음소 및 성조 레이블링 프로그램 개발

이윤경*, 곽철, 권오욱
충북대학교 제어계측공학과

Development of a Phoneme and Tone Labeling Program

Yun-Kyung Lee*, Chul Kwak, Oh-Wook Kwon
Department of Control and Instrumentation, Chungbuk National University

Abstract - Although previous speech analysis programs usually provide speech analysis and phoneme labeling functionalities, they require much time in manual labeling and support only English alphabets. To solve these problems, we develop a new Windows-based program with an improved phoneme and tone labeling method as well as the conventional speech analysis functionalities. The developed program has the unique feature in semi-automatic phoneme and tone labeling based on hidden Markov models.

1. 서 론

한국어 음성의 성조 연구에 있어서 필수적인 첫 단계는 음성신호로부터 말소리의 스펙트로그램, 높낮이(피치), 세기(에너지), 억양(intonation), 포만트 주파수(formant frequency)와 같은 음향학적 특징[1][2][3]을 구하고, 이를 음성 파형과 함께 관찰하는 것이다.

음향학적 특징을 구하기 위하여 컴퓨터를 활용하는 것이 일반적인 추세로서, Praat [4]와 WaveSurfer [5]가 음성학 연구자들에게 널리 사용되고 있으며, 국내에서는 kWaves [6]가 공개되어 있다. 이들 소프트웨어는 높낮이, 세기, 억양, 포만트 주파수를 표시하고 저장할 수 있는 기능을 가지고 있다. Praat와 WaveSurfer는 음성신호를 음소 단위로 자동 분할하는 음소 레이블링 기능이 취약하여 연구자가 일일이 음소 경계를 지정해 주어야 하기 때문에, 연구자들은 대량의 문장 분석을 통한 음성학 및 언어학 연구에서 많은 시간을 레이블링 작업에 뺏기게 된다. 또한 이들 소프트웨어는 영어를 대상으로 개발되었기 때문에 한국어 단어나 음소의 입력이 원활하지 않으며, 성조 및 억양 분석을 위한 기능이 부족한 형편이다. kWaves는 반자동 음소 레이블링과 한글 입력 기능을 지원하고 있으나, 로그파일이 지원되지 않고 성조 분석을 위한 기능이 미비하여 사용자 인터페이스가 불편하다는 문제점을 지니고 있다.

본 논문에서는 한국어의 성조 및 억양 분석에 편리하도록 기존 소프트웨어의 문제점을 해결하고, 반자동 음소 레이블링과 한글 입력 및 성조 연구를 위한 통계적 특성 추출 기능을 보완한 새로운 음성분석 프로그램을 개발하였다.

2. 프로그램 개요

2.1 기능

개발된 소프트웨어는 음성학 연구자들을 위한 음성신호 입출력, 음성분석, 레이블링, 포만트 차트 생성 기능을 갖는 윈도우에서 실행되는 소프트웨어이다.

사운드카드로부터 입력되는 음성신호를 녹음하고 재생하며, 음성파일을 읽고 저장할 수 있다. 파형을 그리고 음성파형에 대하여 자르기/복사하기/삭제하기/실행취소를 할 수 있다. 또한 스펙트로그램을 계산하여 표시하고, 기본주파수(F0)를 추정하거나 혹은 다른 소프트웨어에서 구한 피치 정보를 읽어 표시할 수 있다. 추정되거나 읽어서 표시한 F0를 Q-tone level로 변환하여 표시할 수 있다. 에너지, 포먼트 주파수를 계산하여 표시할 수 있다. 음소에 대한 포만트 주파수 데이터가 주어지면, F1과 F2의 관계를 나타내는 포만트 차트를 그릴 수 있고, 사용자가 음성 파형의 일부를 선택하여 특정 키를 누를 경우 선택 구간의 음향특징 정보를 파일에 저장할 수 있다. 만약 계산된 F0 값에 오류가 있다면 사용자가 정정할 수 있다.

음성학자들의 레이블링 작업을 돕기 위하여 반자동 음소 레이블링 기능을 제공하며, 문장에 대한 텍스트 정보와 음성 신호가 주어지면 음성 인식기를 이용하여 문장/단어/음절/음소/성조 단위로 레이블링 할 수 있다. 레이블의 삽입, 치환, 이동을 지원하며 한글 레이블을 입력할 수 있도록 하였다.

2.2 차별성

개발된 프로그램은 기존 소프트웨어와는 달리 주어진 텍스트 정보로부터 음소 경계를 자동으로 찾아서 레이블링할 수 있는 반자동 음소 레이블링 기능이 있으며 레이블에 한글을 사용할 수 있다. 레이블 삽입, 삭제, 취소를 위한 사용자 인터페이스의 편의성을 향상하여 레이블링 윈도우를 사용하기 쉽게 하였다. 어절 단위로 어절 마지막 음절과 직전 음절 모음의 피치 값을 비교하여 Ha 또는 La 등으로 레이블링하였으며, 어절 구간을 검출하여 강세 어구를 L%, H%, LH%, HL% 등으로 레이블링하였다.

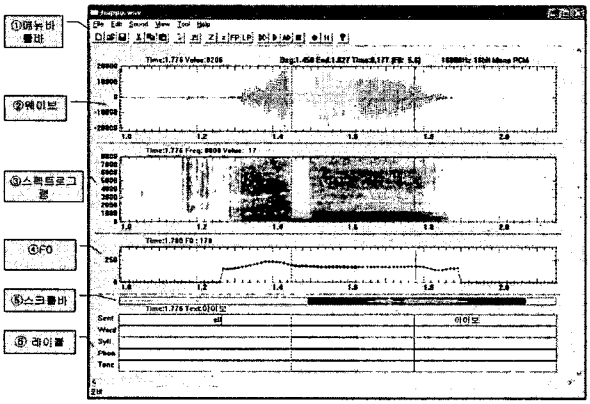
피치 주파수를 한 옥타브를 24단계로 하여 Q-tone level 변환을 하는 기능이 있으며 포먼트 차트 그리기 기능으로 포먼트 주파수(F1, F2)에 대하여 Bark 척도의 2차원 그래프와 포만트 주파수의 샘플 분포를 그리고, 평균점과 분산 타원을 표시한다.

<표 1> 개발 프로그램과 기존 소프트웨어와의 기능 비교

기능	Praat	WaveSurfer	개발 프로그램
음성입출력	o	o	o
파형, 스펙트로그램 표시	o	o	o
피치, 에너지, 포먼트	o	o	o
레이블 정보 입력	o	o	o
로그 파일 지원	o	o	o
수동 레이블링	o	o	o
반자동 레이블링	x	x	o
한글 지원	x	x	o
성조 분석 기능	x	x	o
프로그래밍 지원	o	x	x
학습 알고리듬 지원	o	x	x
통계 분석	o	x	x

3. 음성 분석

음성을 분석하는 데 필요한 파라미터들로서 파형과 스펙트로그램, 에너지, 기본주파수(F0), Q-Tone Level변환, 에너지, 포먼트 주파수를 구하였다. <그림 1>은 개발된 소프트웨어의 실행화면으로 파형, 스펙트로그램, F0, 레이블의 출력을 확인할 수 있다. 스크롤바는 전체 음성파형의 모양이 축소되어 나타나며, 현재 화면에 표시되는 부분이 파랗게 표시된다.



<그림 1> 음성분석 실행화면 예제

3.1 스펙트로그램

주파수 영역에서의 음성신호 분석을 위하여 스펙트로그램을 계산한다. 먼저 음성신호를 프리엠퍼시스 하고, 윈도우를 적용하여 프레임 단위로 분할한다. 한 프레임의 파형을 FFT를 취하여 크기를 구하고, 그 결과를 dB척도로 변환한다. 광대역 스펙트로그램을 그리기 위하여 프레임 크기 5 ms, 프레임 이동 1 ms, 프리엠퍼시스(계수 0.95) 및 해밍 윈도우를 적용하였다.

3.2 F0 및 Q-Tone Level

F0는 음의 높이를 나타내는 파라미터로 AMDF(average magnitude difference function)방법을 사용하여 추출하였다. AMDF는 연산을 절대값과 차분으로 계산하기 때문에 자기상관함수에서 사용하는 계산방법보다 빠르다는 장점이 있다. AMDF를 사용하여 추출된 피치 후보는 프레임 간에 피치가 급격히 변하는 것을 방지하기 위해 현재 프레임의 이전 3프레임부터 이후 3프레임에서 중간 값으로 하는 스무딩(smoothing)을 적용하여 피치를 구한다. 그리고 짧은 구간의 무성음 프레임 구간(1~2프레임)이 유성음 사이에 위치하여 있으면 이전과 이후 프레임의 평균 피치 값을 갖는 유성음으

로 처리한다.

Q-tone level값은 F0 값을 이산적인 등급으로 변환시킨 것으로서, 110Hz를 0으로 정하고 등급간에 주파수 비율이 $^{24}\sqrt{2}$가 되도록 정한 것이다. 무성음과 같이 피치 주파수가 0인 경우의 등급은 편의상 -100으로 정하였다.

3.3 포만트 주파수

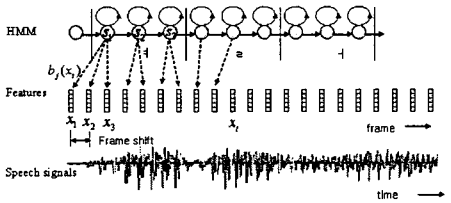
포만트는 음성을 켑스트럼(cepstrum)으로 변환하여 피크를 구해서 추출하는 방법과 LPC(linear predictive coding)으로부터 구하는 방법 등이 있는데 여기에서는 LPC를 이용하였다. 성도가 선형 시스템으로 가정하여 구한 all0pole 필터 H(z)의 극점을 구함으로써 포만트 주파수를 알 수 있다. 계산된 포만트 주파수는 이전 프레임에서의 주파수 값을 참조하여 F1, F2, F3, F4로 분류되고, 같은 포만트 위치로 분류된 값들은 선으로 연결되어 디스플레이 된다.

4. 자동 레이블링

4.1 Hidden Markov Model (HMM)

자동 음소 레이블링[7][8]은 음성학 연구자들에게 초기 음소 레이블링 결과를 제공함으로써 전체 작업 시간을 줄일 수 있게 한다. 최근 음성인식에서 사용되는 HMM(hidden Markov model)을 이용하여 음성신호를 음소 단위로 자동으로 분할하고 있다.

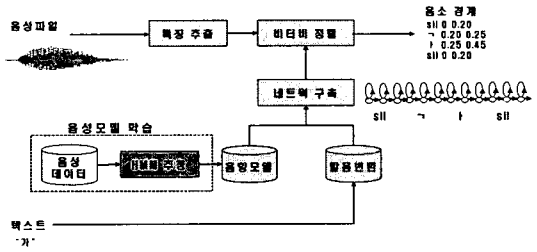
HMM은 시간에 따라서 변화하는 신호를 모델링 하는 방법으로서, 숨겨진 상태천이 확률 과정과 각 상태에서 관측신호를 발생하는 관측 확률과정의 두 개의 랜덤 프로세스로 이루어진다. HMM의 상태천이 확률 및 관측 확률은 시간에 독립이며 현재 상태에만 의존한다고 가정한다. HMM에 의한 음성모델링에서의 음성신호는 아래 그림과 같이 left-to-right 천이만을 갖는 HMM으로 모델링 되며, 각 상태에서는 특징 벡터가 출력된다. 아래 예는 "여"라는 음소는 3개의 상태를 갖는 HMM으로 모델링 되었으며, 각 상태에서 3, 2, 2프레임의 특징벡터를 출력함을 나타낸다.



<그림 2> HMM에 의한 음성신호 모델링

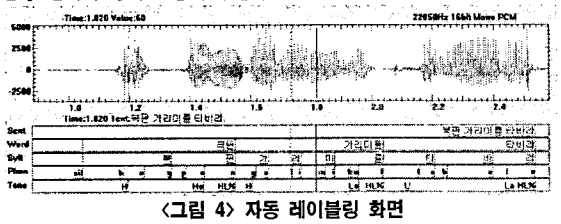
4.2 HMM을 이용한 음소 레이블링

음성인식기가 일반적으로 제공하는 비터비 정렬 기능을 이용하여, 주어진 텍스트로부터 음소열을 발생하고 이에 대한 음향모델을 이용하여 음성신호와 정렬함으로써 음소 단위의 경계를 얻을 수 있다. 이와 같이 HMM을 이용한 음성인식기의 정렬 기능을 이용하여 대략의 음소 경계를 자동으로 구한 다음, 찾아진 음소 경계를 수작업으로 수정할 수 있도록 한다. 특징추출 모듈은 음소 정렬을 위한 특징을 뽑아내며 일반적으로 MFCC(mel scale cepstrum coefficient)가 사용된다.



<그림 3> HMM 기반 음소 레이블링

음소 레이블링의 과정은 다음과 같다. 먼저 훈련데이터로부터 바움-웰치 알고리듬을 이용하여 음소단위의 HMM을 학습한다. 인식단계에서는 주어진 음성신호에 대한 문장 텍스트를 발음열로 변환하고 이를 음소열로 바꾼 다음, 문장에 대한 HMM 네트웍으로 구성한다. 문장모델이 구성되면 전향 알고리듬을 이용하여 입력된 특징벡터열에 대한 관측확률을 계산하여 가장 확률이 높은 경로를 찾아서 상태가 변화하는 시간을 찾으면 그것이 음소 경계가 된다. 음소 단위 레이블링 결과를 문장 텍스트와 정렬하여 음절, 어절, 문장 단위의 경계도 함께 구한다.



<그림 4> 자동 레이블링 화면

4.3 성조 레이블링

성조 레이블링은 음성분석에서 구한 피치 정보를 이용하여 구해진다. 성조 레이블은 액센트구(AP)와 억양구(IP)의 두 계층으로 나누어진다.

AP는 편의상 어절내의 음절 피치를 고려하여 어절의 시작(initial)과 끝(final)음절의 중간 위치에 표기되며, IP는 편의상 피치의 기울기를 고려하여 어절 끝에 표기된다. 표기 방법은 다음 표와 같이 참고문헌[9]를 따른다.

<표 2> 성조 레이블링 방법

종류	기호	피치 모양	위치
AP initial tones	L	둘째 음절의 피치가 첫 음절의 피치보다 10Hz 이상 높을 때	어절의 첫음절 중간
	H	첫 음절의 피치가 둘째 음절의 피치보다 10Hz 이상 높을 때	어절의 첫음절 중간
	+H	둘째 음절의 피치가 셋째 음절의 피치보다 10Hz 이상 높을 때	어절의 둘째 음절 중간
AP final tones	Ha	마지막 음절의 피치가 마지막에서 두 번째 음절의 피치보다 10Hz 이상 높을 때	어절의 마지막 음절 중간
	La	마지막에서 두 번째 음절의 피치가 마지막에서 마지막 음절의 피치보다 10Hz 이상 높을 때	어절의 마지막 음절 중간
	L+	마지막에서 세 번째 음절의 피치가 마지막에서 두 번째 음절의 피치보다 10Hz 이상 높을 때	어절의 마지막에서 두 번째 음절 중간
IP final boundary tones	L%	마지막 두 음절의 기울기가 각각 -0.1 Hz/ms 이하	어절 끝점
	H%	마지막 두 음절의 기울기가 각각 +0.1 Hz/ms 이상	어절 끝점
	LH%	마지막 음절의 기울기는 +0.1 Hz/ms 이상, 마지막에서 두 번째 음절의 기울기가 -0.1과 +0.1 이내	어절 끝점
	HL%	마지막 음절의 기울기는 -0.1 Hz/ms 이상, 마지막에서 두 번째 음절의 기울기가 -0.1과 +0.1 이내	어절 끝점

5. 결 론

음성분석과 음소 및 성조의 레이블링에 이용될 수 있는 음성분석 소프트웨어의 기능에 대하여 개략적으로 논하였다. 음성학 연구자들에게 널리 사용되고 있는 음성분석 소프트웨어가 공개되어 있긴 하지만 영어를 대상으로 개발되었기 때문에 한국어 단어나 음소의 입력, 성조 및 억양 분석에 불편을 감수하여야 하였다.

본 논문에서 제안한 음성 분석 소프트웨어는 음성신호 입출력, 음성분석, 레이블링, 포만트차트 생성 기능을 갖는 소프트웨어로 음성신호를 녹음하고 재생하며, 음성파일을 읽고 저장할 수 있다. 입력된 음성파일에 대하여 파형, 스펙트로그램, 기본주파수(F0), 에너지, 포먼트 주파수를 표시할 수 있다. 또한 한글 입력과 반자동 음소 레이블링 기능으로 한국어 문장을 입력하여 문장/단어/음절/음소/성조 단위의 다단계 레이블링의 결과를 얻을 수 있으며, 성조 및 억양 분석을 위한 기능이 보완되었다.

참 고 문 헌

[1] L.R. Rabiner, R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.

[2] L.R. Rabiner, B.-H. Juang, Fundamentals of Speech Recognition, Prentice-Hall, 1993.

[3] X. Huang, A. Acero, H.-W. Hon, Spoken Language Processing, Prentice-Hall, 2001.

[4] http://www.fon.hum.uva.nl/praat/download_win.html

[5] http://www.speech.kth.se/wavesurfer/

[6] http://speech.chungbuk.ac.kr/~owkwon/srhome/index.html

[7] H. Kawai and T. Toda, "An evaluation of automatic phone segmentation for concatenative speech synthesis," Proc. ICASSP, 2004.

[8] J. Adell, A. Bonafonte, J.A. Gomez, M.J. Castro, "Comparative study of automatic phone segmentation methods for TTS," Proc. ICASSP, 2005.

[9] http://www.linguistics.ucla.edu/people/jun/ktobi/K-tobi.html