

공통 음성DB의 구축 : 현황 및 향후 계획

이 용 주

김 봉 완

원광대학교 컴퓨터공학과
전북 익산시 신룡동 344-2, 우:570-749
yjlee@wonkwang.ac.kr

원광대학교 컴퓨터공학과
전북 익산시 신룡동 344-2, 우:570-749
joe@speech.wonkwang.ac.kr

Construction of Korean Speech DB for Common Use : Status and Future Direction

Yong-Ju Lee

Department of Computer Engineering,
Wonkwang University

Bong-Wan Kim

Department of Computer Engineering,
Wonkwang University

요약

음성의 과학적인 연구를 위해서는 대상이 되는 음성을 체계적으로 수집 정리한 음성 데이터베이스가 필요하다. 본 논문에서는 공동 이용을 목적으로 본 연구팀에서 지금까지 구축한 음성 DB의 중간 결과 및 향후 구축계획에 대하여 기술한다.

1 서론

음성 인식 및 합성 등 우리말 음성 정보 처리 시스템의 개발을 위해서 가장 먼저 확보해야 할 것이 다양한 사람이 발성한 대량의 음성 데이터베이스이다.

따라서 음성 관련 연구자들은 모두 이러한 음성 데이터베이스가 연구 초기부터 필수적으로 필요하나 이의 확보에는 많은 시간과 예산, 그리고 노력이 필요하며 또한 전문적 기술이 필요하다. 따라서 국가적인 차원에서 체계적으로 구축하여 공동으로 활용케 하는 것이 바람직하며 선진 각국에서도 국가나 공공 연구 기관이 주축이 되어 활발하게 구축, 활용되고 있다.

우리 나라도 이러한 추세에 부응하여 공동 이용을 목적으로 한 음성DB가 과기처가 지원하는 STEP 2000 프로젝트 중 국어 정보 처리 기술 개발의 일환으로 구축되고 있다.

본 논문에서는 이와 같은 목적으로 1단계('95 ~ '97)에서 구축된 음성DB의 결과 및 2단계의 1차년도 결과를 중심으로 기술한다.

2 1단계('95 ~ '97) 구축 결과

2.1 기본 음성DB 구축

공동 이용을 목적으로 사용 될 음성 데이터베이스는 적은 발성 목록에 실제 한국어의 발성에 나타나는 음운 현상을 가능한 많이 포함하며, 특정 테스트에 집중되지 않는 것이 바람직하다. 이러한 조건을 충족시키는 발성 목록으로 선진 외국에서는 음운 균형(Phonetically Balanced) 발성 목록을 사용해 왔다[1,2,3].

음운 균형이 취해진 상태란 목록에 포함된 음운 현상들을 확률사상으로 했을 때 엔트로피가 최대인 상태를 말한다. 목록에 나타난 음운 현상의 총 종류를 N , 각 음운 환경의 출현 확률을 $p_i (i=1 \sim N)$ 이라고 할 때, 목록의 엔트로피는 다음 식(1)에 의해 구할 수 있다.

$$E = - \sum_{i=1}^N p_i \log p_i \quad (1)$$

선정된 발성 목록은 다음의 조건을 만족한다. 첫째, 모집단에 나타난 모든 음운 현상을 포함한다. 둘째, 최소한의 발성 목록으로 구성되어 있어야 한다. 셋째, 포함된 음운 현상들간의 확률 분포가 최대한 고르게 분포한다.

따라서 본 연구에서도 공동이용을 위한 음성 데이터베이스의 발성 목록으로 음운 균형 발성목록이 적합하다는 판단아래 100만 어절 이상의 텍스트 코퍼스에서 단어 레벨 PBW 452어절[4], 문장 레벨 PBS 589문장[5]의 발성목록을 작성하고 음성 데이터베이스를 구축하였다.

PBW를 선정하기 위한 음운 환경으로는 발생 가능한 모든 2음소열을 기본 단위로 하였으며, PBS선정 시에는 특히 3음소열로 고려해야 할 유

(제 10회 한글 및 한국어 정보처리 학술대회)

성음화 및 자음약화, 탄설음화의 경우를 유형별로 클러스터링하여 음운 환경으로 정의하고 발성목록을 선정하였다[6].

무제한 어휘의 인식기 및 합성기를 위한 음성 데이터베이스의 발성 목록으로 테스트에 종속적이지 않으면서, 한국어의 음운 환경을 충분히 고려한 음운 균형 발성 목록외에도, 음성 연구자들이 꾸준히 요구하는 것이 숫자음 음성이자. 관련 연구에 의하면 필요로 하는 음성 데이터에 대한 설문 등에서 숫자음 음성은 우선 순위가 높게 나타나고 있다[7]. 따라서 기본 음성 데이터베이스로 음운 균형 발성 목록외에도 단독 숫자음, 4연 숫자음, 단문, 고빈도어를 대상으로 음성 데이터베이스를 구축하였으며 각 발성 목록별 규모는 다음 표와 같다.

표1. 발성 목록별 규모

발성 목록	발성 횟수
단독 숫자음	일반인 70명 X 4회 발성 어나운서 2명 X 2회 발성
4연 숫자음	일반인 70명 X 4회 발성 어나운서 2명 X 2회 발성
단 문	일반인 70명 X 2회 발성 어나운서 2명 X 2회 발성
고빈도 2000어절	어나운서 2명 X 2회 발성
PBW	일반인 70명 X 2회 발성
PBS	일반인 20명

2.2 레이블링 기준안 작성

레이블링된 음성 데이터베이스는 음성 인식 및 합성기의 훈련 및 평가, 성능 개선을 위한 중요한 자료로서 활용된다. 그러나 음성 데이터베이스를 레이블링하는 작업은 대부분 수작업으로 이루어지고 있으며 또한 그 기준마저 통일되지 않고, 각 연구자의 필요에 따라 작성하여 사용하고 있으므로, 기준 및 결과가 공동 이용을 목적으로 사용하기에는 부적합하거나 객관적 검증이 이루어지지 못하고 있는 실정이다.

이에 본 연구에서는 공통적으로 이용할 수 있는 레이블링된 음성 데이터베이스를 구축하기 위한 한국어 레이블링 기준안 초안[8]을 작성하여 PBW 10명분을 레이블 하였다.

이 초안은 향후 여러 분야에서 공통으로 활용할 수 있는 기준안이 되기 위해 다양한 사용자 요구사항을 조사, 대상으로 하는 음성 시료의 확대, 관련 분야 전문가들의 객관적의 평가 및 검증 작업을 거치고 있다.

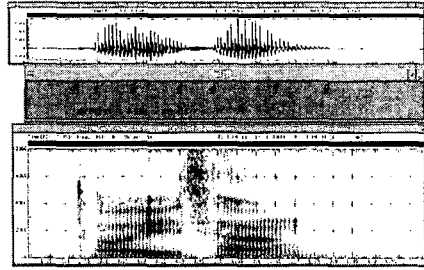


그림1. 음소 레이블링

3 2단계의 중간 결과

3.1 음성DB 구축을 위한 Workbench의 개발

음성 데이터베이스의 구축을 위해서는 다양한 소프트웨어의 도움이 필요하다. 따라서 본 연구에서는 음성DB구축 및 활용을 위한 Workbench를 개발하였다.

Workbench는 음성 언어 자료의 확보를 위한 텍스트 처리 모듈들과 음성 데이터의 처리를 위한 신호처리 모듈들로 구성되어 있다. Workbench에 포함된 모듈로는 발성 목록 선정 모듈, 끝점 검출기를 이용한 음성 데이터 편집 모듈, 다단계 레이블링 시스템, 텍스트에서 원하는 음운 환경을 포함하고 있는 문자열을 다양한 조건으로 검색할 수 있는 음운 환경 검색기를 포함하고 있다.

3.1.1 발성 목록 선정 모듈

발성 목록 선정 모듈은 대량의 텍스트 코퍼스를 입력으로 받아 앞에서 철자 음운 변환을 통해 음소열로 변환한 후, 이로부터 음운 균형 발성 목록을 선정해 주는 모듈이다.

3.1.2 클라이언트/서버 음성 수집 시스템

음성 데이터베이스를 구축하기 위해 음성 시료를 수집할 경우 사용자에게 발성 목록을 어떻게 제시해 줄 것인가 하는 점도 중요한 요소가 된다. 따라서 본 연구에서는 사용자에게 컴퓨터 모니터를 통해 자동으로 발성 목록을 제시해 주고, 또한 동일한 발성 내용을 다양한 환경(다른 PC기종-예를 들면 노트북과 데스크탑 PC, 다른 종류의 사운드 카드, 다른 종류의 마이크 등)에서 수집하기 위해 클라이언트/서버 음성 수집 시스템을 개발하였다.

이를 위해 음성 수집 클라이언트에서는 각각

(제 10회 한글 및 한국어 정보처리 학술대회)

의 환경(사운드 카드, 마이크 등)을 설정하고 네트워크를 통해 수집 서버에 접속한다. 접속 후 클라이언트가 서버의 통제에 의해 수행하는 기능은 다음과 같다.

- 발성 화자 별 음성 데이터의 관리
- 발성 목록의 제시
- 발성 시간 정보의 표시



그림 2. 클라이언트/서버 음성 수집 시스템



그림 3. 음성 수집 클라이언트

수집 서버는 접속된 수집 클라이언트들에 대하여 다음과 같은 기능을 수행한다.

- 클라이언트 별 음성 수집 통제
- 발성 목록의 관리(전송, 재발성)
- 발성 화자의 통제(발성시작, 휴식 등)
- 발성 화자 정보의 관리

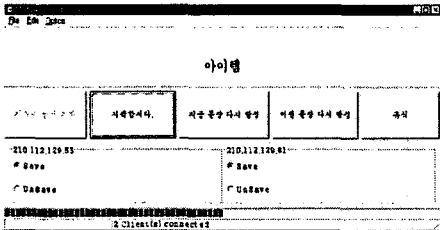


그림 4. 음성 수집 서버

따라서 발성 화자는 복수의 마이크 앞에서 한번만 발성하면 각각의 클라이언트에서는 네트워크를 통해 서버의 통제를 받아 음성 시료를 수집함으로써 음성 시료 수집의 효율성을 높일 수 있다. 또한, 동일한 음성 내용에 대한 다양한 환경의 음성 시료를 확보할 수 있으므로 마이크, 사운드 카드 등 환경에 의한 영향을 분석할 수 있는 시료를 얻을 수 있다.

3.1.3 음성 편집 시스템

음성 편집 시스템은 수집된 음성 데이터가 연속적으로 저장되어 있을 경우, 데이터 중 무음 구

간, 잘못 발성된 구간을 제외하고 원하는 음성 데이터만을 파일로 저장하기 위한 모듈이다.

음성 편집 시스템에서는 단구간 영교차율, 에너지, 주파수 에너지의 파라미터를 사용하여 발성 구간을 자동 검출하고 앞과 뒤에 일정 구간의 무음 영역(400ms)을 확보하여 사용자에게 보여 줌으로써, 사용자는 양호한 데이터인지를 판별하여 파일로 저장하면 된다.

다음 그림은 본 모듈을 이용하여 자동차 소음 환경에서 끝점을 검출하고 음성 데이터의 편집을 수행하는 그림이다.

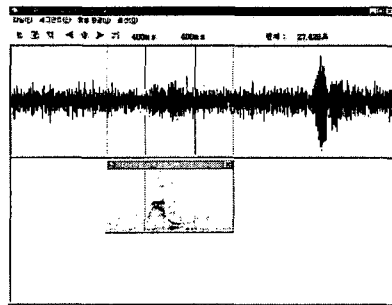


그림5. 음성 편집 시스템

3.1.4 다단계 음성 레이블링 시스템

레이블링을 위해서는 레이블링 틀이 필요하나 현재 대부분의 소프트웨어가 유닉스(Unix)기반으로 되어 있거나 전용 하드웨어를 요구하므로 보급이 미흡한 실정이고, 그나마 대부분이 고가이므로 연구자들이 손쉽게 접하기 힘든 면이 있다.

구현된 레이블링 시스템은 경계 분할 정보, 음성 정보, 언어 정보 등의 다른 계층을 두고 레이블링을 수행할 수 있도록 되어 있다. 레이블링 시스템에서 제공하는 개략적 기능은 다음과 같다.

- 웨이브 폼 보기, 확대
- 음성 구간 듣기(전체, 확대, 선택 영역)
- 스펙트로그램 보기

다음 그림은 구현된 레이블링 시스템을 이용하여 레이블링을 수행하는 화면을 나타낸 것이다

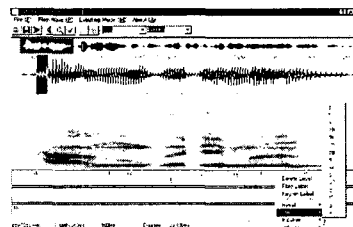


그림6. 다단계 레이블링 시스템

(제 10회 한글 및 한국어 정보처리 학술대회)

3.1.5 음운 환경 검색 시스템

공동 음성 데이터베이스의 활용도를 높이기 위해서는 음성 데이터베이스에 포함된 음운 환경을 검색하고 분석할 수 있는 방안이 필요하다.

예를 들어, 한국어의 유성음화에 대한 분석을 수행할 경우, 음성 데이터베이스의 발성 목록에서 유성음화가 발생할 수 있는 조건인 “모음 + 연자음 + 모음”의 음운 환경을 포함하고 있는 발성 목록을 선별하고 분석을 수행할 수 있다면 음성 데이터베이스의 활용도를 높일 수 있다.

본 시스템은 이러한 목적을 달성하기 위하여 발성 목록을 입력으로 받아, 원하는 조건을 만족하는 발성 목록을 포함하고 있는 문장 또는 단어를 선별해 주는 시스템이다.

본 시스템에서 지원하는 검색 방법은 문자열 검색, 음소열 검색 및 음운 환경 검색의 3가지이다. 문자열 검색은 발성 목록에서 원하는 문자열을 포함하고 있는 목록을 선정하여 출력하는 방법이고, 음소열 검색은 발성 목록에서 원하는 음소열을 포함하고 있는 목록을 선정하여 출력하는 방법을 말하며, 이러한 경우 발성 목록을 철자음운 변환을 거쳐 변환한 음소열에서 검색을 수행하게 된다.

음운 환경 검색은 특정 음소열을 지정하지 않고 위의 예와 같이 “모음 + 연자음 + 모음”처럼 음운 환경을 만족하는 유형을 지정하여 검색을 수행하는 방법을 말한다.

본 시스템에서 “고모음 + 파열 연자음 + 고모음”을 검색하는 경우를 다음 그림에 나타내었다.

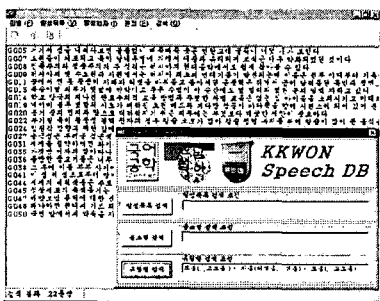


그림7. 음운 환경 검색 시스템

3.2 기본 음성DB의 심층화

현재 진행하고 있는 2단계에서는 이미 구축된 음성 데이터베이스의 심층화를 목표로 하여 진행하고 있다.

먼저 레이블링 데이터베이스의 규모를 확대하여 추가로 PBW 50명분의 음성 데이터 레이블링

작업을 완료하였다.

최근 방음 부스에서 녹음된 클린 스피치 뿐만 아니라 전화 환경, PC 환경 등 다양한 환경에서 발생한 음성 데이터베이스에 대한 요구가 증가하고 있다. 따라서 이러한 목적의 음성 데이터베이스를 설계하고 프로토타입을 구축하였으며 그 특성은 다음과 같다.

- 발성 목록
 - 4연 숫자 (35종)
 - 지역명 (200종)
- 발성 환경
 - 핸드폰
 - 유선전화(사무실, 가정, 공중전화)
 - 코드리스 전화
- A/D 환경
 - PC 환경
- 화자별 발성량
 - 35명 X 65 단어 = 2275단어

또한 이미 구축된 문장 레벨 음성 데이터베이스인 PBS의 양이 충분치 않다고 판단되기 때문에 추후 음성 데이터베이스의 양을 확장 할 예정이다.

3.3 레이블링 기준의 검토 및 보완

초안으로 작성된 레이블링 기준을 여러 분야에서 공통으로 활용할 수 있도록 관련 분야 전문가들의 객관적의 평가 및 검증작업을 통해 지속적으로 개선해 나가고 있다. 현재 검토하고 있는 내용은 다음과 같다.

- 단모음 + 단모음
- 연속된 동일한 모음
- 음절말 파열음
- 파열음 + 파열음
- 파열음 + 파찰음
- 비음 + 비음
- 공명 음화 표기
- 어절말 경계

4 결론

음성 인식 및 합성 등 우리말 음성 정보 처리 시스템의 개발을 위해서 가장 먼저 확보해야 할 것이 다양한 사람이 발성한 대량의 음성 데이터베이스이다.

본 논문에서는 공통으로 이용하기 위한 음성 데이터베이스의 현재 구축 현황과 앞으로의 구축 방향에 대하여 기술하였다.

현재 구축된 음성 데이터베이스의 내용으로는 음운 균형 발성 목록으로 70명이 2회씩 발성한

(제 10회 한글 및 한국어 정보처리 학술대회)

단어 레벨의 PBW 452어절, 20명이 발성한 PBS 589문장, 기타 단독 숫자음, 4연 숫자음이 있다.

앞으로의 구축 계획은 구축된 음성 데이터베이스의 양을 지속적으로 확대하고, 레이블링 작업을 통하여 심층화할 계획이다.

참고문헌

- [1] K. Shikano, "Phonetically balanced word list based on information entropy," Proc. of ASJ., 1984
- [2] Hayamuzu, "Generation of VCV / CVC balanced word sets for speech data base," Bulletin of Electrotechnical Laboratory, 1986
- [3] K. Iso, et al, "Design of a Japanese Sentence List for a Speech Database," Proc. ASJ., 1988
- [4] 이용주, 김봉완 외, "공통 이용을 위한 음성 DB의 설계 및 구축에 관한 연구," 한국 음향학회 논문지, 16(4), 1997
- [5] 이용주, 김봉완 외, "공통 음성DB를 위한 PBS의 설계," 1997년도 한국음향학회 학술발표대회 논문집, 1997
- [6] Yong-ju Lee, et al, "Design and Construction of Korean Speech Database for Common Use," Proc. of ICSP '97, 1997
- [7] 정유현, 최준혁 외, "공통 음성 데이터베이스 구축을 위한 사전 조사연구," 전자공학회 하계 학술대회, 1992
- [8] 이용주, 김종진 외, "한국어 음성 DB 구축을 위한 한국어 레이블링 기준에 관한 연구," 제 13회 음성통신 및 신호처리 워크샵 논문집, 1996
- [9] 이용주, "음성언어코퍼스," 정보과학회지 제16권 제2호, 1998. 2