음성정보와 문법정보를 이용한 한국어 운율 경계의 자동 추정*

김선희(서울대), 전재훈(UTD), 홍혜진(서울대), 정민화(서울대)

<차 례>

<Abstract>

Automatic Detection of Korean Prosodic Boundaries Using Acoustic and Grammatical Information

Sunhee Kim, Je Hun Jeon, Hyejin Hong, Minhwa Chung

This paper presents a method for automatically detecting Korean prosodic boundaries using both acoustic and grammatical information for the performance improvement of speech information processing systems. While most of previous works are solely based on grammatical information, our method utilizes not only grammatical information constructed by a Maximum-Entropy-based grammar model using 10 grammatical features, but also acoustical information constructed by a GMM-based acoustic model using 14 acoustic features. Given that Korean prosodic structure has two intonationally defined prosodic units, intonation phrase (IP) and accentual phrase (AP), experimental results show that the detection rate of AP boundaries is 82.6%, which is higher than the labeler agreement rate in hand transcribing, and that the detection rate of IP boundaries is 88.7%, which is slightly lower than the labeler agreement rate.

* Keywords: Prosodic boundary detection, Acoustic information, Grammatical information.

* 이 논문은 2006년도 정부(교육인적자원부)의 재원으로 제1저자에게 수여된 한국학술진흥재단의 지원을 받아 수행된 연구입니다(KRF-2006-353-A00063).

1. 서 론

운율(prosody)이란 피치(pitch), 길이(duration), 세기(intensity) 등과 같은 자질에 의하여 실현되는데, 이러한 자질은 화자가 발화할 때 강세, 리듬, 억양 등과 같은 운율 현상으로 구조화된다. 이러한 운율 현상은 의미를 분화하는 언어학적인 정보와 화자의 태도, 의도, 혹은 감정 등을 나타내는 비언어학적인(paralinguistic) 정보를 표현하게 된다.

운율은 언어학뿐만 아니라, 음성정보처리 영역에서도 주요 연구 주제로 알려져 있다. 언어학에서의 운율 연구는 음운론 분야에서 [1]을 기점으로 하는 자립분절음운론(autosegmental phonology)으로부터 최적성이론(optimality theory)[2]에 이르기까지 많은 연구가 수행되었는데, 특히, 운율 구조와 통사 구조와의 관계에 대한 대표적인 연구로는 [3]-[5]가 있다. 이 연구들은 운율과 통사 구조와의 관계를 기반으로 하는 운율 모델링 방법에 이용되었다.

음성정보처리에 있어서 운율은 음성합성의 자연성을 결정하는 가장 중요한 요소로 지적되어 왔고[6][7], 음성인식의 성능 향상에 기여하며[8][9], 단어의 중의성이나 문장의 종류를 구분할 수 있는 요소로서 음성언어이해(spoken language understanding)에 있어서도 그 역할의 중요성이 강조되었다 [8][10]-[12]. 운율 모델로는 자립분절음운론에 기반을 둔 tone and break indices (ToBI)[13]를 비롯하여 Hirst 모델[14][15], rise/fall/connection (RFC) 모델[16], 경사도(tilt) 모델[17], IPO 모델[18], Fujisaki 모델[19] 등이 있는데, 이 가운데 ToBI는 음성합성과 음성인식의 성능 향상을 위하여 구축된 대용량 음성 데이터의 전사를 위하여 음성학자들과 음성공학자들이 합의하여 공통적인 규약으로 집대성한 결과물로서, 현재 가장 일반적인 운율 모델로 알려져 있다.

그러나 실제로 음성정보처리 시스템에서 운율은 아직까지도 거의 이용되지 않거나 이용되는 경우에 있어서도 제한적이라고 평가되는데, 이는 운율정보가 어떻게 음성정보처리 시스템에 이용될 수 있는지에 대한 구체적인 방법이 설정되지 않았기 때문이라고 한다 [12]. 현재까지 운율에 관한 대부분의 연구는 ToBI에서 제시하고 있는 운율정보를 대용량 음성 데이터로부터 자동으로 추출하는 연구가 주로 이루어져 왔다. 그러나 이를 실제 시스템에 이용한 예는 위에서 언급한 경우들을 포함하여 그리 많지 않으며, 실제로 이용된 경우에 있어서도 제한적인 이용에 의해 그 효과가 두드러지지 않은 편이다. 실제 사람들 간의 언어생활에 있어서 운율 요소가 많은 언어적·비언어적 기능을 갖는다는 점을 감안할 때, 음성언어처리 분야에서 운율에 관한 연구는 앞으로 많은 연구가 필요한 분야라고 할 수 있다.

언어에 따라서 영어와 같은 강세 언어의 경우는 단어의 강세 예측에 대한 연구와 운율 경계에 대한 연구가 개별적으로 혹은 통합적으로 수행되었고, 중국어나

일본어와 같이 철자 상 띄어쓰기가 되어 있지 않은 언어의 경우는 운율 경계 연구가 주로 수행되었다. 한국어의 경우는 Korean ToBI (K-ToBI)[20]를 기반으로 한 운율 경계 추정에 관련된 연구가 대부분인데, 이러한 연구들은 대부분 문법정보를 기반으로 한 추정 방법으로서 실제 음성 데이터의 분석에 의하여 추출된 음향학적 자질들은 거의 이용되지 않고, 주로 음성합성을 위한 운율 모델링을 위하여 텍스트 기반의 문법정보를 이용한 것들이 많았다.

본 연구는 음성언어처리 분야에서 운율 연구가 어떻게 진행되어 왔는지 선행 연구들을 살펴보고, 한국어의 음성정보처리 시스템의 성능 향상을 위하여 음성정보와 문법정보를 기반으로 한 자동 운율 경계 추정 방법을 제안한다. 본 연구는 한국어의 운율 경계 추정에 있어서 어떤 자질들이 유용한지를 고찰하는 기초 연구로서, 음성정보와 문법정보를 이용한 운율 경계 추정의 유용성을 보이는 것을 그 목적으로 한다. 이러한 연구는 궁극적으로 음성정보처리 시스템의 성능향상에 기여할 것으로 기대한다.

2. 선행 연구

2.1. 운율 경계 추정 방법론

운율 경계 자동 추정에 관한 연구는 크게 통사적 구조와 운율 구조가 유사하다고 보는 입장에서 텍스트로부터 얻어낸 문법(언어)정보를 주로 이용하는 연구와 음성정보만을 이용하는 연구, 그리고 문법정보와 음성정보를 모두 이용하는 연구로 나누어 볼 수 있다.

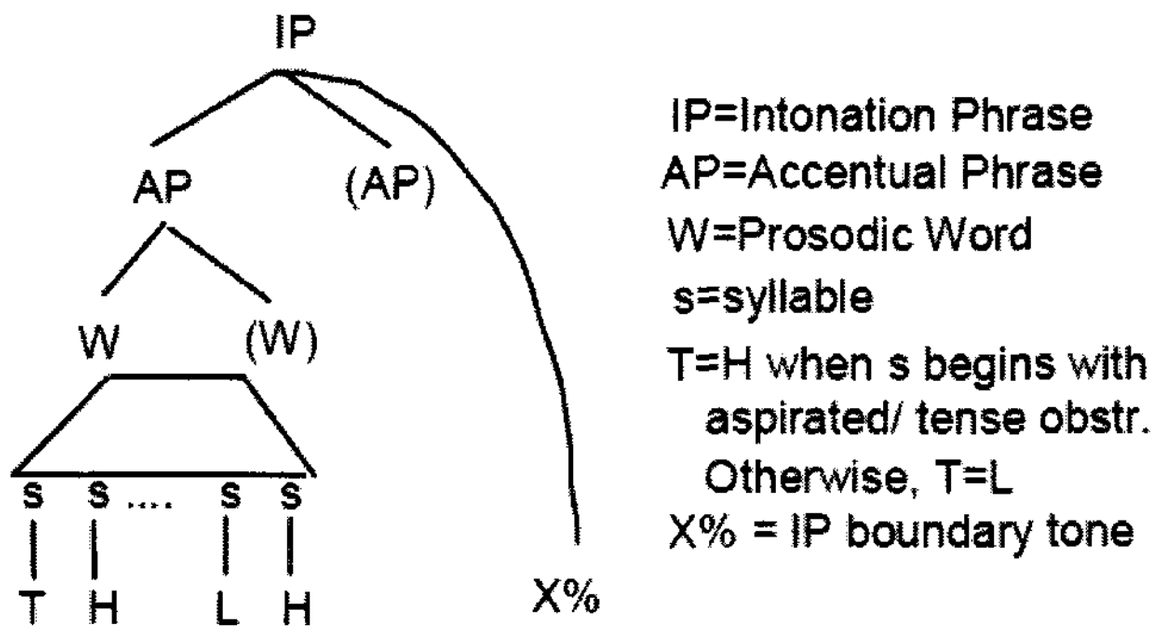
문법정보만 이용하는 경우는 일반적으로 텍스트에서 추출한 품사(part-of-speech: POS) 정보를 이용하여 모델링하는데, 모델링 방법으로는 classification and regression tree (CART)를 이용한 경우가 많고[6][21]-[24], hidden Markov model (HMM)을 이용하는 방법도 제안되었다 [25]. N-gram을 이용하는 경우는 독자적으로 N-gram만 이용하는 경우[26], N-gram과 기타 확률 모델을 결합하여 사용하는 경우[27], 결정트리와 N-gram을 함께 사용하는 경우[28][29]가 있었다. 또한, 문법정보를 이용한 방법인 최대엔트로피(maximum entropy) 모델도 제안되었다 [30].

음성정보만을 이용한 경우는 [31]-[33]이 있는데, [31]은 피치 정보를 모델링하여 한국어의 경계(break) 강도를 예측하였고, [32]는 피치 정보와 함께 휴지기(pause)를 이용하여 한국어의 억양구(intonational phrase: IP)와 강세구(accentual phrase: AP)를 검출하였다. [33]은 피치 정보와 휴지기 및 지속시간을 가우시안 모델(Gaussian model)과 베이지안 의사결정(Bayesian decision) 방법으로 모델링하여 중국어에서 경계의 종류를 예측하였다.

문법정보와 음성정보를 함께 이용한 경우로는 [34]-[37]을 들 수 있다. [34][37]은 음성정보와 문법정보를 모두 HMM으로 모델링한 방법을 소개하였다. [35]는 N-gram으로 문법정보를 모델링하고, 음성정보로는 휴지기의 지속 시간을 모델링하였다. [36]은 문법정보는 인공 신경망(artificial neural network: ANN)을 이용하여 모델링하고, 음성정보는 가우시안 혼합 모델(Gaussian mixture model: GMM)을 이용하여 모델링하였다.

2.2. 한국어의 운율 경계 추정

한국어의 운율에 관한 대표적인 연구로는 K-ToBI[20]를 들 수 있는데, K-ToBI에 의하면 한국어의 운율 체계는 강세구와 억양구로 구성된다. 다음 <그림 1>은 [20]에서 제안한 한국어의 운율 체계이다.



<그림 1> 한국어의 운율 구조

한국어의 운율 단위는 억양구와 강세구로 구성된다. 강세구는 억양구의 하위 단위이고, 단어(word 혹은 prosodic word)의 상위 단위이며 구 성조(phrasal tone)를 갖는다. 억양구는 경계 성조(X%)를 가지며 구의 마지막 음절은 장음으로 실현된다. 4음절 이상의 강세구는 LHLH 구 성조로 실현되고, 3음절 이하인 경우는 오름조(LH, LLH, LHH)로 실현된다. 또한, 강세구의 첫 성조는 첫 분절음의 특성에 따라 결정되는데, 첫 분절음이 평음인 경우는 L로 실현되고, 경음이나 격음인 경우는 H로 실현된다. Break index (BI)는 각 단어 사이의 결합 정도를 나타내는 것으로 4개를 설정하였다. 3의 경우는 억양구 경계, 2는 강세구 경계, 1은 단어 경계, 그리고 0은 단어보다 작은 경계를 표시한다.

따라서 한국어의 운율 경계 추정이란 결국 강세구와 억양구의 경계 추정, 혹은 BI의 종류를 추정하는 일에 해당된다. 다른 외국어와 마찬가지로 한국어의 경우에 있어서의 운율 모델링은 대부분 음성합성 시스템을 위한 연구로서 텍스트를 기반

으로 한 운율 추정 연구가 많이 이루어졌다. 텍스트 기반의 문법정보 기반 운율 경계 추정 연구를 보면, 억양구 경계는 일반적으로 통사적 경계와 일치함으로써 텍스트 분석만으로도 어느 정도 좋은 결과를 얻어낼 수 있으나, 강세구 경계의 경우는 그 추정률이 매우 낮다. 이는 강세구의 경우 통사적인 정보보다는 발화 속도와 같은 다른 요인에 의하여 결정되기 때문이다.

한국어의 경우에도 운율 경계 추정 및 운율 레이블링에 대한 연구가 다양하게 진행되었다. 운율 경계 추정을 위해 해당 단어의 품사, 어절의 길이, 문장에서의 어절의 위치와 같은 언어적 정보를 이용하여 결정 트리를 이용하거나[21][24][38], N-gram을 이용한 경우가 있었고[26][39], conditional random fields (CRFs) 모델을 구성하는 방법이나[40], 문장 성분의 의존 관계를 이용하여 구문 관계에 따른 규칙을 정의하고 문장의 길이, 통사구의 문장 내 위치 등의 정보에 따라 가변적으로 운율 경계를 추정하는 방법이 제안되었다 [41]. 음성정보만을 이용한 경우는 휴지기와 피치 궤적을 이용한 경우[31]와 피치 파라미터를 이용한 경우가 있었으나[32], 문법정보와 음성정보를 모든 이용한 연구는 거의 없었다.

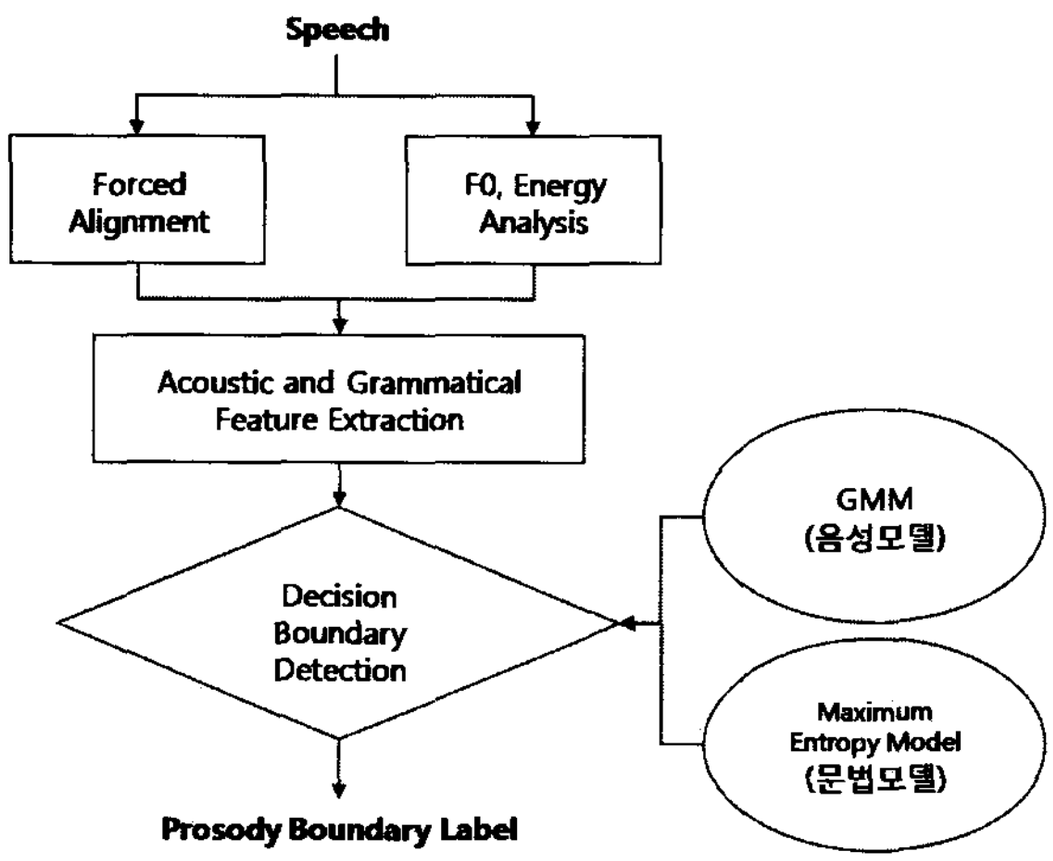
이와 같이 기존의 연구는 대부분 텍스트 기반으로 품사 정보나 구문 관계를 분석하여 제한된 정보만을 이용함으로써, 운율 경계 추정에 피치나 지속시간, 강도 등의 실제 음성 발화에 대한 정보를 적극적으로 이용하지 못하였다. 특히, 운율구는 화자의 발화 속도나 호흡의 길이 등에 따라서 그 실현이 달라지기 때문에 그러한 요소 등을 함께 고려해야 하지만, 기존 연구에서는 발화 속도 등에 따른 운율구 형성에 있어서의 차이를 반영하지 못하였다. 이러한 대부분의 연구에서 억양구 경계의 경우 90% 전후의 높은 성능을 보이나, 강세구 경계의 추정에서는 그리 높지 않은 성능을 보이는 것으로 보고되었는데, 실험에 사용된 데이터들의 특성이 매우 상이하여 절대적으로 그 성능을 비교하기에는 어려움이 있다. 또한, 성능 평가에 있어서 대부분 음성 데이터가 아닌 텍스트 데이터를 사용하였는데 [21][26][38]-[40], 구 성조와 경계 성조, 휴지기 등과 같은 여러 운율 특성에 의하여 결정되는 강세구와 억양구의 운율 경계를 텍스트 데이터만을 이용하여 평가한 것은 문제가 있다고 할 수 있다.

3. 운율 경계 추정 방법

본 논문에서는 한국어 운율 자동 레이블링의 기초 연구로 낭독체 음성 코퍼스(EUROM1)[42]를 사용하여 운율 경계 자동 추정 방법을 제안한다. 운율 경계 추정은 주어진 발화에 대해서 올바른 운율 경계를 찾아내는 것으로, 주어진 단어(형태소) 순서 $W=\{w_1,\cdots,w_n\}$와 음성 특징 $A=\{a_1,\cdots,a_n\}$을 이용하여 최적의 운율 경계 순서 $L=\{L_1,\cdots,L_n\}$를 찾는 것으로 아래와 같이 표현할 수 있다.

$$L = \arg\max_L \; p(L|W,A) = \arg\max_L \; p(L|W)p(A|L,W) \quad (1)$$

위의 식에서 $p(L|W)$는 단어의 문법적(언어적) 정보를 이용한 문법모델이며, $p(A|L,W)$는 음성 특성을 이용한 음성모델이다. 다음 <그림 2>와 같이 문법모델은 최대엔트로피모델을 이용하였고, 음성모델은 GMM을 이용하였다.



<그림 2> 운율 경계 추정 방법

국내외 기존 연구에서 품사 정보나 구문 정보 등의 문법적 정보만을 사용하여 운율 경계 추정을 시도하여 상당히 높은 정확도를 획득하였다 [40][41][43]. 그러나 운율 단위가 통사 단위와 높은 상관관계가 있기는 하지만 두 단위가 정확하게 일치하지는 않는다. 또한, 운율 단위는 화자의 신체적인 조건이나 감정, 발화의 목적이나 종류 등에 따라서 다르게 형성되기 때문에 문법적인 정보만으로는 완벽하게 운율 경계를 추정할 수 없다. 따라서 본 논문에서는 문법모델과 음성모델을 결합하여 운율 경계 추정에 이용하는 방법을 제안하고자 한다.

본 연구에서는 운율 경계를 기준으로 하여 단어를 다음의 세 가지로 분류하였다. 즉, (1) 운율 경계 내부 단어, (2) 강세구 경계 단어, (3) 억양구 경계 단어로 정의하고, 각 단어가 나타나는 경계를 각각 (1) 비경계, (2) 강세구 경계, (3) 억양구 경계라 하여 운율 경계를 추정하도록 하였다.

3.1. 음성모델

음성모델은 음성 특징을 GMM으로 구성하였다. 본 연구에서는 학습 데이터의 부족으로 각 단어에 대한 GMM을 학습하지 않고 아래와 같이 전체 단어에 대해 하나의 GMM을 구성하였다.

$$p(A|L,W) \approx p(A|L) \tag{2}$$

음성정보는 화자 사이 또는 같은 화자라도 발화 지속 시간에 따른 변이가 존재하므로, 피치 약화에 대한 보정 후 정규화(z-score) 하여 사용하였다. 음성모델을 구성하기 위해 다음 <표 1>과 같은 자질 14개를 사용하였다.

<표 1> 음성모델 구성에 사용된 자질

자질	자질 내용
Max_F0	해당 단어 마지막 음절의 F0 최대값
Min_F0	해당 단어 마지막 음절의 F0 최소값
Mean_F0	해당 단어 마지막 음절의 F0 평균값
ΔMax_F0	해당 단어 마지막 음절과 다음 단어 첫 음절의 F0 최대값 변화도
ΔMin_F0	해당 단어 마지막 음절과 다음 단어 첫 음절의 F0 최소값 변화도
ΔMean_F0	해당 단어 마지막 음절과 다음 단어 첫 음절의 F0 평균값 변화도
Gradient	단어 경계부의 F0 기울기
Intensity	해당 단어 마지막 음절의 강도
ΔIntensity	해당 단어 마지막 음절과 다음 단어 첫 음절의 강도 변화도
Final_Syl_Duration	해당 단어 마지막 음절 길이 비율
Speaking_rate	해당 단어와 다음 단어의 발화 속도 비율
Silence	경계부 묵음 지속 시간
Prev_Duration	이전 운율 경계로부터 해당 단어까지의 지속 시간
Prev_Duration_Next	이전 운율 경계로부터 해당 단어 다음 단어까지의 지속 시간

3.2. 문법모델

문법모델은 아래의 식과 같이 문법적 정보를 최대엔트로피모델로 구성하는 방법을 사용하였다.

$$p(L|W) \approx p(L|\Phi(W)) \tag{3}$$

위 식에서 $\Phi(W)$는 해당 단어의 자질을 나타내는 것으로 총 10개의 자질이 이용되었다. 자질의 종류는 다음 <표 2>와 같다.

<표 2> 문법모델 구성에 사용된 자질

자질	자질 내용
POS±2	해당 단어와 전후 각 두 개 단어의 품사 정보
Word_length+1	해당 단어와 다음 단어의 음절 수
Prev_Word_length	이전 운율 경계로부터의 지속 음절 수
Prev_POS_length	이전 운율 경계로부터의 지속 형태소 수
Spacing	해당 단어 전후의 띄어쓰기 정보

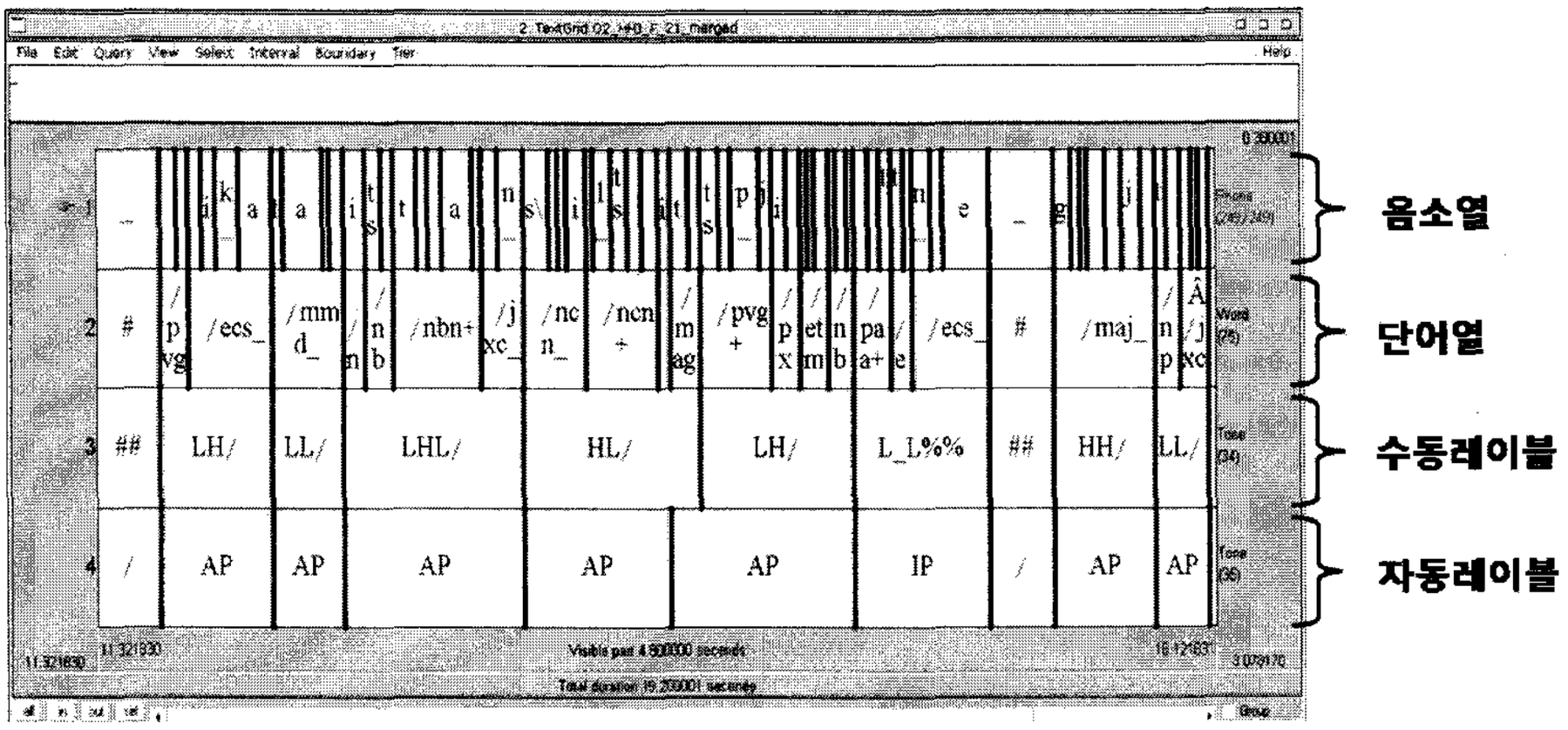
자질의 종류는 해당 단어의 품사 정보와 음절 수, 다음 단어의 음절 수, 이전 운율 경계로부터의 지속 음절 수와 형태소 수, 띄어쓰기 정보를 포함하였다. 품사 정보는 해당 단어의 앞뒤 연속하는 각 두 개의 단어도 동시에 사용하였다.

4. 실험 결과 및 논의

본 논문에서는 유럽의 표준 낭독체 코퍼스인 EUROM1 코퍼스[42]의 기준에 따라 제작된 한국어 낭독체 음성 코퍼스(EUROM1 Korean)를 사용하였다. EUROM1 코퍼스는 유럽의 ESPRIT 과제 가운데 음성 및 음성 시스템의 평가를 위하여 제안된 "음성 평가 방법론(speech assessment methodology: SAM)"의 수행을 위하여 유럽의 11개국이 참여하여 공동으로 개발한 음성 코퍼스로서 단어, 숫자, 문장, 문단 등의 여러 형식으로 구성된다. 한국어 EUROM1 음성 코퍼스는 이 가운데 문단만을 녹음한 것으로 발성 목록은 EUROM 텍스트(화자 당 40문단, 총 166문장)이고, 총 녹음시간은 약 2시간이다. 제보자로는 서울 토박이 20대 남녀 각 5명으로 총 10명이 참여하였다. 녹음은 Marantz PMD 670(녹음기)과 Shure SM-58(마이크)을 이용하여 방음시설이 된 녹음실에서 진행하였다.

본 논문에서 사용된 음성 코퍼스는 [44]에서 사용한 것과 동일한 것으로, [44]에서 제안한 방법을 바탕으로 하여 레이블링되었다. 음소 레이블링은 다국어 합성기인 MBROLA[45]를 이용하여 자동으로 speech assessment methods phonetic alphabet(SAMPA)[46]로 전사되었고, 이러한 음소 레이블링 결과는 Praat[47]의 전사창에서(TextGrid) 해당 음소의 경계와 강세구와 억양구 경계를 정렬시켜 레이블링하였다. 강세구와 억양구 경계 레이블링은 운율 경계 레이블링에 숙련된 3명의 음성학 전문가에 의하여 수행되었다. 본 실험에서는 전체 음성 코퍼스 가운데 1시간 분량인 20문단을 사용하였는데, 학습에는 총 18문단이 사용되었으며 실험에는 학습에 사용되지 않은 2문단을 사용하였다.

다음 <그림 3>은 운율 경계 추정 결과의 예이다.



<그림 3> 운율 경계 추정 결과 예

운율 경계 추정 실험 결과는 아래 <표 3>과 같다.

<표 3> 각 모델별 운율 경계 검출 실험 결과[1)]

		Precision	Recall	F-Measure
음성모델	강세구	67.0	60.2	63.4
	억양구	74.5	78.1	76.3
문법모델	강세구	76.0	75.6	75.8
	억양구	79.7	71.8	75.5
음성모델+문법모델	강세구	85.3	80.2	82.6
	억양구	84.3	93.5	88.7

강세구의 경계 추정에 있어서는 음성모델을 사용한 경우에 F-Measure 63.4%, 문법모델을 사용한 경우는 75.8%로 음성모델을 사용한 경우가 우월한 성능을 보

1) F-Measure는 일반적으로 정보이론에서 효율성을 평가하기 위하여 사용되는 척도로서 다음과 같이 Precision(percent correct of total detected)과 Recall(percent detected of total correct)의 조화평균으로 나타낸다.

$$\mathrm{F-Measure} = \frac{2 \times \mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$$

본 논문에서 Precision과 Recall은 각각 다음과 같이 계산되었다.

$$\mathrm{Precision} = \frac{\text{추정된 운율경계 중 실제 운율 경계}}{\text{추정된 운율 경계}}$$

$$\mathrm{Recall} = \frac{\text{추정된 운율경계 중 실제 운율 경계}}{\text{추정되어야 할 실제 운율 경계}}$$

인다. 반면, 억양구의 경계 추정에 있어서는 음성모델을 사용한 경우는 76.3%, 문법모델을 사용한 경우는 75.5%로서 음성모델을 사용한 경우가 다소 우월한 성능을 보인다. 이는 강세구 경계와는 달리 억양구 경계에서는 마지막 음절이 장음화되고 묵음 구간의 길이 역시 길어지는데, 이러한 음성적 특성을 음성모델이 적절하게 반영하고 있음을 보여준다고 할 수 있다. 운율 경계 추정 성능을 비교해 보면 문법모델만을 사용하는 경우가 강세구와 억양구 각각 75.8%와 75.5%로, 음성모델만을 사용하는 경우인 63.4%와 76.3%보다 좋은 성능을 보인다고 할 수 있겠다. 그러나 음성모델과 문법모델을 결합한 경우가 각 모델을 단독으로 사용한 경우에 비해 강세구의 경우 82.6%, 억양구의 경우 88.7%로 모두 성능이 향상됨을 볼 수 있다.

앞에서 이미 지적한 바와 같이 기존의 연구들이 대부분 음성 데이터가 아닌 텍스트 데이터를 이용하여 그 성능을 평가하였으므로, 기존의 연구 결과와 본 논문의 실험 결과를 비교하는 것은 의미가 없다고 판단된다. 따라서 본 실험으로 얻어진 결과는 K-ToBI 레이블링에서 전사자들 간의 일치도와 비교하는 것이 의미가 있다고 할 수 있다[12]. K-ToBI 레이블링의 전사자 간의 일치도를 비교한 [48]은 강세구 경계에 대한 일치도를 78%, 억양구 경계에 대한 일치도는 91%로 보고하고 있다. <표 3>에서 보는 바와 같이, 본 실험 결과는 음성모델과 문법모델을 결합하여 사용하는 경우에 강세구 경계의 경우는 82.6%로 전사자 간의 일치도보다 높고, 억양구의 경우는 88.7%로서 전사자 간의 일치도에 비하여 다소 낮다. 강세구 경계의 경우 구 성조를 기준으로 전사하는 것이 용이하지 않아 훈련된 전사자 간에도 그 일치도가 높지 않은데 반하여, 본 논문에서는 음성정보와 결합하였을 때 특히 그 성능이 많이 향상된 것을 볼 수 있었다. 억양구 경계의 경우와 관련하여 최근의 연구[41]에서 문법모델만으로도 90% 정도의 추정 성능을 보고하였고, [48]에서는 전사자 간의 일치도가 91%로 나타났는데, 이러한 결과들에 비하여 본 연구의 성능이 근소하게 낮게 나타남을 볼 수 있었다. 이는 본 연구가 기본적인 문법정보 자질만을 사용한 것에 기인한 것으로, 문법 자질의 보완을 통하여 성능 향상을 꾀할 수 있을 것으로 보인다. 결과적으로 본 논문에서 제안한 바와 같이 문법정보와 음성정보를 이용하여 운율 경계를 추정하는 경우에는 비교적 우수한 성능으로 자동 레이블링이 가능하다는 점에서 수동 레이블링을 하는 것에 비하여 효율적임을 알 수 있다.

5. 결 론

본 논문에서는 문법정보와 음성정보를 이용하여 한국어 운율 경계를 추정하는 방법을 제안하였다. 제안된 방법으로 얻어진 강세구와 억양구의 경계 추정 결과를

K-ToBI 레이블링의 전사자 간의 일치도와 비교해 볼 때, 강세구 경계의 경우는 82.6%로 전사자 간의 일치도보다 높고, 억양구의 경우는 88.7%로서 전사자 간의 일치도에 비하여 다소 낮았다. 이는 통사 구조와 운율 구조가 상관관계가 있다고 해도 정확하게 일치하지는 않기 때문에, 그 상관관계에 따라 규칙을 설정하고 경계를 추정하는 문법정보 기반의 방법보다는, 본 논문에서 제안된 방법과 같이 문법정보 이외에도 실제 음성정보를 이용함으로써 더 좋은 성능을 얻을 수 있었다. 그러나 영어를 대상으로 한 운율 경계 추정 연구에서의 결과와 비교하였을 때 낮은 성능을 보이고 있기 때문에, 본 논문에서 음성모델 및 문법모델을 구성하기 위해 사용한 자질을 검토하여 각 모델을 개선할 필요가 있다. 현재 본 논문에서 사용한 음성정보 자질과 문법정보 자질은 경험에 바탕을 둔 것으로 문법정보 자질과 음성정보 자질을 선택하는 기준과 방법에 대한 후속 연구가 필요하다. 또한, 운율 경계의 자동 추정을 포함하는 전체 운율 자동 레이블링에 대한 연구를 진행할 계획이다.

참 고 문 헌

[1] D. Kahn, *Syllable-Based Generalizations in English Phonology*, Unpublished Ph.D. Dissertation, MIT, Boston, 1976.

[2] A. Prince, P. Smolensky, *Optimality Theory: Constraint Interaction in Generative Grammar*, Center for Cognitive Science Technical Report 2, Rutgers University, 1993.

[3] E. Selkirk, *Phonology and Syntax: The Relation Between Sound and Structure*, Cambridge: MIT Press, 1984.

[4] E. Selkirk, "On derived domains in sentence phonology", *Phonology Yearbook 3*, pp. 371-405, 1986.

[5] M. Nespor, I. Vogel, *Prosodic Phonology*, Dordrecht: Foris, 1986.

[6] M. Q. Wang, J. Hirschberg, "Automatic classification of intonational phrase boundaries", *Computer Speech and Language*, Vol. 6, No. 2, pp. 175-196, 1992.

[7] K. Ross, M. Ostendorf, "Prediction of abstract prosodic labels for speech synthesis", *Computer Speech and Language*, Vol. 10, No. 3, pp. 155-185, 1996.

[8] E. Shriberg, A. Stolcke, "Prosody modeling for automatic speech recognition and understanding", *Proc. ISCA Workshop on Prosody in Speech Recognition and Understanding*, pp. 13-16, 2001.

[9] K. Hirose, N. Minematsu, Y. Hashimoto, K. Iwano, "Continuous speech recognition of Japanese using prosodic word boundaries detected by mora transition modeling of fundamental frequency contours", *Proc. ISCA Workshop on Prosody in Speech Recognition and Understanding*, pp. 61-66, 2001.

[10] A. Batliner, R. Kompe, A. Kieβling, M. Mast, H. Niemann, E. Nöth, "M=Syntax+Prosody: A syntactic-prosodic labelling scheme for large spontaneous speech databases", *Speech*

Communication, Vol. 25, No. 4, pp. 193-222, 1998.

[11] H. Niemann, E. Nöth, A. Batliner, J. Buckow, F. Gallwitz, R. Huber, A. Kieβling, R. Kompe, V. Warnke, "Using prosodic cues in spoken dialog systems", *Proc. International Workshop on Speech and Computer*, pp. 17-28, 1998.

[12] S. Ananthakrishnan, S. S. Narayanan, "Automatic prosodic event detection using acoustic, lexical, and syntactic evidence", *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 16, No. 1, pp. 216-228, 2008.

[13] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C.Wightman, P. Price, J. Pierrehumbert, J. Hirschberg, "ToBI: A standard scheme for labeling prosody", *Proc. ICSLP*, pp. 867-869, 1992.

[14] D. J. Hirst, R. Espesser, "Automatic modelling of fundamental frequency curves using a quadratic spline function", *Travaux de l'Institut de Phonétique d'Aix*, Vol. 15, No. 1, pp. 71-85, 1993.

[15] D. J. Hirst, A. Di Cristo, R. Espesser, "Levels of representation and levels of analysis for the description of intonation systems", *Prosody: Theory and Experiment*, Berlin: Kluwer Academic Press, pp. 37-88, 2000.

[16] P. A. Taylor, "The rise/fall/connection model of intonation", *Speech Communication*, Vol. 15, No. 1-2, pp. 169-186, 1995.

[17] P. A. Taylor, "Analysis and synthesis of intonation using the tilt model", *Journal of the Acoustical Society of America*, Vol. 107, No. 3, pp. 1697-1714, 2000.

[18] J. t'Hart, R. Collier, "Integrating different levels of intonation analysis", *Journal of Phonetics*, Vol. 3, No. 4, pp. 235-255, 1975.

[19] H. Fujisaki, "Dynamic characteristics of voice fundamental frequency in speech and singing", *The production of speech*, Heidelberg: Springer-Verlag, pp. 39-55, 1983.

[20] S.-A. Jun, "K-ToBI(Korean ToBI) labeling conventions: Version 3.1.", *UCLA Working Papers in Phonetics*, pp. 149-173, 2000.

[21] 강선미, 권오일, "한국어 음성합성기의 운율 예측을 위한 의사결정트리 모델에 관한 연구", *음성과학*, 제14권, 제2호, pp. 91-103, 2007.

[22] A. Stolcke, E. Shriberg, R. Bates, M. Ostendorf, D. Hakkani, M. Plauche, G. Tür, Y. Lu, "Automatic detection of sentence boundaries and disfluencies based on recognized words", *Proc. ICSLP*, pp. 2247-2250, 1998.

[23] S. Lee, Y.-H. Oh, "Tree-based modeling of prosodic phrasing and segmental duration for Korean TTS systems", *Speech Communication*, Vol. 28, No. 4, pp. 283-300, 1999.

[24] K. Yoon, "A prosodic phrasing model for a Korean text-to-speech synthesis system", *Computer Speech and Language*, Vol. 20, No. 1, pp. 69-79, 2006.

[25] P. Taylor, A. W. Black, "Assigning phrase breaks from part-of-speech sequences", *Computer Speech and Language*, Vol. 12, No. 2, pp. 99-117, 1998.

[26] 김상훈, 성철재, 이정철, "운율구 경계현상 분석 및 텍스트에서의 운율구 추출", *한국음향학회지*, 제16권, 제1호, pp. 24-32, 1997.

[27] E. Sanders, P. Taylor, "Using statistical model to predict phrase boundaries for speech synthesis", *Proc. Eurospeech*, pp. 1811-1814, 1995.

[28] 김병창, 이근배, "자연어 처리 기반 한국어 TTS 시스템 구현", *말소리*, 제46호, pp. 51-64, 2003.

[29] X. Sun, T. H. Applebaum, "Intonational phrase break prediction using decision tree and N-gram model", *Proc. Eurospeech*, pp. 537-540, 2001.

[30] X. Zhang, J. Xu, L. Cai, "Prosodic boundary prediction based on Maximum Entropy model with error-driven modification", *Proc. ICSLP*, pp. 149-160, 2006.

[31] 강평수, 김진영, "피치 정보를 이용한 운율 경계 강도 예측", *대한전자공학회 신호처리 합동학술대회 논문집*, 제11권, 제1호, pp. 689-692, 1998.

[32] 이기영, 송민석, "한국 표준어 연속음성에서의 억양구와 강세구 자동 검출", *음성과학*, 제7권, 제2호, pp. 209-224, 2000.

[33] R. Cai, Z. Y. Wu, L. H. Cai, "Annotation of Chinese prosodic level based on probabilistic model", *Proc. ISCSLP*, pp. 375-378, 2002.

[34] S. Ananthakrishnan, S. Narayanan, "An automatic prosody recognizer using a coupled multi-stream acoustic model and a syntactic-prosodic language model", *Proc. ICASSP*, pp. 269-272, 2005.

[35] Y. Gotoh, S. Renals, "Sentence boundary detection in broadcast speech transcripts", *Proc. ISCA Workshop on Automatic Speech Recognition*, pp. 228-235, 2000.

[36] K. Chen, M. Hasegawa-Johnson, A. Cohen, "An automatic prosody labeling system using ANN-based syntactic-prosodic model and GMM-based acoustic-prosodic model", *Proc. ICASSP*, pp. 509-512, 2004.

[37] K. Iwano, "Prosodic word boundary detection using mora transition modeling of fundamental frequency contours", *Proc. Eurospeech*, pp. 231-234, 1999.

[38] 권오일, 홍문기, 강선미, 신지영, "코퍼스 방식 음성합성에서의 개선된 운율구 경계 예측", *음성과학*, 제9권, 제3호, pp. 25-34, 2002.

[39] 엄기완, 김진영, 김선미, 이현복, "품사셋에 의한 운율경계강도의 예측", *말소리*, 제35-36호, pp. 145-155, 1998.

[40] 김병창, 김승원, 이근배, "CRF를 이용한 운율경계추정 성능개선", *말소리*, 제57호, pp. 139-152, 2006.

[41] 정영임, 조선호, 윤애선, 권혁철, "구문 관계와 운율 특성을 이용한 한국어 운율구 경계 예측", *인지과학*, 제19권, 제1호, pp. 89-105, 2008.

[42] D. J. Hirst, H. Cho, S. Kim, H. Yu, "Evaluating two versions of the Momel pitch modeling algorithm on a corpus of read speech in Korean", *Proc. Interspeech*, pp. 1649-1652, 2007.

[43] 김효숙, 김정원, 김선주, 김선철, 김삼진, 권철홍, "국어 낭독체 발화의 운율경계 예측", *말소리*, 제43호, pp. 1-9, 2002.

[44] 김선희, 유현지, 홍혜진, 이호영, "Momel을 이용한 한국어의 억양 연구", *말소리*, 제63호, pp. 85-100, 2007.

[45] http://tcts.fpms.ac.be/synthesis/mbrola.html.

[46] http://www.phon.ucl.ac.uk/home/sampa/.

[47] http://www.fon.hum.uva.nl/praat/.

[48] S.-A. Jun, S. Lee, K. Kim, Y. J. Lee, "Labeler agreement in transcribing Korean intonation with K-ToBI", *Proc. ICSLP*, pp. 211-214, 2000.

접수일자: 2008년 5월 21일
게재결정: 2008년 6월 24일

▶ 김선희(Sunhee Kim)
주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 인문관 5동 313호
소속: 서울대학교 인문정보연구소
전화: 02) 880-7735
E-mail: sunhkim@snu.ac.kr

▶ 전재훈(Je Hun Jeon)
주소: Richardson, Texas 75083-0688, USA
소속: Department of Computer Science, University of Texas at Dallas
E-mail: jehun.jeon@gmail.com

▶ 홍혜진(Hyejin Hong)
주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 인문관 1동 426호
소속: 서울대학교 언어학과
전화: 02) 880-9039
E-mail: souble1@snu.ac.kr

▶ 정민화(Minhwa Chung) : 교신저자
주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 인문관 3동 406호
소속: 서울대학교 언어학과
전화: 02) 880-9195
E-mail: mchung@snu.ac.kr