음성 언어 자료 확보를 위한 Workbench의 설계 및 구현

김 태환, 박 순철, 김 봉완, 이 용주
원광 대학교 컴퓨터공학과

Design and implementation of workbench for spoken language data acquisition

Tae-hwan Kim, Soon-chul Park, Bong-wan Kim, Yong-ju Lee
Computer Engineering, Wonkwang Univ.
yjlee@wonnms.wonkwang.ac.kr

요 약

음성 언어 자료의 확보 및 활용을 위해서는 다양한 소프트웨어의 도움이 필요하다. 본 논문에서는 본 연구실에서 설계 및 개발한 PC용 Workbench에 대하여 기술한다.

Workbench는 음성 언어 자료의 확보를 위한 텍스트 처리 모듈들과 음성 데이터의 처리를 위한 신호처리 모듈들로 구성되어 있다. Workbench에 포함된 모듈로는 텍스트를 자동 읽기 변환하는 철자 음운 변환기, 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.

1. 서 론

음성 언어의 인식 및 합성 등 우리말 음성 정보 처리 시스템의 개발을 위해 가장 먼저 확보해야 할 것이 다양한 사람이 발성한 대량의 음성 언어 자료이다. 이 음성 언어 자료는 인식 연구에서는 알고리즘의 훈련 및 평가용, 합성에서는 합성 단위 제작을 위한 기본 자료이며 음운 및 운율 규칙의 생성을 위한 기본적인 분석 자료의 대상이 된다.

이러한 음성 언어 자료의 확보를 위해서는 발성 목록의 설계에서부터 음성 자료의 수집 및 편집, 레이블링, 검색 환경의 제공 등 다양한 소프트웨어의 도움이 필요하다.

본 논문에서는 본 연구실에서 음성 언어 자료의 확보를 위한 Workbench를 설계하고, 구현한 결과에 대하여 기술한다. 구현된 모듈로는 텍스트를 자동 읽기 변환하는 철자 음운 변환기, 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 검색기가 있다. 2.1에서 Workbench의 개요, 2.2 ~ 2.7에서는 개발된 각각의 모듈에 대하여 기술한다.

2. 음성 언어 자료 확보를 위한 Workbench

2.1 Workbench 개요

음성 언어 자료의 확보를 위한 Workbench는 발성 목록의 설계 및 음성 수집 모듈, 음성 편집 모듈, 레이블링 시스템, 음운 환경 검색 모듈로 구성되어 있으며 각 모듈들간의 관계는 다음 그림과 같다.

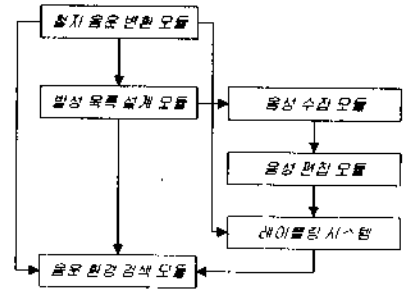

그림1. Workbench 개요

2.2 철자 음운 변환 모듈

철자 음운 변환기는 한글을 소리나는 대로 바꿔주는 읽기 규칙을 구현한 것이다. 읽기 규칙의 기본 원칙은 정부에서 고시한 표준어 규정의 표준 발음법을 따르고 있다[1]. 읽기 규칙은 크게 4가지로 구별되며, 모음의 발음, 장음처리, 음운 변동 규칙, 음의 첨가로 나눌 수 있다. 구현된 모듈은 표준어 규정에 있는 원칙 중 장음 처리를 제외한 3가지의 규칙을 적용하고, 두 가지 이상으로 적용 가능한 경우에는 보다 일반적인 경우를 적용하도록 하였다. 또한 규칙으로 구현할 수 없는 경우에는 예외 발음 사전에 추가하여 처리하도록 하였다.

2.3 음운 균형 발성 목록 선정 모듈

특정 테스크에 종속적이지 않은 음성 데이터베이스를 구축하기 위해서는 가급적 적은 발성 목록에 한국어의 다양한 음운 현상이 포함되는 것이 바람직하다.

본 모듈은 대량의 텍스트 코퍼스를 입력으로 받아 앞에서 기술한 철자 음운 변환 모듈을 통해 음소열로 변환한 후, 이로부터 음운 현상들이 균형 있게 분포하는(Phonetically balanced) 발성 목록을 선정해 주는 모듈이다.

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.

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

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.

본 연구실에서는 구현된 모듈을 이용하여 PBW 452어절, PBS 589문장을 선별하여 음성 데이터베이스로 구축한 바 있다[2, 3].

2.4 클라이언트/서버 음성 수집 모듈

음성 데이터베이스를 구축하기 위해 음성 시료를 수집할 경우 사용자에게 발성 목록을 어떻게 제시해 줄 것인가 하는 점도 중요한 요소가 된다.

본 모듈은 사용자에게 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 개발된 것이다.

그림2. 클라이언트/서버 음성 수집 모듈

이를 위해 음성 수집 클라이언트에서는 각각의 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

수집 서버는 접속된 수집 클라이언트들에 대하여 다음과 같은 기능을 수행한다.

- 클라이언트 별 음성 수집 통제
- 클라이언트 간 동기 유지
- 발성 목록의 전송
- 잘못 발성된 발성 목록의 재발성
- 발성 화자의 통제(발성 시작, 휴식 등)
- 발성 화자 정보의 관리
- 샘플링 주파수의 설정
- 목록별 발성 시간의 설정

따라서 발성 화자는 복수의 마이크 앞에서 한 번만 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.

다음 그림은 음성 수집 클라이언트와 2개의 클라이언트가 접속되어 있는 상태에서 음성 수집을 통제하고 있는 음성 수집 서버를 나타낸 것이

다.

그림3. 음성 수집 클라이언트

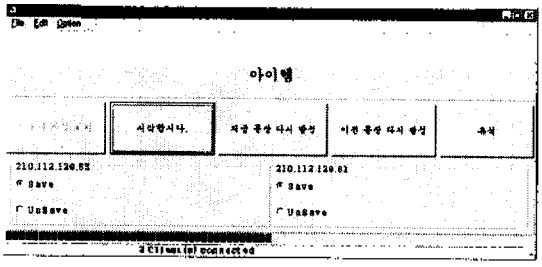

그림4. 음성 수집 서버

2.5 음성 편집 모듈

음성 편집 모듈은 수집된 음성 데이터가 연속적으로 저장되어 있을 경우, 데이터 중 무음 구간, 잘못 발성된 구간을 제외하고 원하는 음성 데이터만을 파일로 저장하기 위한 모듈이다.

연속된 음성 데이터에서 음성 구간을 검출하기 위해서는 끝점 검출 알고리즘이 필요하며, 특히 음성 인식기의 경우 끝점 검출이 성확하지 않을 경우 성능이 저하된다는 사실은 잘 알려져 있다[4].

일반적으로 고립 단어의 끝점 검출을 위해서는 단구간 에너지와 영교차율을 파라메터로 사용한다[5]. 그러나 이러한 끝점 검출 과정은 소음이 존재하지 않는 경우에는 좋은 성능을 나타내지만, 자동차 소음과 같이 SNR이 0dB에 가까운 소음 환경에서는 끝점 검출이 거의 불가능하다. 이러한 경우 주파수 영역의 에너지를 파라메터로 이용하면 보다 좋은 성능을 얻을 수 있다.

따라서 본 모듈에서는 단구간 영교차율, 에너지, 주파수 에너지의 파라메터를 사용자가 선택하여 사용할 수 있도록 구현 하였다.

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

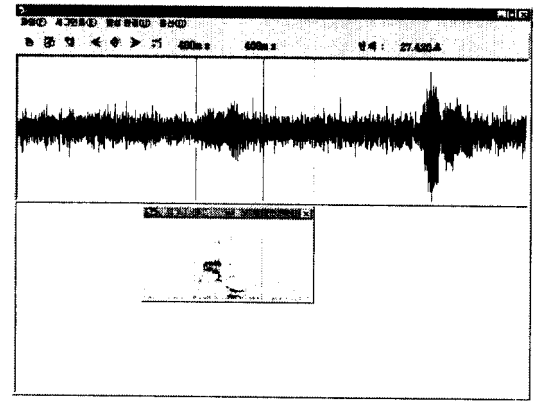

그림5. 음성 편집 모듈

2.6 다단계 음성 레이블링 시스템

레이블링 된 음성 데이터베이스는 음성 인식 및 합성기의 훈련 및 평가, 성능 개선을 위한 중요한 자료로서 활용된다. 그러나 현재 음성 데이터베이스를 레이블링하는 작업은 대부분 수작업으로 이루어지고 있으며, 자동 레이블링을 위한 연구가 활발하게 진행되고 있으나 자동 레이블링 시스템의 훈련을 위해서도 수동 레이블링된 데이터는 필요한 실정이다.

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

본 절에서는 PC에서 사운드카드를 이용하여 손쉽게 레이블링을 할 수 있도록 다단계 음성 레이블링 시스템을 구현한 결과에 대하여 기술한다.

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
- 스펙트로그램 보기
- 스펙트로그램 명암 조절
- FFT 분석 포인트의 조절
- 전체 구간 듣기
- 확대 구간 듣기
- 선택 구간 듣기
- 레이블 구간 듣기
- 다단계 레이블

- 레이블 심볼의 사용자 정의

본 레이블링 시스템에 사용된 스펙트로그램의 특징은 다음과 같다.

- Frame size : 8ms(125Hz bandwidth)
- Frame advance rate : 1ms
- Window : Hanning window
- 기본 FFT 포인트 : 512 포인트

다음 그림은 구현된 레이블링 시스템을 이용하여 레이블링을 수행하는 화면을 나타낸 것이다.

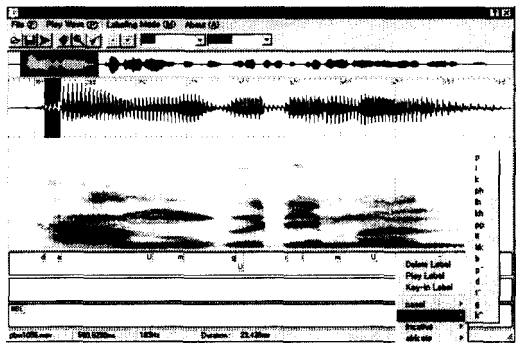

그림6. 다단계 레이블링 시스템

2.7 음운 환경 검색 모듈

특정 음운 현상을 분석하기 위해서는 원하는 조건을 갖춘 발성 목록을 설계하고 음성 시료를 수집하여 이를 분석하는 방법이 사용된다.

그러나 최근 공통으로 사용하기 위한 다종 다양한 환경을 포함하고 있는 대량의 음성 데이터베이스[2, 3, 6]가 배포됨에 따라 이러한 음성 데이터베이스로부터 원하는 음운 환경을 포함하고 있는 발성 목록을 추출하고 이를 분석할 수 있는 방안도 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 발성목록을 선별하고 분석을 수행할 수 있다면 음성 데이터베이스의 활용성을 높일 수 있다.

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

본 모듈에서 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 위에서 기술한 철자 음운 변환 모듈을 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. 음운 환경을 정의하는 방법은 다음과 같다.

-자음 : 자음(<조음방법>,<조음위치>,<강세>)

<조음방법>

조음방법으로는 비음, 파열음, 마찰음, 파찰음, 유음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 조음방법 유형은 어떠한 조음방법도 다 포함할 것(Don't care)을 나타낸다.

<조음위치>

조음위치로는 양순음, 치음, 경구개음, 성문음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 조음위치 유형은 어떠한 조음위치도 다 포함할 것(Don't care)을 나타낸다.

<강세>

강세로는 연음, 격음, 경음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 강세유형은 어떠한 강세도 다 포함할 것(Don't care)을 나타낸다.

<조음방법>, <조음위치>,<강세>에 해당하는 토큰들의 어떠한 조합도 사용 가능하다. 단, 각 토큰들의 위치는 지켜져야 하며, 각 토큰은 쉼표에 의해 구분되어야 한다.

-모음 : 모음(<종류>,<수평위치>,<수직위치>)

<종류>

종류로는 모음, 이중모음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 모음과 이중모음을 구분하지 않겠다는 것(Don't care)을 나타낸다.

<수평위치>

수평위치로는 전설음, 중설음, 후설음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 수평위치 유형은 어떠한 수평위치도 다 포함할 것(Don't care)을 나타낸다.

<수직위치>

수직위치로는 고모음, 중모음, 저모음, 공백을 사용할 수 있다. 만일 공백을 사용한다면 수직위치 유형은 어떠한 수직위치도 다 포함할 것(Don't care)을 나타낸다.

<종류>, <수평위치>,<수직위치>에 해당하는 토큰들의 어떠한 조합도 사용 가능하다. 단 각 토큰들의 위치는 지켜져야 하며, 각 토큰들은 쉼표에 의해 구분되어야 한다.

위와 같이 정의된 음운 환경을 이용하여 본 모듈에서 "고모음 + 파열 연자음 + 고모음"을 검색하는 경우를 다음 그림에 나타내었다.

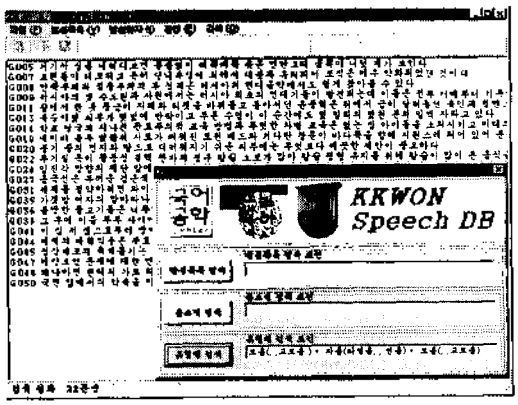

그림7. 음운 환경 검색 모듈

3. 결 론

음성 언어의 인식 및 합성 등 우리말 음성 정보 처리 시스템의 개발을 위해 가장 먼저 확보해야 할 것이 다양한 사람이 발성한 대량의 음성 언어 자료이다.

이러한 음성 언어 자료의 확보를 위해서는 디지털 신호처리 기술뿐만이 아니라 자연 언어 처리 기술을 포함한 다양한 소프트웨어의 도움이 필요하다.

본 논문에서는 본 연구실에서 음성 언어 자료의 확보를 위한 Workbench를 설계하고, 구현한 중간 결과에 대하여 소개하였다. 구현된 모듈로는 텍스트를 자동 읽기 변환하는 철자 음운 변환기, 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. 향후 다양한 분석 기능도 포함하여 종합적인 음성 언어 처리 Workbench로 발전시켜 나갈 계획이다.

참 고 문 헌

[1] 최운천, 지민제, 이용주, "문장음성 변환 시스템 글소리II를 위한 읽기 규칙," 1992년도 제4회 한글 및 한국어 정보처리 학술발표 논문집, '92. 10

[2] 김봉완, 이용주 외, "공동 이용을 위한 음성 DB의 설계 및 구축에 관한 연구," 한국음향학회논문지, 16(4) : 35 ~ 41, '97. 5

[3] 김봉완, 이용주 외, "공통 음성DB를 위한 PBS의 설계,", 1997년도 한국음향학회 학술발표대회 논문집, '97. 7

[4] J.G. Wilpon, et al, "An improved word-detection algorithm for telephone-quality speech incorporating both syntactic and semantic constraints," AT&T Tech. J., 63(3) : 479 ~ 493, '84. 3

[5] L.R. Rabiner, et al, "An algorithm for determining the endpoint of isolated utterance," Bell syst. Tech. J., 54(2) : 297 ~ 315, '75. 2

[6] 이용주, "음성 데이터베이스의 현황과 과제," 한국음향학회 제13회 음성 통신 및 신호처리 워크샵 논문집, '96. 8

[7] 이용주, "음성언어코퍼스," 정보과학회지, 16(2) : 41 ~ 48, '98. 2