through the speech technology standard forum during this year.

* Keywords: speech DB, standardization, XML, metadata

* 본 연구는 정보통신부 출연 "언어정보처리기술개발" 사업으로 수행되었습니다.

1. 서 론

최근 각종 휴대형 정보단말기, CTI 및 텔레메틱스 등 사용자 편의성을 증대하기 위한 음성인터페이스(VUI: Voice User Interface)가 매우 중요한 요소 기술로 부각되고 있다. 특히, 각종 정보기기의 소형화와 이동성에 따라 언제, 어디서나 정보를 획득할 수 있는 정보 검색 서비스가 향후 2-3년 내에 일반화될 것으로 예상되고 있고(가트너그룹 조사에 의하면 2004년까지 e-commerce 고객 중 40%가 무선을 통해 접속할 것으로 예측), 이러한 정보기기의 자연스러운 입력방식으로 음성 인터페이스의 적용이 필수적인 조건으로 될 것이다.



<그림 1> Voice User Interface 응용

최근 국제표준화 기관인 ETSI, ITU-T, IETF에서는 분산음성인식기술(DSR) 국제표준화를 추진하고 있으며, 마이크로소프트/CISCO/Intel이 주도하고 있는 SALT (Speech Application Language Tag) 포럼에서는 멀티모달 환경에서의 음성인터페이스에 대한 표준화 작업을 수행하고 있는 바, 가까운 시일 내에 음성기술은 점점 상용화되어 일반인들에게 매우 친숙한 기술이 될 것으로 예상된다.

이와 같이 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 DARPA를 통하여 수집한 DB를 기반으로 1990년대 초에 LDC (Linguistic Data Consortium)를 구성하였고[2], 유럽에서는 1995년에 비영리기관으로 ELRA (European Language Resource Association)를 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 개

발에 활용하여 왔으나, DB 구축에 많은 시간과 비용이 소요되어 외국업체에 비해 음성 기술 경쟁력이 떨어진다는 지적을 받아왔다. 이에 따라 정부도 음성 DB의 필요성을 인식하고 ETRI, SiTEC을 통해 정부 출연 대량의 공통 음성/텍스트 DB 구축이 진행되고 있다.

공통 음성 DB는 사용자가 필요로 하는 음성언어 정보가 충실히 표현된 DB, 사용자들이 쉽게 가져다 활용할 수 있는 구조화된 DB 구성이 요구된다. 따라서 대량의 음성 DB를 효율적으로 활용할 수 있도록 음성 DB 표준화의 필요성이 제기되고 있다. 현재 다양한 형태로 존재하는 음성 DB의 음성언어 정보 구조 및 표현 내용, 음성 데이터 포맷 등 기관별 각기 달리 표기되고 있는 음성언어 정보의 통일된 표기 방법이 필요하다. 이러한 표준화 작업은 음성인식 및 음성합성 기술 개발시 음성 DB의 활용도를 높일 수 있고, 제품 개발 기간을 단축시킬 수 있다. 특히 다양한 기관에서 구축한 음성 자원을 공동으로 활용할 수 있게 해줌으로써 국가적으로 자원의 효율적 사용이 가능해지며, 국내 업체의 해외 경쟁력이 대폭 강화될 수 있는 계기가 될 수 있다.

2. ETRI 음성 DB 구축 현황

2002년도에 구축한 음성 DB는 PC, 통신망 등 환경별로 25개 분야에 걸쳐서 수집되었다[4][5]. 통신망 환경은 유/무선 전화망, VoIP, PC 환경으로 세분화된다. PC 환경은 헤드셋과 마이크로 구분하였고, 성능에 따라 중가, 저가로 분류하였다. 응용에 따라 DB는 다시 음성인식용 단어, 숫자, 문장으로 나눠진다. 문장에는 낭독체 문장, 대화체 문장, 언어모델링용 문장, 음성합성용 문장, 화자인식용 단어/문장으로 다양하다.

음성인식용 단어는 10,000 단어를 대상으로 1,000명의 화자가 발성한 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 발성 목록은 상장 회사명, 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. 구성 화자는 성별, 연령별, 지역에 따라 적정 비율로 분포되었다. 한편, 구축된 DB는 80%를 훈련용으로, 20%를 평가용으로 구분하여 패키징하였다. 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 투자 방지하고 음성 서비스 시장을

활성하는데 크게 기여할 것으로 판단된다. 본 DB는 HCI (Human Computer Interface), CTI (Computer Telephony Interface), 텔레매틱스(Telematics), 생체 정보 인식, 시각 장애인용 음성 응용, 자동 통역 등 각종 음성인식 및 합성 엔진 개발에 활용될 것이다.

3. 국내외 표준화 동향

국외 동향으로, 유럽에서는 프랑스의 ELRA (European Language Resource Association)를 주축으로 DB 구축, DB 표준화, 평가, 배포 목적으로 SpeechDat 프로젝트를 수행해 오고 있다. 표준화 대상으로는 수집 시스템, 수집 환경, 음성언어 정보, 음성 DB, 전사 방법, 검증 방법, 평가 방법 등 다양한 분야에 걸쳐 표준화를 수행하고 있다. 본 프로젝트에 참여한 기관은 벨기에 L&H, 독일의 Siemens, 이탈리아의 CSELT 등 모두 17개국 기관이 참여하고 있다.

COCOSDA (International Coordinating Committee on Speech Databases and Assessment)는 "음성 입출력 평가법 및 음성 데이터베이스"에 관한 워크샵으로 주로 음성인식 및 합성 시스템의 성능 평가 방법 및 음성 데이터베이스를 다루고 있다. EAGLE (Expert Advisory Group on Language Engineering Standards)는 EU 지원으로 구성된 전문가 그룹으로 언어 공학적 응용을 위한 각종 스펙의 가이드라인을 제시하고 음성 및 언어 과학에 대한 연구 및 응용에 대한 광범위한 컨설팅과 기존에 구축된 음성언어 자원의 평가 및 효율적인 구축을 위한 방법론을 제시해 주고 있다. 한편 ISO/TC 37/SC 4 에서는 XML 기반 LR (Language Resource) 메타데이타(morpho-syntactic annotation 등) 표준화 작업을 추진하고 있으며 텍스트 자원에 대한 표준화 작업은 국제적으로 상당히 진행이 되고 있다.

국내 동향으로는, 2001년 KAIST 전문용어공학연구센터에서는 과학기술부와 한국과학기술평가원의 기술 용역 사업에 의한 "대용량 음성(음향)/언어/영상 DB 구축 및 표준화"과제에 대한 사업 발표회를 개최하였다. 이 발표회에서 음성/언어제품 개발 및 제품 평가를 위한 평가용 DB와 개발용 DB에 대한 설명회와 공개 토론회를 갖고 시범 CD-ROM 및 자료집을 배포한 바 있다. 한국정보보호진흥원과 한국전자통신연구원은 생체 인식용 생체 특징(지문, 홍채, 얼굴, 음성)의 데이터 교환을 지원하기 위한 생체 데이타 포맷 표준화 작업을 진행 중에 있고 표준 데이타 구축도 병행하여 추진하고 있다.

최근에는 산업자원부 지원 표준화 포럼에서 음성 정보 처리 관련 용어 정리를 추진하였고, 현재 보완 작업 중이다. DB 표준화와 관련, 원광대 SiTEC에서는 음소 분할 기준, 운율 정보 표기 체계화 등 주로 학술적인 목적으로 표준화를 시도하고 있으며[3], ETRI 음성/언어정보연구센터에서는 SiTEC과 공동으로 음성 DB 단체 표

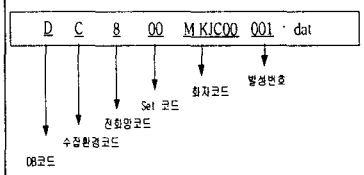
준화를 추진하고 있다.

4. 음성 DB 표준화 제안

음성 DB를 용이하기 다루기 위해서는 각 기관에서 정의된 표현 방식을 우선 통일하는 것이 사용자에게는 큰 도움이 된다. 특히 음성 DB가 대용량일수록 전처리에 소요되는 시간이 길어지고, 또한 각 기관마다 표현 방식이 다를 경우, 기관마다 다른 프로그램으로 파일 처리를 해야하는 애로를 겪게 된다. 따라서 표준화 이전에 데이터 저장 또는 표현 방식을 통일하는 것이 실질적인 작업을 위해 무엇보다 필요하다. 이에 ETRI와 SiTEC에서 대량의 음성 DB를 구축하고 있는 바 양 기관에서 서로 상이하게 사용하고 있는 부분을 통일하고자 각 기관에서 사용하고 있는 파일 명명 방법, 디렉토리 구조, 음성언어 부가 정보 표현 방식 등을 <표 1>과 같이 정리 비교하였다. 대표적으로 ETRI의 경우, 파일명에 파일에 관한 모든 정보(DB 종류, 수집 환경, 통신망 종류, 성별, 화자명 등)를 표현하고 있으며, SiTEC의 경우, 디렉토리명에 모든 정보를 표현하고 있고 파일명은 화자명과 일련번호로 표현하고 있다. 그 외 파일 확장자, 파일에 헤드 정보를 포함하는지, 전사 표기 방법 등에서 서로 상이한 부분이 있다. 이중 어느 표현 방식이 더 편리하냐 하는 것은 SiTEC과 ETRI간 협의를 통해 통일안을 작성할 예정이다.

음소기호의 경우, 국내 대표 기관이 사용하고 있는 현황은 <표 2>와 같다. 현재 각 기관이 고유의 음소기호를 정의하여 사용하고 있으나 기관간에 사용하고 있는 음소가 1:1로 주로 일치가 되기 때문에 단순히 매핑해서 사용하더라도 현재로서는 큰 문제가 없다. 그러나 통일된 기준을 마련하고 앞으로 점진적으로 통일할 수 있도록 하는 것이 대용량 데이터에 대한 음소기호 처리시 시간적인 낭비 요인을 미리 막을 수 있을 것이다. 또한 음성 관련 학계나 중소 업체에 음소기호의 사용 기준을 제시함으로써 연구 결과에 대한 객관적인 성능 비교가 가능할 것으로 보인다. 음소기호 통일안은 업체나 학계의 의견을 좀 더 수렴하여 진행할 예정이다.

<표 1> SiTEC와 ETRI간 공통 음성 DB 표현 방식 비교

구 분		SiTEC	ETRI
디렉터리 표기	구조	-/DB명/환경/성별/지역/나이/화자	-/DB명/환경/세트/화자
	명명법	(성별): m | f (지역): 2자리 코드 지정 (나이): 1자리 코드 지정 (화자): Initial 3자리+숫자 2자리	(환경): 유/무선 | 마이크 | 헤드셋 | VoIP (세트): 'set'+2자리 숫자 (화자): Initial 3자리+숫자 2자리 + (반복횟수 2자리
파일 표기	구조	-(발성목록명)(발성번호).??? 예) PBS0001.wav	-(DB명)(수집환경)(전화망종류)(세트)(성별)(발성번호).??? 예) DC800MKJC00001.dat 
	명명법	-(발성목록명): pbs | pbw | prw -(발성번호): 4자리 숫자 예) PBS0001.wav	-(수집환경): 1자리 코드 지정 -(전화망코드): 1자리 코드 지정 -(set 코드): 2자리 숫자 -(화자코드): 성별 1자리+Initial 3자리+숫자 2자리 -(발성번호): 3자리 숫자 예) DC800MKJC00001.dat
확장자 표기	음성	- *.wav (with header)	- *.dat (without header)
	전사	- *.trn	- *.txt
	래링고	- *.lar	- *.lay
전사 표기	구조	-1개 파일을 사용하고 2개 층을 소리층, 철자층으로 분리하여 표기	-철자전사와 발음전사를 2개 파일로 분리하여 표기
	표기법	-숫자/기호/외래어 표기 심볼 사용 예) <소리전사> 키 <일><미터> 이다 <철자전사> 키 <1><m> 이다	-한글로 모두 변환 예) 키 일 미터이다
	전사 기호		-잡음기호 정의

<표 2> 각 기관별 음소기호 표기법 비교

구분	국어 표기법	문화관광부 (고시 제2000-8호)	IPA	연구계		산업계			학계
				ETRI		S사	K사	B사	SITEC
				합성	인식				
모음	ㅏ	a		a	a	a	aa		a
	ㅓ	ae		v	v	v	vv		v
	ㅗ	o		o	o	o	oo		o
	ㅡ	eu		U	U	_	xx		U
	ㅣ	i		i	i	i	ii		i
	ㅐ	ae		E	E	E	ai		E
	ㅔ	e		e	e	e	ee		e
	ㅚ	oe		we	we	oi	oi		O/wA
	ㅟ	wi		wi	wi	ui	ui		Y/wi
	ㅞ	we		we	we	we			we/wA
	ㅜ	u		u	u	u	uu		u
	ㅑ	ya		ja	ja	ya			ja
	ㅕ	yeo		jv	jv	yv			jv
	ㅛ	yo		jo	jo	yo			jo
	ㅠ	yu		ju	ju	yu			ju
	ㅒ	yae		jE	jE	yE			jE/jA
	ㅖ	ye		je	je	ye			je/jA
	ㅘ	wa		wa	wa	wa			wa
	ㅙ	wae		wE	wE	we			wE/wA
	ㅝ	wo		wv	wv	wv			wv
	ㅢ	ui		Wi	Wi	_i	xi		xi
자음	ㄱ	g/k		g/K	g	g/KK	gl		g
	ㄲ	kk		G	G	gl	gg		G
	ㅋ	k		k	k	k	kh		k
	ㄷ	d/t		d	d	d/TT	dl		d
	ㄸ	tt		D	D	dl	dd		
	ㅌ	t		t	t	t	th		t
	ㅂ	b/p		b/P	b	b/PP	bl		b
	ㅃ	pp		B	B	bl	bb		B
	ㅍ	p		p	p	p	ph		p
	ㄹ	r/l		r/L	r	r/LL	ll		r/l
	ㅈ	j		z	z	j	jl		z
	ㅉ	jj		Z	Z	jl	jj		Z
	ㅊ	ch		c	c	c	ch		c
	ㅅ	s		s	s	s	sl		s
	ㅆ	ss		S	S	sl	ss		S
	ㅎ	h		h	h	h	hl		h
	ㄴ	n		n/N	n	n/NN	nl		n
	ㅁ	m		m/M	m	m/MM	ml		m
	ㅇ	ng		0	0	0	ng		N
변이음	유성음화								V
	무성음화								Q
기타	묵음				#		ql		sil
	탈락				-				-
	잡음						zz		

음성언어 정보 파일의 부가 정보 포맷인 경우, SAM project에서 정의되었던 표기 방법 <표 3>을 참조하되[6] XML (Extensible Markup Language) 형식으로 확장 정의하여 표준화 할 계획이다.

<표 3> SAM project에서 사용된 음성언어 부가 정보 포맷

DBN: <DB명>
SRC: <파일명>
SEX: <성별>
AGE: <나이>
DLT: <지역>
SAM: <sampling frequency>
SBF: <byte order>
OTS: <철자전사>
PTS: <발음전사>
EOF:

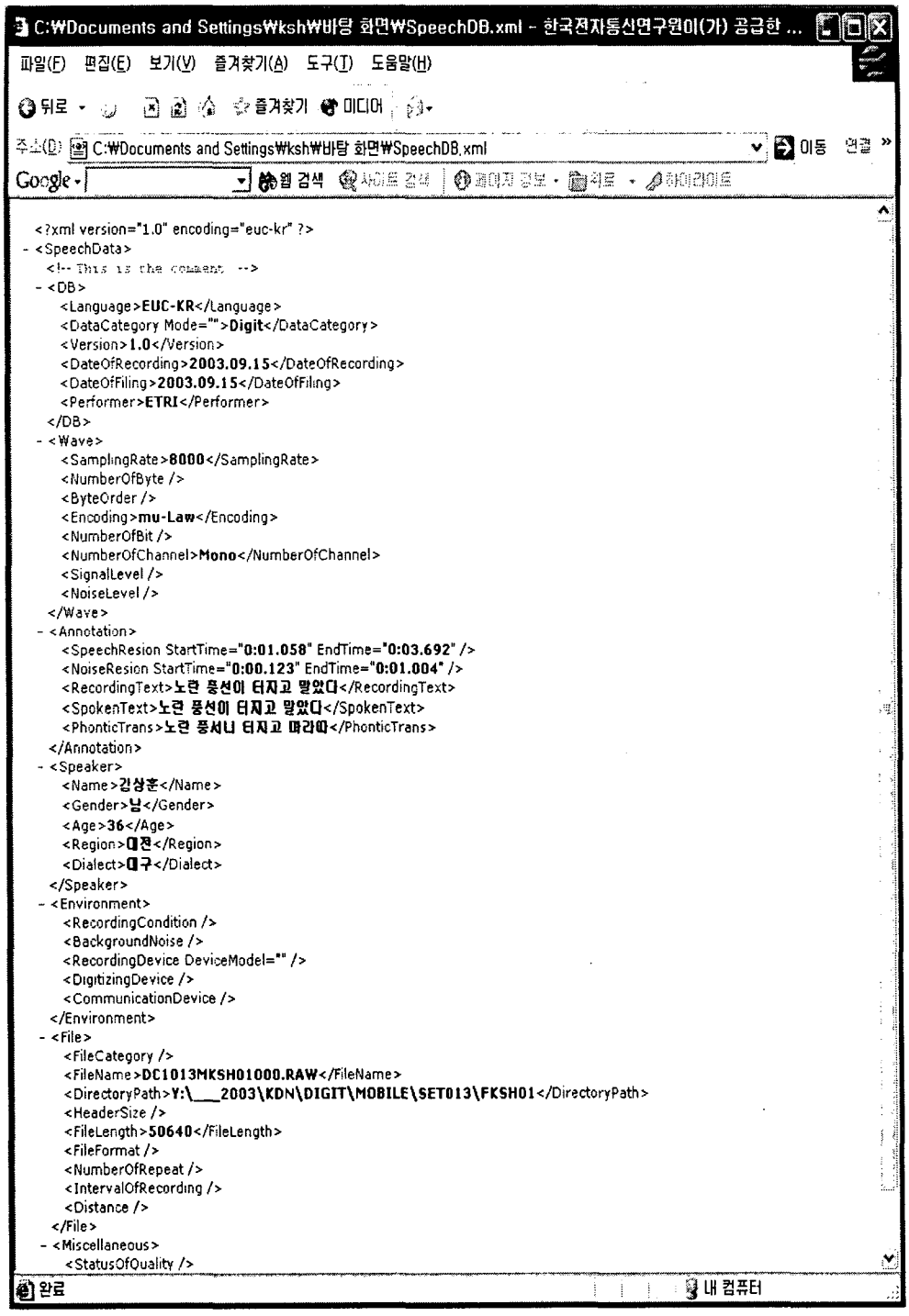
음성 DB를 XML로 표기하기 위해서는 메타데이타 표기를 위한 의미 태그 선정, 메타데이타 표현 구조 그리고 편집/검색 하기 위한 도구가 필요하다. 이들 중 현재 음성 DB 메타데이타 표현용 의미 태그 선정 작업을 음성 정보 처리 기술 포럼을 통해 추진하고 있으며 1차 선정된 결과는 <표 4>와 같다.

<표 4> 제안된 부가 정보 포맷

대분류	속 성		세부 속성	비고
	영문 표기	의 미		
기본 정보 (DB_Info)	Language	언어		
	Version	버전		
	ApplicationCategory	응용분야		
	NumberOfSpeaker	발성화자수		
	NymberOfUtterance	발화수		
	DataCatagoy	DB종류		
	RecordingDate	녹음날짜		
	FilingDate	수정날짜		
	RevisionHistory	수정기록		
	Distributer	수행기관		

대분류	속성		세부 속성	비고
	영문 표기	의 미		
음성 정보 (Wave_Info)	SamplingRate	샘플링주파수		
	NumberOfByte	샘플당바이트수		
	ByteOrder	바이트순서		
	EncodingLaw	인코딩방법		
	NumberOfBit	비트수		
	NumberOfChannel	채널수		
	SignalToNoiseRatio	SNR		
전사 정보 (Annotation_Info)	SpeechSection	음성구간	StartTime EndTime	
	NoiseSection	잡음구간	StartTime EndTime	
	RecordingTrans	녹음문장		
	SpokenTrans	철자전사		
	PhoneticTrans	발음전사		
화자 정보 (Speaker_Info)	SpeakerName	화자명		
	Sex	성별		
	Age	나이		
	Region	지역		
	Dialect	방언		
환경 정보 (Environment_Info)	RecordingEnviron	발성환경		
	NoiseEnviron	잡음환경		
	RecordingDevice	녹취장비	DeviceModel	
	DigitizingDevice	A/D장비		
	CommunicationDevice	통신종류		VoIP
파일 정보 (File_Info)	FileCategory	파일종류		
	FileName	음성파일명		
	DirectoryPath	파일위치		
	HeaderSize	헤더크기		
	FileLength	파일길이		
	FileFormat	파일포맷		
	NumberOfRepeat	반복차수		
	TimeInterval	녹음주기		
	Distance	발성거리		
기타 정보 (Miscellaneous_Info)	QualityStatus	품질상태		

1차 안으로는 기본 정보 등 7개의 대분류와 42개의 속성으로 의미 태그를 분류하였다. 의미 태그 선정은 현실적으로 가능한 정보를 우선 표기하고 추후 필요로 하는 정보에 따라 태그를 정의하여 추가할 계획이다. <그림 2>는 정의된 메타데이타 표기용 의미 태그를 이용하여 XML 스크립트를 작성한 예를 보여주고 있다.



<그림 2> XML기반 음성 DB 메타데이타 표현 예

5. 결 론

ETRI 음성/언어정보연구센터에서는 정보통신부 출연 "언어 정보 처리 기술 개발" 사업의 일환으로 음성정보처리산업협의회 산하 음성정보처리 포럼을 통해 음성 DB 표준화 작업을 수행하고 있다. 첫 번째, 각 기관간에 상이하게 사용하고 있는 DB 표현 방식(파일 명명, 디렉토리 구조 등)을 통일하고자 하며, SiTEC과 공동 작업을 통해 추진하고자 한다. 두 번째, 음성 DB의 부가 정보인 음성언어 메타데이타를 XML로 표기하기 위한 국외 표준화 현황을 조사하고, 이로부터 포럼에서 자체적으로 국내 표준을 새롭게 작성해야 하는지 아니면 국제 표준을 따라야 하는지 등 음성 DB의 XML 표기 표준화 방향을 결정하고자 한다. 국내 표준을 작성할 경우, 국내 공통 음성 DB의 음성언어 부가 정보 XML 표기 first draft를 2003년 11월까지 관련 전문가의 의견 수렴 후 작성하고 2003년 12월 공청회를 거쳐 표준화 가능한 부분을 우선 표준화하고자 한다. DB 표준화는 음성 DB의 사용을 매우 용이하게 해주며, 각 기관간의 DB 공유도 활성화할 수 있어 국내 음성 자원의 효용성을 극대화할 수 있는 계기가 될 것으로 보인다.

참 고 문 헌

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

[2] LDC home page: http://www.ldc.upenn.edu.

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

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

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

[6] SpeechDat Technical Report "Specification of speech database interchange format".

접수일자: 2003년 8월 22일
게재일자: 2003년 9월 17일

▶ 김상훈(Sanghun Kim)
주소 : 305-761 대전광역시 유성구 전민동 464-1 엑스포아파트 2405-907
소속 : 한국전자통신연구원 컴퓨터소프트웨어연구소 음성/언어정보연구센터
음성DB연구팀
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

▶ 한민수(Minsoo Hahn)
주소: 305-732 대전광역시 유성구 화암동 58-4번지
소속: 한국정보통신대학원대학교 음성/음향정보연구실
전화: 042) 866-6123
FAX: 042) 866-6110
E-mail: mshahn@mail.icu.ac.kr