음성언어연구와 음성 데이타베이스

이용주, 정유현, 한남용, 최준혁
한국전자통신연구소

요약

한국어 음성 및 언어정보처리를 위해 필수적인 우리말 음성 데이타베이스의 구축을 위하여 먼저 각국의 동향을 살펴보고, 용도, 필요성, 기술적 고려사항에 관하여 검토한후 현재 전자통신연구소에서 수행하고있는 관련연구활동및 계획을 소개한다.

1. 서론

음성을 맨머신인터페이스의 수단으로 활용하기위한 음성정보처리 연구는 관련기술의 진보에 따라 한정어휘의 인식및 합성시스템들은 실용화에 접어들고 있고, 이제 임의어휘를 대상으로하는 시스템들이 연구 개발되고 있다. 특히, 임의어휘의 인식시스템 개발에는 음소단위 인식기술 개발이 필수적이지만, 연속음성중의 음소는 발성자에 따른 개인차는 물론이고 전후에 발성되는 음소의 영향에 의한 조음결합에 따라 그 특성이 크게 변화한다. 이러한 개인차 및 조음결합의 현상을 분석하기 위해서는 많은 사람이 발성한 다양한 음성데이타가 필요하다. 또한 시스템의 객관적인 평가를 위해서 표준적인 음성데이타도 필요하다. 전자를 연구용 데이타베이스, 후자를 공통음성데이타 라고 부르고 있다. 이와같은 목적으로 외국에서는 이미 여러기관에서 이를 제작하여 CDROM 등에 수록하여 공동으로 이용할수 있도록 하고 있으나 우리나라는 아직 거기까지는 이르지 못하고 있으며 일부 기관을 제외한 대부분의 기관에서 나름대로 자체 제작하여 내부적으로만 사용하고 있어 데이타량 및 이용형태가 제한되고 또한 각 연구자가 발표한 인식시스템의 성능 및 분석방식의 평가를 각 연구자의 데이타에 의존하고 있어서 객관적으로 이루어지지않고 있는 실정이다. 따라서 음성데이타베이스의 체계적인 구축이 시급하다.

본고에서는 음성데이타베이스에 관한 각국의 동향과 함께 용도, 필요성, 구축시의 기술적 고려사항등을 고찰한후 우리말 데이타베이스 구축을 위한 전자통신연구소의 관련활동에 대해서도 간략히 소개하고자 한다.

2. 외국의 동향

2.1 미국

미국에서의 음성 데이타베이스는 음성인식 장치의 객관적 평가를 목적으로 그 필요성이 본격적으로 제기되었으며, 1981년 미국 TI사의 음성 연구팀이 시판되고 있는 음성인식 장치를 대상으로 성능 비교를 실시하였다. 1982년 미국에서 음성 입출력 기술 workshop이 개최되어, 음성 입출력 장치의 제조 회사, 음성연구자 및 이용자까지 참가하여 음성 입출력 기술의 표준화를 위한 기본적인 문제가 토의되었다. 미국 연방표준국(NBS)은 음성인식장치의 성능평가를 위한 발성자, 환경, 어휘, 평가법,평가에 이용되는 음성데이타, 녹음과 시험법, 평가항목 등의 지침을 작성하였다. 미국 과학 아카데미는 1984년 많은 정부기관의 요청을 받아 컴퓨터 음성 인식기술 위원회를 발족시켰다. 이 위원회의 보고서에서 "소수의 실험적인 데이타베이스이외에 앞으로 널리 쓰일 공통 음성 데이타베이스와 평가방법을 만들어 중앙의 책임하에 보관 및 배포를 담당하도록 하자"는 제안을 하였고, 이에 따라 공동으로 사용할 필요가 있는 음성 데이타를 국립표준기술연구소(NIST)가 중심이 되어 제작, 배포하고 있다. 현재 미국은 단어음성 데이타베이스뿐만 아니라 연속음성 데이타베이스 구축에 관한 연구도 활발히 수행하고 있으며, 특히 DARPA의 연속음성인식 Project에서의 공통의 task인 "자원관리"에 대한 연구를 중심으로 TI, MIT, CMU, NIST, AT&T의 Bell Lab., IBM, NYNEX가 아래와 같은 내용을 수록하고, NIST에서 이를 CD-ROM으로 제작하여 국립기술정보서비스(NTIS)를 통해 유료로 관련연구기관에 배포하고 있다.

- o 불특정화자 학습용 3,360문장[80명 x 42문장]
- o 특정화자 학습용 7,344문장[12명(남7명,여5명) x 612문장]
- o 평가용 4,320문장[특정화자 12명, 불특정화자 개발용 40명, 불특정화자 평가용 20명]
- o 특정화자 연속음성 Corpus 10,508문장[남2명 x 여2명 x 2,627문장]
- o Acoustic-Phonetic 연속음성 Corpus(TIMIT)[음성인식시스팀 개발 및 평가용, 630명분의 8개 방언 문장과 10개의 보통 문장으로 구성]
- o 불특정화자 연속 숫자 Corpus(TIDIGITS)
  - 326명 x 77개 숫자음

- 발성자 구성 비율 : 남자 111명, 여자 114명, 소년 50명, 소녀 51명

향후 추진 계획으로는 목표지향의 task로서 "항공여행 안내 데이타베이스", "항공제어 데이타베이스", "전화교환 데이타베이스" 등을 구상하고 있다.

이외에도 CMU에서는 1,000명의 10,000문장의 음성을 수록하고 있으며, MIT는 100명이 발성한 회화문 9,692문장을 수록하고 있다.

2.2 유럽

유럽에서는 NATO 음성처리연구 그룹에서 NATO 가맹국중 6개국이 참가하여 음성인식 기술의 유용성에 대해서 검토를 하고, 1983년 19명이 발성한 자국어 및 타국어의 연속음성 35,000어에 대한 음성 데이타베이스를 만들어 각종 인식 시스팀의 성능평가에 이용하고 있다.

프랑스에서는 GRECO Project에 의해 1987년에 아래와 같은 시스팀 평가용 공통 음성데이타와 연구용 음성 데이타베이스를 구축하였다.

o 공통 음성 데이타
  - 시스팀 학습용 2분간의 문장
  - 숫자(100개), 3연숫자(50개), 4연숫자(50개), 5연숫자(50개), 연속음성(50개),
  - 0~99숫자음(100개), 고립발성 단어(50개), 연속음성(50개), 음성합성용 minimal pair list

o 연구용 음성 데이타베이스
  - 2분간의 테스트 패턴, 단모음 15개
  - CVCVC형 음절, CVCV형 음절

1990년 7월부터는 GRECO-PRC CNRS Communication Man-Machine program, ACCT(Agencee Cooperation Culturelle et Technique), and ESPRI POLYGLOT project에 의해 불특정 및 특정화자 음성인식 시스팀 개발용으로 발성자당 10,000단어씩 100명분을 목표로 하는 음성데이타베이스 개발을 시작하였다.

2.3 일본

일본의 음성 데이타베이스에 관한 연구는 주로 전자기술총합연구소(ETL), 동북대학, 일본전자공업진흥협회(JEIDA), ATR 자동번역전화 연구소, 오오사까 대학 등에서 활발히 수행되었다. 전자기술총합연구소에서는 1973년부터 음성 데이타 파일 제어 시스팀을 만들어 사용하였으며, 그 후에 단음절, 연속음성 등을 수록하고 단어음성에 대해서는 음소보다 작은 단위로써 레이블링한 음성 데이타베이스를 구축하였다. 동북대학에서는 단어 음성 데이타를 수집하기 위한 시스템을 개발하여 자기 디스크와 자기테이프에 음성 데이타베이스를 구축하고, 최근에는

이를 CD-ROM판으로 만들었다. 오오사카 대학은 음성 연구자들의 공동 이용을 목표로 1983년 범용 DBMS INQ를 이용한 음성 데이타베이스 SPEECH-DB를 개발하여 각 대학의 대형 계산기 센터의 네트워크를 통해 TSS 단말로 엑세스할 수 있도록 구성하였다. 일본전자공업진흥협회에서는 1982년에 음성연구에 관련되는 기업, 대학, 정부의 연구자로 구성된 위원회에서 음성 입출력의 표준화를 논의하였으며, 주로 음성 데이타베이스에 관해 검토하고 이를 일본어 공통 음성 데이타로 구축하여 현재 일본 국내 50개 기관에 배포하여 사용하고 있다.

현재 일본에서 사용되고 있는 중요 음성 데이타베이스는 다음과 같다.

가. ETL 음성 데이타베이스

- o phoneme-balanced 단어 1542개 x 남 10명
- o Phonetic labeling 음성 데이타베이스

나. 동북대학 음성 데이타베이스

- o 212 단어 : 남자 32명, 여자 40명, 16,800단어
- o 3000단어 : 남자 15명, 여자20명, 113,480단어
- o CD-ROM & Optical Disk에 저장

다. JEIDA 음성 데이타베이스(일본어 공통 음성 데이타)

- o 323단어 x 150명(남 75명, 여 75명) x 4회
- o 발성자 구성 비율 : 20대 45명, 30대 45명, 40대 30명, 50-60대 30명
- o 323단어 종류

  단음절, 도시명, 숫자,기능어
- o 일본 국내 40여 기관에 배포

라. JEIDA Noise 데이타베이스

- 0 18종류의 소음데이타를 DAT에 수록

마. ATR 음성 데이타베이스

- o Isolated word JSDB(Japanese Speech Database)
  - 중요어(5,229개), phoneme-balanced 단어 (216개), alphabet (35개)
    숫자 (25개), 단음절 (101개)
  - 대화문 115개 : 음소 레이블링
  - 발성자 : 남녀 각 10명의 어나운서와 나레이터
  - 음성인식, 합성 연구용
  - 대학, 연구기관에 유료로 배포
- o Isolated sentence JSDB
  - 503개 문장 : 신문, 잡지 참조

- acoustic-phonetic transcription & grammatical information 포함
- 발성자 : 8명
- prosodic characteristics 분석용

o Mixed word and sentence JSDB
- 단어 750개와 150개 문장
- 불특정화자 음성인식연구용
- 92년까지 100명분 제작

o Text JSDB
- 250 문장 : 교과서, NHK TV text book 참조
- 발성자 : 2명의 나레이터
- 음성합성연구용

o ATR의 향후 추진 계획
- 20문장 x 100명분 이상의 연속음성 데이타베이스 구축
- 전문가의 스펙트로그램 관찰에 의한 6계층으로 레이블링

그외에 기업연구소에서 음성인식 및 합성장치의 연구 개발 과정에 대규모의 음성 데이타베이스를 구축하여 사용하고 있지만 아직 공개하지는 않고 있다.

일본은 앞으로 본격적인 연속음성 데이타베이스를 구축할 계획을 추진중에 있으며, 이를 위해 1990년 6월에 일본음향학회에 "연속음성 데이타베이스 조사위원회"가 발족되었다. 동시에 국가적 규모의 "음성언어 데이타베이스 연구소"가 설립되어야 한다는 안도 제시되고 있다.

3 국내 현황

우리나라의 경우에는 아직 외국처럼 체계적으로 만든 음성 데이타베이스가 보급되고있는 상태는 아니며, 대학 및 연구소에서 실험에 필요한 숫자음, 단음절, DDD지역명, 고립단어 등의 소규모의 음성 데이타를 연구자 각자가 자체 제작하여 사용하고 있다. 단지 전자통신연구소에서 시험 제작한 일부 데이타만이 몇몇대학에 배포되어 시험적으로 사용되고 있는 정도이다.

4. 음성데이타베이스의 구축을 위한 고려사항

4.1. 음성 데이타베이스의 의의

음성정보처리 연구에 공통으로 이용 가능한 대량의 각종 음성 데이타를 수집, 편집,배포하는 일은 다음과 같은 점에서 의의가 매우 크다고 생각된다.

- 연구 개발자 입장에서는 분석, 합성, 인식 등의 알고리즘 개발 평가에 이용

할 수 있다.

- 음성인식.합성 시스템의 사용자 입장에서는 각종 시스템의 성능 평가를 할 수 있다.

4.2. 음성데이타베이스의 내용

우리말 음성 데이타베이스의 구축을 위해서는 발성내용, 발성자의 범위, 녹음조건 등의 여러가지 조건을 고려하여야 한다.

가. 발성내용

- 단음절 : 단모음, 자음+모음, 모음+자음, 자음+모음+자음 등
- 단어 : 단독 숫자음, 지명(시도명, 전철역명 등), 최소음소쌍, 고빈도단어, Phoneme Balanced Word, 기능어(은행서비스용, 컴퓨터조작용, 가전제품용 등)
- 연속 단어 : 연속숫자음, DDD번호, 우편번호 등
- 문장 : 일기예보, 음운 발란스 문장 등

나. 발성자

연령별, 직업별, 출신지별(12세 이전 거주지를 기준으로 함), 학력별로 고루 분포되어야 하며 1차적으로는 표준말만을 고려하나 불특정화자의 인식등 화자의 개인성연구를 위해서는 앞으로 방언도 그 대상이 되어야 한다.

다. 데이타량

- 특정화자용 : 소수화자, 최저 2회 이상 발성
- 불특정화자용 : 다수화자, 최저 1회 이상 발성(2회 이상이 바람직)

4.3. 음성 데이타 수록과 편집

공통음성 데이타베이스를 구축하기 위하여 필요한 음성 데이타의 수록과 편집시 고려해야 되는 중요 사항들은 다음과 같다.

가. 수록조건

- 녹음장소 : 무향실, 방음실, 조용한 사무실, 잡음이 있는 방, 컴퓨터실 등
- 잡음의 종류 : 백색잡음, 자동차 잡음 등
- 입력장치 : 마이크, 전화기
- A/D변환 : 필터특성, 표본화 특성, 양자화 비트수

나. 기록매체

아날로그 테이프, DAT, CD, Flopy disk, CD-ROM, 광 디스크 등이 있으나, 보존성, 기억용량 및 대량복사를 고려하여 볼 때 DAT, CD-ROM이 유리.

다. 발성지시방법

- 발성내용 지시: 리스트, 화면, 음성

- 발성타이밍 지시: 발성자가 자유롭게 발성,신호음, 화면

라. 녹음

- 음성 수록에 있어서는 음성 데이타 수록시 추후에 데이타 편집을 쉽게하기 위하여 단어사이에 1초이상의 간격을 둔다.
- 마이크로폰에서 입까지의 거리는 10-30cm을 기준으로 한다.
- 수록단어 앞에 발성자를 식별하기 위하여 발성자 번호, 이름, 날짜 등을 녹음한다.
- 주변의 소음 레벨은 50dBA 이하를 목표로 한다.

마. 데이타체크

녹음시 다음과 같은 무제가 발생하기 쉽다.

- 발성리스트와 상이(순서변경, 오독, 중복), 무의미 발성, 잡음혼입, 발성레벨, 발성속도, 발성간격

따라서 반드시 녹음후 체크가 필요하며 틀린곳은 재발성.

바. 편집

1)편집하지 않음 : 작업량은 적지만 기록매체의 량은 많아진다(2초 이상의 무음구간은 1초 정도로 단축하는 일은 해야 한다.)

2)무음구간의 편집 : 각 단어의 시작점 및 끝점 전후에 300ms의 무음구간을 둔다(단어간의 간격을 600ms로 편집한다). 편집작업의 자동화가 가능하다.

3)라벨을 붙임 : 각 테이프에 단어명, 발성자, 날짜, 수집장소 등의 정보를 부가하여 편집을 한다.

5. 음성 데이타베이스 구축을 위한 ETRI의 활동

국내에서의 조직적인 음성DB의 구축을 위한 방안들이 5~6년전부터 제안되어 왔으나 여건의 불비로 만족할만한 결과는 아직없다. 그러나 이 분야에 꾸준한 관심을 보여왔던 ETRI의 경우, 음성인식 및 합성을 위한 다음과 같은 데이타들이 내부 용도로 수집되어 사용되어 오다가 최근에 일부 대학들에게 시험용으로 제공된 바 있다. 즉,

- 단독숫자 22종 x 10명
- 4연숫자 35종 x 10명
- CV단음절 140종 x 10명
- PBW 445단어 x 4명 (음소단위 레이블링)

그러나 위 데이타들은 공개를 목적으로한 것들이 아니므로 사용에 주의를 요한다. 최근에 과학기술처의 지원을 받아 공개를 목적으로한 공통음성데이타의 제작

이 시도되어 1차적으로 100명이 9회 발성한 단독숫자및 4연숫자 60종의 데이타가 수집되어 정리중에 있다. 이밖에도 공통음성데이타의 지속적인 확장을 위하여

- 고빈도단어
- 음운 벨런스 단어 (PBW)
- 음운연쇄

등과 같은 고립단어의 수집과 함께

- 음운벨런스문장
- 한정된 테스크에서의 대화문 (예: 호텔예약)

과 같은 연속음성의 수록도 계획하고있다.

이러한 작업을 위하여 음성자동수집및 편집시스템(발성내용제시 시스템 포함)과 음성데이타베이스 관리시스템이 구축되어 있다.

한편, 고빈도 단어 추출을 위한 300만어절 규모의 텍스트데이타베이스를 구축중에 있으며, 2음소열, 3음소열 등의 통계적 분포를 활용한 음운벨런스단어(PBW)의 추출, 최소음소쌍이나 임의의 음소열을 포함한 단어의 추출, 발음규칙의 확인등에 활용할수 있는 발음사전 DB가 구성되어 있다.

이밖에도 계속 추진되어야 할 사항으로는

o 대량의 음성 데이타 저장 및 관리 방법

o 연구목적에 적합한 음성 데이타 검색 시스템 개발

o 표준 녹음기기, 녹음환경 등의 설정과 기재준비 및 기록매체 선택

o 음성 데이타의 추가변경 관리 시스템 개발

o 음성 편집 시스템 개발(잡음제거, 정정발성, 발음편집, 부수정보의 입력등)

o 음소 단위의 레이블링 기술

초기에는 레이블링하지 않은 데이타가 필요하나 , 음소단위 인식 기술 개발을 위해서는 음소 레이블링 음성 데이타베이스가 필요하므로 레이블링 전문가의 확보 및 자동 레이블링에 관한 기술 개발이 필요하다.

o 대상 단어의 선정을 위한 분석 및 조사

o 방언 데이타베이스 구축

불특정화자 음성인식기술 개발을 위해 향후 추가되어야 할 사항이다.

6. 결론

지금까지 음성정보처리 연구에 있어서 기본적인 연구 개발 도구인 동시에 개발내용의 객관적인 평가의 기준이 되는 음성 데이타베이스에 관하여 기술하였다. 음성정보처리 연구분야에서 앞서가는 나라들은 이러한 음성 데이타베이스의 중요성

을 인정하고 80년대 초부터 국가적인 차원에서 구축하여 이를 CD-ROM 또는 DAT에 수록, 배포하고 있으며, 계속적으로 확장(발성자 수 증가, , 연속문장 , 방언 포함)하고 있다.

음성 데이타베이스는 구축시 많은 시간과 노력을 필요로 하기 때문에 우리나라에서도 음성정보처리연구의 저변확대 및 활성화를 기하고 동시에 음성분석, 음성합성 및 음성인식의 개발내용에 대한 객관적인 평가 기준을 제공하는 음성 데이타베이스 구축에 관한 국가 차원의 집중적이고 지속적인 연구 개발이 이루어져야 할 것으로 생각된다.

참고문헌

[1] 정유현,이용주, " 음성데이타베이스의 연구동향 및 전망 "
한국음향학회지 제10권 4호, 1991

[2] Joon-Hyuk Choi, et al, "Construction of A Large Korean Speech Database and Its Management System In ETRI, Proc., 24.2.1, ICSLP 90.

[3] 이용주, 이정철, 김경태, "음성 데이타베이스 구축에 관하여",
한국음향학회지, 제7권 제5호, 1988

[3] 이용주, 김경태 외, "단어음성 데이타 수집 및 DB 구성 시스템",
대한전자공학회 추계 종합학술대회 논문집, Vol. 9, NO.2, 1986.12

[4] 이용주, 김경태 외, "대용량 발음사전 표제어에 나타난 음소의 통계적 성질"
대한전자공학회 통신교환연구회 발표논문집, 1987.11

[5] 김경태, 최준혁, 이용주, " 음성 데이타베이스 관리 시스템의 구축",
Korea-Japan Joint Symposium on Acoustics, 1991. 7

[6] 정유현, 최준혁 외, "공통 음성 데이타베이스 구축을 위한 사전 조사 연구",
대한전자공학회 하계 학술 대회, 1992.6.

[7] 한국전자통신연구소, "자동통역전화를 위한 요소 기술 개발 "
최종보고서, 1991.12.

[8] 한국전자통신연구소, "보급형 음성데이타베이스 구축에 관한 연구"
최종보고서, 1992.7

[9] Shuichi Itahashi, "Recent Speech Database Project in Japan",
Proc., 24.1.1, ICSLP 90.

[10] NIST: Speech Copora Produced on CD-ROM Media by The National Institute of Standards andTechnology(NIST), April, 1991

[11] Seiichi Nakagawa, "Assessment and Database of Speech Recognition /Understanding Systems,, (in Japanese), 일본전자정보통신학회, 90.2

[12] Y. Sagisaga, et al, "A Large-Scale Japanese Speech Database",
Proc., 24.4.1, ICSLP 90.

[13] K. Iso, et al, " Design of a Japanese Sentence List for a Speech Database", Proc. ASJ. 2-2-19, March, 1988

[14] T. Ehara, et al, " ATR Dialogue Database", Proc., 24.5.1, ICSLP 90

[15] A. Kurematsu, "ATR Japanese speech database as a tool of speech recognition and synthesis",Speech Communication, 9, 1990

[16] R. Mizoguchi, et al, "Speech Database with an Intelligent Access Mechanism - SPEECH-DB", 일본 정보처리학회논문집, May 1983

[17] J. Gauvain, et al, "Design Considerrations and Text Selection for BREF, a large French read-speech corpus, Proc., 24.6.1, ICLSP 90

[18] K. Tanaka, "The Speech Database for Speech Analysis and Recognition Research", Proc., 24.7.1, ICLSP 90

[19] S. Makino, et al, "A Distributed Speech Database with an Automatic Acquisition System of Speech Information", Proc. 24.91, ICLSP 90

[20] K. Shikano, :"Phonetically balanced word list based on information entropy", Preprints Autumn Meeting Acous. Soc. Japan, Paper Mar. 1984

[21] S. Hayamizu, et al, "Genetation of VCV/CVC Balanced Word Sets for Speech Database", 일본 전자기술총합연구소, 제49권 제10호, 1985

[22] K. Akiba, et al, "Speech Database for Research of Japanese Speech Recognition", 일본음향학회 강연논문집, 1982.3

[23] G. Doddington, "The next generation DARPA speech recognition/natural language database", Proc. of the ESCA Workshop, Speech I/O Assessment and Speech Database, Sep. 1989

[24] J.m. Baker, et al, "Speech Recognition Performance Assessments and Available databases", Proc.,ICASSP 83