

운율구 단위의 음성인식을 위한 운율구 개수 추정

강 지 영 · 김 진 영
전남대학교 전자공학과

Prosodic Phrase Boundary Estimation for Continuous Speech Recognition

Ji-Young Kang · Jin-Young Kim

Dept. of Electronics Engineering, Chonnam National University
(jykang@dsp.chonnam.ac.kr, kimjin@dsp.chonnam.ac.kr)

요 약

본 연구에서는 한국어 음성인식기의 향상을 위한 방법으로서 운율구 단위의 음성인식을 제안하고 운율구 경계를 예측하는 방법을 제시하였다. 실험을 위해서 서울 말씨를 쓰는 남자가 보통속도로 읽은 100개의 문장과 학교 방송국 여자 아나운서가 읽은 100개의 문장에 대해서 운율구 청취 테스트한 데이터를 기준으로 사용했다. 피치 정보와 휴지기 경계 정보를 이용해서 강한 운율경계강도가 나타나는 지점을 운율구의 경계로 예측했을 때 평균 70% 정도의 예측률을 보여주었다.

1. 서 론

음성 인식에 있어서 연속음 인식을 하는데 있어서 하나의 문제점은 발화된 문장의 길이가 길어질수록 언어학적으로 찾아야 할 문장의 수가 기하 급수적으로 늘어난다는 사실이다. 그런데, 만약, 문장보다 작은 단위로 발화된 문장을 잘라낼 수 있다면, 연속음 인식의 복잡도는 줄어들 것이다. 발화된 문장은 연속된 어절들의 연결로서 이루어지는데, 어절들간에는 운율적 경계가 존재한다. 물론, 느슨하게 묶일 수도 있으며, 불연속성이 아주 클 수도 있다. 연속음 인식을 할 때, 가장 좋은 방법은 음소간의 경계를 확실하게 결정할 수 있어 음소

단위 인식을 하는 것이다. 그러나, 음소단위의 경계를 정하는 것이 쉽지가 않으며, 더 나아가 음절 단위, 어절 단위로 끊는 것도 어려운 작업이다. 왜냐하면, 운율경계강도가 어절간에 크지 않다면, 어절의 경계를 찾는 것은, 음성인식의 도움 없이 거의 불가능하기 때문이다. 따라서, 본 연구에서는 운율구(prosodic phrase)단위의 음성인식의 기초연구로서 운율구 경계를 찾는 방법을 제안한다. 왜냐하면, 운율구 단위는 주로 운율경계강도가 커서 운율정보에 의하여 결정하는 것이 가능하기 때문이다. 이 때 음소의 길이정보를 고려하기는 쉽지 않은데, 이는 음소의 길이정보는 일반적으로 음성인식이 수행되어야 알 수가 있기 때문이다. 따라서 운율구 단위의 음성인식을 위한 방안으로서 다음과 같은 방안을 제시한다.

- 첫째, 피치정보와 휴지기 정보를 사용하여 운율구의 개수와 경계를 대략 정한다.
- 둘째, 시간에 따른 에너지 정보와 스펙트럼 변화를 계산하여 이 정보를 바탕으로 운율구 경계를 교정한다.

위의 방안에서 피치정보와 휴지기 정보를 사용하는 이유는 이들이 운율의 불연속 점들을 잘 반영하기 때문이다. 본 연구에서는 운율구 경계를 찾기 위한 기본 아이디어로서 두 가지의 방법을 제안한다. 첫째는 피치의 기저선 개선 이론에 기반한 방법이며, 두 번째 방법

제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

은 구문론적인 휴지기의 유무와 피치악센트의 존재유무를 사용하는 방법이다. 특히, 운율경계강도가 클 경우 피치는 대체적으로 falling tone을 보이는 경향이 있는데 이러한 사실을 기반으로 하여 운율경계를 찾자는 것이다.

우리가 살펴본 운율구 단위의 음성인식의 전처리구조는 다음과 같은 그림 1과 같이 표현할 수 있다.

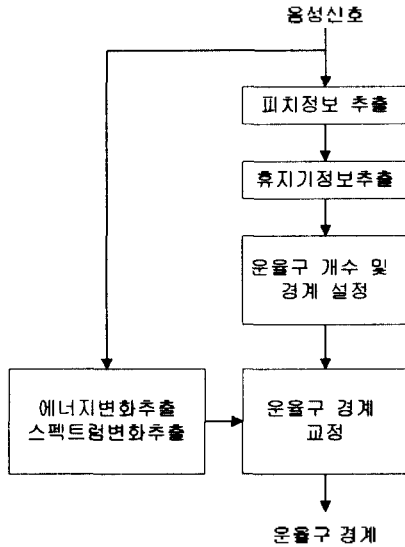


그림 1. 운율구 단위의 음성인식을 위한 전처리 구조

한편, 위의 그림에서, 휴지기 정보와 피치정보에 의한 운율구 경계예측은 상당한 경계지점의 오차가 발생할 수 있는데, 이는 에너지 윤곽선과 스펙트럼 변화도를 추출하여 교정된다. 그러나, 운율경계의 교정에 대해서는 다음의 연구로 미루도록 한다.

2. 피치윤곽선 계곡기반의 운율구 예측

그러면, 피치정보를 이용하여 운율구 경계를 정하는 문제를 생각하여 보자. 앞에서 간단히 설명했던 바와 같이, 강한 운율경계는 RD¹⁾(Right Dependent boundary) 값이나 To.Right²⁾의 값이 큰 경우에 발생한다.

그런데, Fujisaki의 억양발생 모델에 의하면, RD값이 큰 경우 구명령(phrase명령)이 발생하여 피치의 기저선(base line)이 상승하게 된다. 따라서, 기저선의 최소값은 강한 운율경계에서 최소값을 보이게 된다. 즉, 기저선의 재설정(재설정)이 일어나게 된다.

1) RD(Right Dependent boundary) : 수형도에서 경계의 값이 (j-1) : i, j는 수형도에 왼쪽, 오른쪽가지를 나타냄)
 2) To.Right : 피 수식어와 수식어 사이의 어절의 수

기저선의 재설정(재설정)이 나타나는 모습은 다음 그림과 같다.



그림 2. 기저선 재설정

한편, 우리 나라 말의 경우 피치악센트가 LHLH의 형태로 실리기 때문에 전체의 피치값의 최소값은 거의 어절의 첫음절에 오는 경우가 많다. 물론, 휴지기가 없다는 가정하에서 그러하다. 또한, LHLH모델에 의하면, 피치정보에 국부 최소값(local minimum)이 강한 운율경계의 사이에 발생할 수 있다. 본 연구에서는 이러한 피치현상에 대한 관찰을 바탕으로 하여 계곡기반의 운율경계추정 (Valley Based Prosodic Boundary Estimation, VBPBE)을 제안하였다. 다시 정의하자면 다음과 같다.

(계곡기반 운율경계추정 : VBPBE) 주어진 문장에 대하여 피치 경계(contour)를 구한 다음, 피치의 국부 최소값을 결정한다. 다음 국부 최소값 둘로부터 다시 한번 피치 최소값을 구하고 이 최소값이 발생하는 위치를 운율경계라고 단락적으로 추정한다.



그림 3. 운율경계 예측된 샘플문장

(빈 사각형은 국부 최소값, 검은 사각형은 강한운율경계)

위의 그림은 샘플문장

요즘 ~ 우리 나라 전체의 / 갖가지 시끄러운 / 결국 ~ 어려운 경제를 회생시키자는 / 여러 가지 형태의 ~ 노력의 표출 일 것이다.

에 대한 피치윤곽선과 국부 최소값을 이은 직선을 보여주고 있으며, 검정색 사각형으로 표시한 부분이 강한 운율경계가 나타나는 지점이다. 다음의 표 1은 본 접근방법의 타당성과 문제점을 더욱 검증하기 위하여 임의의 샘플문장 200개에 대하여 실험한 결과를 요약한 것이다. 예측된 데이터는 청취 테스트 결과와 비교하여

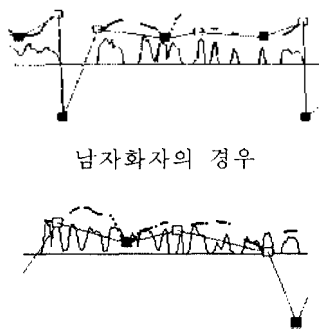
운율구 단위의 음성인식을 위한 운율구 개수 추정

본 결과 강한 운율경계가 있는 지점과, 거의 한 음절 정도의 정확도로 일치하고 있음을 확인할 수 있었다.

표 1. 100문장에 대한 삽입, 삭제 에러

	Insertion	Deletion	운율구 총개수
남자화자	131	68	406
여자화자	41	56	406

위의 표를 살펴보면, 여자화자의 경우에는 개곡기반 운율경계 추정이본이 잘 맞는 반면에 남자화자의 경우에는 많은 삽입에러를 유발함을 알 수 있었다. 남자화자의 경우 말씨의 특성이 각 어절 끝에서 피치가 상승하는 형태이고 여자화자의 경우 어절 끝에서 피치가 떨어지는 모습을 하고 있었다. 어절 끝에서 피치가 상승하는 남자의 경우에는 상승피치 바로 앞에서 피치의 국부 최소의 최소값이 생겨서 잘못된 운율 경계가 발생함을 알 수 있었다. 이 부분은 보상하기 위해서 상승피치의 바로 앞에서 발생하는 운율경계는 상승피치가 끝나는 부분으로 옮김으로써 보상하였다. 하지만 남자화자의 경우 운율구 중에서 나타나는 LHLH형태의 모델에 의해 운율경계가 아닌 곳에서도 국부 최소의 최소값이 발생하는 경우가 많은 것 같다.



남자화자의 경우

여자화자의 경우

그림 4. 두 화자의 피치곡선 비교

또 각 예측 데이터는 운율경계의 시점을 정확하게 찾지는 못한다는 사실을 알 수 있었다. 예를 들어 기음이나 경음에 이어지는 피치에서는 돌발현상이 생겨 에러를 발생시킴을 확인할 수 있었으며 약 20피치 정도(단, 여자음성)의 에러가 발생하기도 했다. 그러나, 그 범위는 구문론적 휴지기의 길이나 한음절 정도의 시간 차이임을 확인할 수 있다. 또한 위의 모델에 덧붙여 설명하자면, 운율경계 강도가 큰 경우는 피치가 특정 주파수 범위에서 발생한다는 추가적인 정보를 이용하여 에러를 줄일 수 있을 것으로 예상된다.

위와 같은 사실로부터 개곡기반의 모델이 타당성이 있다는 결론을 내리고 위의 결과를 구축한 남자 여자 화자의 발화문장 200개에 적용시킨 결과 다음의 표 2와 같은 결과를 얻을 수 있었다. 표를 살펴보면, 평균 탐지율(detection rate)이 여자의 경우 76%로서 상당히 높은 결과를 얻고 있음을 알 수 있었다. 남자화자의 경우는 51%로서 좀더 낮게 나타났는데, 이는 삽입오류(insertion error)의 발생률이 더 많기 때문인데, 삽입오류는 운율구를 과도하게 찾은 것에 해당하는 데, 이는 인식할 때 문장을 더 세분하여 자른 경우에 해당하므로 오류로 생각하기는 힘들다.

표 2. 개곡기반모델의 탐지율과 오류

	탐지율	삽입오류	missing 오류
남자화자	83%	32%	17%
여자화자	86.6%	10%	14%

지금까지의 실험결과를 검토하여 볼 때, 우리는 개곡기반의 방법이 아주 간단하면서도 강한 운율경계강도를 찾는데, 매우 유용하다는 점을 알 수 있다.

3. 휴지길이와 피치악센트 정보를 이용한 운율경계예측

서론에서 간략하게 밝힌 바와 같이 강한 운율적 불연속성은 청취테스트 결과 마치 긴 휴지기가 어절 사이에 존재하는 것처럼 청취자는 판단한다. 그런데, 실제 음성 데이터를 살펴보면, 긴 휴지기가 존재하는 경우가 많지만 그렇지 않은 경우도 종종 발생한다. 그런데, 그러한 경우에는 피치가 하나의 음소에서 LHL의 형태를 보이면서, 음소의 길이가 길어지는 경향을 보인다. 따라서 본 연구에서는 이러한 사실을 바탕으로 하여 운율구 경계를 찾는 방법을 제안하였다. 그 방법은 다음과 같다.

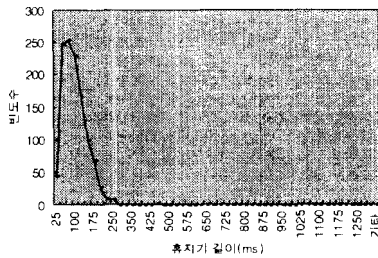
첫째, 주어진 음성 데이터를 분석하여 긴 휴지기가 존재하는 구간을 찾아 이를 일차적으로 운율구의 경계로 정한다.

둘째, 선정된 운율구 경계 내에서 LHL의 톤이 존재하는 지를 체크하여, 만약 이러한 톤이 발생하면, 이를 운율구 경계라 생각하여 자른다.

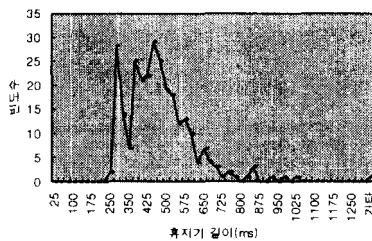
본 연구에서는 위와 같은 방법을 위하여 먼저 휴지기 분석을 하였다. 실험문장 100문장에 대해서 구분론

제15회 음성통신 및 신호처리 워크샵(KSCSP '98 15권1호)

적 휴지기와 음운론적 휴지기를 나타내는 히스토그램을 그려보면 아래와 같다.



a. 음운론적 휴지기



b. 구문론적 휴지기

그림 5. 휴지기 히스토그램

이 그림에서 두 곡선이 만나는 지점을 운율경계를 자르는 기준으로 삼을 수 있다. 긴 휴지기가 나타나는 샘플문장에 대해 분석한 그림의 예는 그림 6과 같다.



그림 6. 피치의 변화가 큰 곳에서 나타나는 강한 운율경계

휴지기 정보 및 피치악센트 정보를 이용한 운율구 개수 예측모델의 신뢰도 검증을 위해서 200개의 문장에 대해서 첫 번째와 동일한 실험을 했을때, 그 결과를 아래 테이블에 표기했다.

표 3. 100문장에 대한 삽입,삭제 예러

	Insertion	Deletion	운율구 총개수
남자화자	118	41	406
여자화자	38	43	406

한편, 위 표에서 삽입에러가 많은 데, 삽입에러가 많다는 것은, 주어진 문장을 강한 운율구 이상으로 세분하여 자를 수 있는 단서를 보여주는 것으로서, 이를 잘 이용하면 오히려 음성인식에 도움이 될 수 있을 것이다.

4. 결 론

본 연구에서는 피치운곽선을 가지고 운율구 경계를 찾는 방법과 휴지기 정보 및 피치 악센트를 사용하여 추정하는 두 가지 방법을 제안하고 200여 개의 발화문장을 대상으로 그 성능을 검토하였다.

아직까지 본 연구에서는 운율구의 경계를 100% 완벽하게 찾는 알고리즘을 개발하지는 못한 상태에 있다. 그러나 위에서 언급했던 바와 같이 발화문장의 포락선 정보와 스펙트럼변화 정보를 사용하여 경계를 교정하는 것이 가능하리라고 생각되며 이는 향후에 연구를 해야 할 과제이다.

참 고 문 헌

- [1] ETRI, "다중 매체 환경에서의 대화체 음성 번역 통신 기술 개발", 정보통신부, 1996.
- [2] 한국통신 연구개발본부, "음성 언어 시스템 개발을 위한 한국어의 운율구조 및 담화구조 연구", 1997.
- [3] 김선미, "한국어의 리듬단위와 문법구조", 서울대학교 박사학위 논문, 1997.
- [4] 전남대학교 공업기술 연구소, "한국어 운율구조 연구", 1996.
- [5] KAIST AI, "음성언어 시스템 개발을 위한 한국어의 운율구조 및 담화구조 연구", 1996.
- [6] A.J.Hunt, "Models of Prosody and Syntax and their Application to Automatic Speech Recognition", 1995.

* 본 논문은 '97대학기초연구지원사업'의 연구결과중 일부입니다.