

Momel을 이용한 한국어의 억양 연구*

김선희, 유현지, 홍혜진, 이호영(서울대)

<차 례>

- | | |
|-------------------------|-----------------------|
| 1. 서론 | 3. 결과 및 논의 |
| 2. 연구 방법 | 3.1. K-ToBI 데이터 분석 결과 |
| 2.1. Momel-INTSINT | 3.2. EUROM 코퍼스 분석 결과 |
| 2.2. Momel을 이용한 억양 레이블링 | 3.2.1. 수동 검토 |
| 2.3. 음성 코퍼스 | 3.2.2. AP 패턴 |
| 2.3.1. K-ToBI 데이터 | 3.2.3. IP 경계 성조 |
| 2.3.2. EUROM 코퍼스 | 4. 결론 |

<Abstract>

A Study on Korean Intonation Using Momel

Sunhee Kim, Hyunji Yoo, Hyejin Hong, Ho-Young Lee

This paper aims to propose how to extract intonation patterns using Momel, a pitch stylization algorithm, and to present results of analyzing speech corpora in comparison with those in earlier researches. Two speech corpora are used: one is the sound files obtained from the K-ToBI web site, and the other consists of 80 passages pronounced by 4 speakers (2 male and 2 female). The results show that Momel provides significant pitch targets which can be labeled as H and L tones within prosodic units such as Accentual Phrase (AP) and Intonation Phrase (IP). The resulting AP patterns and IP boundary tone patterns correspond to those in earlier researches. Thus, this study will contribute to the study of intonation as well as to the development of automatic intonation labeling systems.

* Keywords: Korean, Intonation, Momel, INTSINT, Tone, Pitch target.

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

1. 서 론

억양(intonation)은 언어학에서 중요한 연구 분야일 뿐만 아니라, 음성인식과 음성합성과 같은 음성처리 분야에서도 필수적인 연구 분야이다. 현재까지 수행되어 온 억양 연구 가운데 대용량 음성 코퍼스의 억양을 전사하기 위한 억양 전사 표준안으로 제안되었던 ToBI(Tone and Break Indices)[1]가 영어 외에도 여러 다른 언어들의 억양 전사에 이용되면서 음성처리 분야에서는 하나의 대표적인 억양 전사 시스템으로 인정되고 있다. ToBI는 합성음의 자연성 향상과 자연 발화에 대한 인식을 향상을 위한 대용량 음성자료의 구축을 위하여 음성학자들과 음성공학자들이 합의하여 공통적인 규약으로 집대성한 결과물이다.

ToBI 기반의 한국어 억양 모델은 K-ToBI라고 하는데[2], 이는 운율 단위로서 강세구(Accentual Phrase: AP)와 억양구(Intonational Phrase: IP)로 구성되고, 강세구에는 구억양(phrasal tone)이 없히고 억양구의 마지막 음절에는 억양구의 경계 성조(IP Boundary Tone)가 추가된다. 구억양과 경계성조는 L과 H 두 가지 성조(tone)로 표기한다.

그러나 K-ToBI에 의한 음성 데이터의 분석은 실제 레이블링 작업이 장시간 훈련받은 전문가에 의하여 수행되어야 하고 전문가들 사이에도 일치도가 높지 않은 문제가 지적되었고, 그 결과 실제 대용량 음성자료의 구축에 이용되는 데 한계가 있어서 음성합성이나 인식 시스템에 이용된 사례가 많지 않았다. K-ToBI 레이블링 과정에서 발생하는 어려움의 원인 중 하나는 문장에 무성 자음들이 섞여 있거나 모음이 무성화 되어 실제 F0 값이 잘 나타나지 않아 전체 억양 곡선에 상응하는 억양 목표점(pitch target)을 찾는 것이 무척 어렵다는 것이다. 또한, 여러 다른 특성을 가진 음성들이 섞여 있고 음성들이 하나의 연쇄로 발화되는 경우에 동시 조음에 의한 여러 가지 변화를 포함하므로 의미 있는 억양 곡선을 재구성하는 것이 쉽지 않게 된다.

본 논문은 프랑스에서 개발된 자동 억양 궤적 추출 알고리즘인 Momel[3, 4]을 이용하여 억양 패턴을 추출하는 방법을 제시하고, 제시한 방법론에 따라 음성 데이터를 분석한 다음 그 결과를 기존의 연구[2, 5, 6]와 비교하는 것을 목적으로 한다. 자동으로 억양 궤적을 추출하는 Momel을 이용하면 수동 작업에 대한 의존도를 줄일 수 있어서 ToBI 레이블링의 문제점으로 지적된 훈련 받은 전문가에 의한 작업의 필요성을 경감시키고, 전사자 간에 불일치가 야기되었던 부분이 적어짐에 따라 결과적으로 보다 일관성 있는 데이터 분석이 가능해진다. 또한, 이러한 연구는 억양 레이블링 및 AP/IP 경계 추출의 자동화에 기여하여, 실제 대용량 음성 데이터의 구축에 이용되고, 나아가 음성합성이나 인식 시스템에도 이용되길 기대한다.

2. 연구 방법

2.1. Momel-INTSINT

억양은 개별 분절음의 음향적 특성이 반영된 미시적인 부분(microprosodic component)과 전체 억양 패턴을 보여주는 거시적인 부분(macrosodic component)으로 구성되어 있다[3, 7]. 언어학적으로 의미 있는 억양은 거시적인 부분에 의하여 표현되는데, 억양의 거시적인 억양 곡선을 모델링하는 방법으로는 Hirst 모델[3, 4], RFC(rise/fall/connection) 모델[8], 경사도(Tilt) 모델[9], IPO 모델[10] 등이 있다. 미시적 변이와 거시적 변이를 하나로 모델링한 것으로는 중첩 모델(superpositional model)이라고 불리는 후지사키 모델[11]이 있다.

여러 가지 억양 모델 가운데 Momel을 본 연구에서 채택한 이유는 두 가지이다. 먼저, Momel은 거시적인 억양 궤적을 추출하는 면에서 위에서 언급한 다른 모델들과 비교하여 우수한 성능을 보이는 것으로 알려져 있다[7]. 뿐만 아니라, Momel은 Unix, Windows, Linux, MacIntosh 등 여러 다른 환경에서 MES (Unix), SFS (Windows), Praat (Macintosh, Windows, Linux, FreeBSD, SGI, Solaris, HPUX) 등의 음성 분석 도구에 설치하여 사용할 수 있도록 설계되어, 억양을 비롯한 여러 음향음성학적 및 언어학적 분석 연구에 매우 유용하게 이용할 수 있다. 특히 Windows 환경에서 가장 보편적으로 쓰이고 있는 Praat에서 여러 다른 기능들과 함께 억양 분석을 할 수 있다는 것이 큰 장점이라고 할 수 있다.

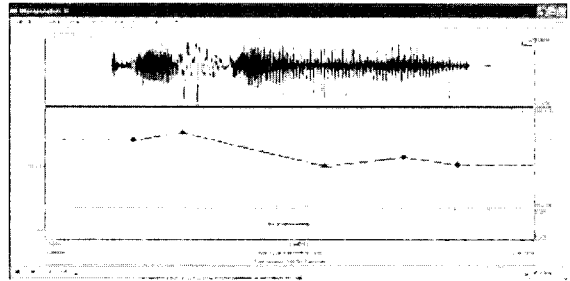
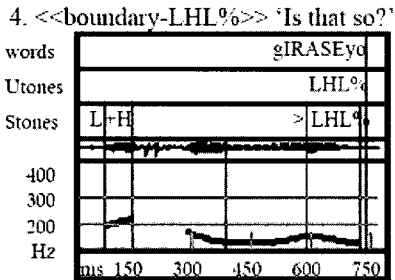
Momel은 [3]에서 quadratic spline function을 기반으로 한 억양 궤적 추출 알고리즘으로서 제안되었는데, [3]에서 제안된 방법론을 자동화하여[4], 현재 웹상에서는 [4]의 제안에 따른 자동화된 Momel을 얻을 수 있다[12].

자동화된 Momel의 성능은 [13]에서 영어, 프랑스어, 독일어, 스페인어, 불어 등의 언어들을 대상으로 하여 평가되었다. [13]에 따르면 각각의 언어에 대하여 Momel을 이용하여 자동으로 추출된 억양 목표점의 숫자와 수동으로 수정한 억양 목표점의 숫자를 비교하여, 정보 이론(Information Theory)에서 사용되는 F-척도(F-measure)¹⁾에 의하여 그 결과를 보고하였는데, 이 다섯 가지 언어에서 평균 F=95.5를 얻어 비교적 높은 성능을 보였다고 한다.

다음 <그림 1>에서 왼쪽은 [2]에 수록된 예제 음성 “그랬어요?”(LHL% 경계 성조)의 성조 레이블링한 것을 나타낸 그림이고, 오른쪽은 동일한 음성 파일을 Windows 환경에서 Praat를 이용하여 Momel로 거시적 억양 곡선으로 변환한 다음

1) F-척도(F-measure)는 정보 검색 등 정보 이론에서 사용되는 성능 평가 척도로서 precision (여기에서는 모든 목표점과 Momel에 의하여 옳게 추출된 목표점의 비율)과 recall(옳게 추출한 모든 목표점과 Momel에 의하여 옳게 추출된 목표점의 비율)과의 조화 평균으로 나타내는데, F 값이 100에 가까울수록 좋은 결과로 평가된다.

이를 자동 알고리즘에 의하여 추출된 억양 목표점의 연결로 나타낸 것이다. 즉, 입력한 음성 파일에 대하여 Momel을 이용하면 자동으로 기본주파수의 거시적인 부분을 정의하는 억양 목표점의 연쇄를 얻을 수 있는데, 이 각각의 억양 목표점을 H나 L로 레이블링하면 왼쪽과 같은 레이블링 결과를 얻을 수 있음을 알 수 있다. 이때 도출된 억양 목표점은 억양 궤적 상에서 변곡점이 되며, 언어학적으로 의미 있는 부분에 해당한다.



<그림 1> K-ToBI[2]에 수록된 한 예제 음성 “그랬어요?”(LHL% 경계 성조)의 성조 레이블링 결과(왼쪽)와 이에 대한 Momel 분석 결과(오른쪽)

이와 같이 Momel을 이용하면 음성 파일로부터 자동으로 전체 억양의 궤적을 추출하고, 그 궤적 위에 있는 억양 목표점을 찾아서 표시해 주므로, 그 억양 목표점을 이용하여 손쉽게 억양을 레이블링할 수 있게 된다.

Momel에 의하여 추출된 억양 목표점에 대한 억양 레이블링을 위하여 [4, 14]는 INTSINT(International Transcription System for INTonation)를 제안하였다. [4, 14]에 따르면, INTSINT는 초기에는 운율적으로 다른 특성을 보이는 두 언어인 프랑스어와 영어의 억양 패턴을 동시에 나타낼 수 있는 방법으로 제안되었는데, 현재는 억양에서의 IPA(International Phonetic Alphabet)로서 모든 언어에 적용이 가능하다[4]. INTSINT는 Momel에 의하여 도출된 각각의 억양 목표점들에 대하여 음역 내에서 차지하는 절대값(Top, Bottom, Mid)과 앞 음절의 높이와 비교한 상대값(Higher, Lower, Same, Upstepped, Downstepped)을 이용하여 그 레이블을 결정하게 된다. 이러한 방법은 [14]에서 9개 언어의 분석에 적용하였고, 실제 음성 합성에 이용되는 경우에 매우 효과적이라고 보고되었다. 하지만, [14]에서도 지적된 바와 같이 이는 음성학적인 레이블링 시스템으로서 음운론적인 해석을 포함하는 억양 패턴을 기술하는 데에는 적합하지 않으므로, 본 연구에서는 억양 패턴을 추출하기 위해서 Momel만을 이용하는 것으로 제한하기로 한다.

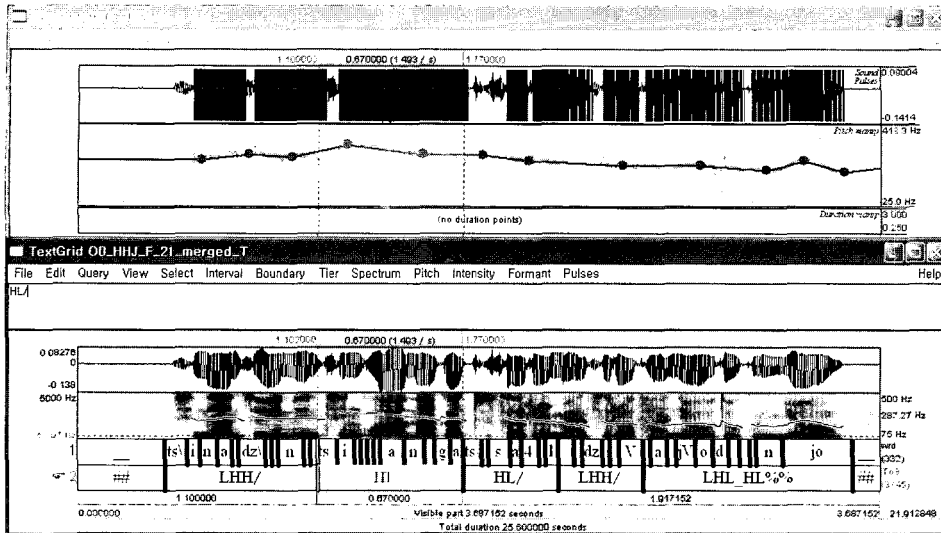
2.2. Momel을 이용한 억양 레이블링

본 논문에서 제안하는 Momel을 이용한 억양 레이블링 방법은 다음과 같이 4단계로 이루어진다.

- 1) 음성 파일의 Momel 분석
- 2) Momel로 검출된 억양 목표점 수정
- 3) AP, IP 경계 표시
- 4) 성조 레이블링: Momel에서 검출된 억양 목표점을 상대적인 F0 값을 기준으로 H, L로 레이블링
 - 4.1) AP 성조 레이블링
 - AP 내에서 가장 높은 목표점을 H로 하여 상대적으로 H, L을 표시
 - 4.2) 수평조의 판단: 두 목표점의 차이가 $\pm 10\text{Hz}$ 이내인 경우²⁾
 - 4.3) IP 경계 성조 레이블링
 - 마지막 음절의 시작 지점을 확인해서 마지막 음절에 표시된 억양 목표점을 H, L로 레이블링

다음 <그림 2>는 위 방법에 따라 Praat를 이용하여 Momel로 검출된 억양 목표점을 수정한 결과 파일[(2)번 단계]과 음소 레이블링된 TextGrid 창을 연결하고, AP/IP 경계를 표시한 다음[(3)번 단계] 억양 목표점을 레이블링한 결과[(4)번 단계]를 보인 것이다.

2) 수평조에 해당하는 경우를 동일한 F0 값을 갖는 경우만으로 한정하는 것은 극단적인 태도로써 본 연구에서는 수평조를 판단하는 기준을 전후 목표점과 비교하여 $\pm 10\text{Hz}$ 이내인 경우로 하였으나, 실제로 자동화하는 경우는 남성과 여성 화자 혹은 개인별 피치 폭에 따라 비율로 정하는 것이 합리적일 것이다.



<그림 2> 음성 파일 “지난주에는 친구 하나가 주사를 맞으러 다녀 오더군요”에 대하여 Momel로 검출된 억양 목표점 수정한 결과 파일[(2)번 단계]과 음소 레이블링된 TextGrid 창을 연결하고, AP/IP 경계를 표시한 다음[(3)번 단계] 억양 목표점을 레이블링한 결과[(4)번 단계]

2.3. 음성 코퍼스

2.3.1. K-ToBI 데이터

본 연구에서 이용한 음성 코퍼스는 두 가지인데, 먼저 K-ToBI[2]에서 제시하고 있는 음성 파일들을 Momel을 사용하여 분석하고 이를 근거로 성조를 레이블링한 결과를 원래의 ToBI 레이블링 결과와 비교한다. K-ToBI 데이터는 [2]에서 내려 받은 음성 파일들로서 K-ToBI로 레이블링이 되어 있는 20개의 음성 파일(낭독체 16개, 드라마 4개)이다.

2.3.2. EUROM 코퍼스

이 연구를 위해 유럽의 표준 낭독체 코퍼스인 EUROM 코퍼스[15]의 기준에 따라 한국어 낭독체 음성 코퍼스(EUROM Korean)를 제작했다. 이 작업에는 서울 토박이 20대 남녀 각 5명으로 총 10명이 제보자로 참여하였다. 발성 목록은 EUROM 텍스트(화자 당 40문단, 총 166문장)이고, 총 녹음 시간 약 2시간이다. 녹음은 Marantz PMD 670(녹음기)와 Shure SM-58(마이크)를 이용하여 방음시설이 된 녹음

실에서 진행하였다.

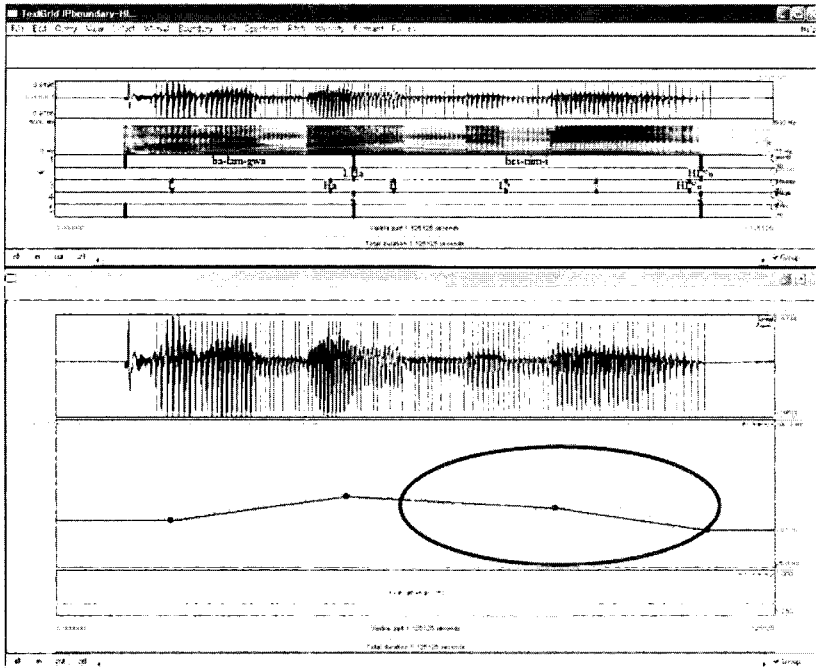
본 논문에서는 전체 코퍼스 가운데 4명(남녀 각 2명)이 각각 20개의 문단을 발화한 총 80개 문단, 320개 문장을 그 분석 대상으로 하였다. 녹음된 음성 데이터는 다국어 합성기인 MBROLA[16]를 이용하여 SAMPA(Speech Assessment Methods Phonetic Alphabet)[17]로 음소 레이블링을 실행하였으며 AP와 IP 경계는 해당 음소의 경계와 정렬시켰다.

3. 결과 및 논의

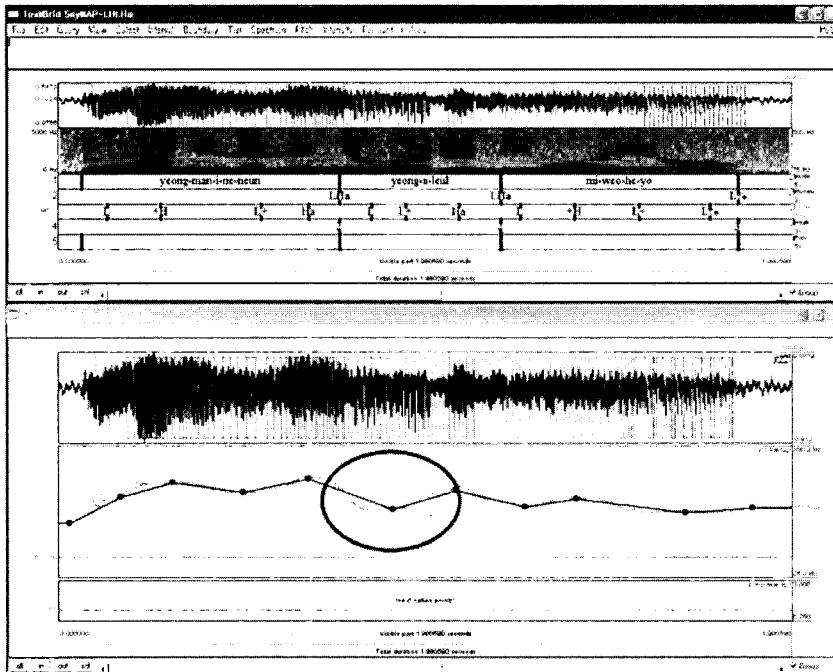
3.1. K-ToBI 데이터 분석 결과

[2]의 음성 데이터를 Momel을 이용해 레이블링할 때, AP/IP 경계는 [2]에서 이미 레이블링 되어 있는 그대로를 이용하므로, 2.2.에서 제안한 4단계 가운데 AP/IP 경계 표시 과정을 제외한 3단계로 진행하였다. 이와 같은 방법에 의하여 레이블링을 수행하는 경우에 확인이 필요한 부분은 2) Momel로 검출된 억양 목표점의 수정 단계와 4) 성조 레이블링의 부분이다.

분석 대상이 되는 20개의 발화는 [2]에서 69개의 AP와 29개의 IP로 총 98개의 운을 단위로 레이블링 되어 있다. Momel을 이용하여 추출된 총 목표점 수는 176개인데, AP와 IP의 성조를 레이블링할 때 ToBI에 의한 레이블링 결과와 비교하여 관찰되는 차이 가운데 가장 문제가 되는 것은 ToBI에서의 분석에 비하여 목표점 수가 충분하지 않아 목표점 추가가 필요한 경우이다. Momel에 의하여 분석된 결과를 ToBI 결과와 일치시키기 위해서는 총 72개의 목표점의 추가가 필요하였다. 목표점 추가가 필요한 경우는 다시 2가지 경우로 분류된다: (1) LH 패턴이 반복될 때, 반복되는 성조의 중간값으로 하나의 목표점을 잡는 경우(<그림 3> 참조), (2) 연속한 둘 이상의 음절이 거의 같은 높이(수평조)일 때 목표점을 하나만 잡는 경우(<그림 4> 참조). 이 두 경우는 모두 두 개의 목표점으로 나타내어야 원음과 유사한 억양을 얻을 수 있는데, 하나의 목표점으로 나타나 원음의 억양과 차이가 나타나는 부분이다.



<그림 3> LH 패턴이 반복될 때, 반복되는 성조의 중간값으로 하나의 목표점을 잡는 경우(점선: 원래의 F0 값, 직선: Momet에 의하여 추출된 억양 목표점들을 연결한 억양 궤적)



<그림 4> 연속한 둘 이상의 음절이 거의 같은 높이(수평조)일 때, 목표점을 하나만 잡는 경우(점선: 원래의 F0 값, 직선: Momet에 의하여 추출된 억양 목표점들을 연결한 억양 궤적)

이와 같이 목표점의 추가가 필요한 경우가 많은 이유는 K-ToBI가 미시적인 억양 변화도 모두 구별해서 레이블링하는 반면에, Momel을 이용하는 경우에는 인접하는 두 개의 목표점의 F0 값이 일정한 임계값 이상의 차이가 있는 경우에만 최종적인 목표점으로 선택되어 거시적인 억양 곡선을 추출하도록 되어 있는 Momel 알고리즘 자체에 기인한다고 볼 수 있다.³⁾

이 밖에도 발화 경계 밖에서 의미 없는 목표점을 잡는 경우도 있었는데, 이러한 발화 경계 밖의 목표점은 음성에 포함된 소음에 기인한 것으로 보인다. 또한, 목표점이 해당 음절에 있지 않은 경우, 경계 전후에 목표점이 밀린 경우들이 있었는데, 이들은 전체 억양 패턴에는 영향을 주지 않았다.

K-ToBI에 의한 억양 레이블링과 Momel의 억양 목표점을 이용하여 억양을 레이블링할 때 야기되는 가장 큰 차이는 K-ToBI가 강세구의 특정 음절들을 기준으로 하여 레이블링하는 데 반하여, Momel을 이용하는 경우는 운율 단위 내부에서 특정 음절과는 무관하게 의미 있는 억양 목표점만을 추출하여 레이블링한다는 것이다. K-ToBI에서 4음절 이상으로 이루어진 AP를 레이블링하는 경우에는 기본적으로 AP 내부에서 앞 두 음절과 뒤 두 음절, 모든 음절의 성조를 기준으로 하는데, Momel을 이용하는 경우에는 음절 수와는 상관없이 의미 있는 목표점만을 추출하게 된다. 다시 말하자면, K-ToBI에 의하여 레이블링하는 경우는 성조가 일정한 음절 위치에 나타나는 것에 반하여 Momel을 이용하는 경우는 그렇지 않게 된다. 그러나, 이러한 차이에도 불구하고 Momel에 의하여 추출되는 억양 목표점을 기반으로 성조를 레이블링하는 경우에 최종적으로 얻어지는 성조 패턴은 K-ToBI에 의하여 얻을 수 있는 성조 패턴과 거의 일치하는 것을 볼 수 있었다.

3.2. EUROM 코퍼스 분석 결과

분석 대상이 되는 80개의 발화를 위 2.2.에서 제시한 방법에 따라 AP와 IP경계를 표시한 다음 억양 패턴을 레이블링한 결과, 최종적으로 1729개의 AP와 718개의 IP로 총 2447개의 단위로 레이블링되었다. 이때 얻어진 AP의 평균 지속시간은 485.04ms, 평균 음절수 3.40이었고, IP의 평균 지속시간은 764.23ms, 평균 음절수는 4.67이었다.

3) <그림 3>에서 보는 것과 같은 문제는 Momel 알고리즘의 문제이지만 <그림 4>에서 보는 것과 같은 문제는 K-ToBI 자체의 문제에 기인한다고도 할 수 있다. [18]에서 논의했듯이 현재의 K-ToBI는 청각적으로 구별되지 않는 미세한 억양 차이도 구별해서 전사하기 때문에 Momel의 분석 결과에 대해 목표점을 추가하여야만 동일한 결과를 얻을 수 있게 된다.

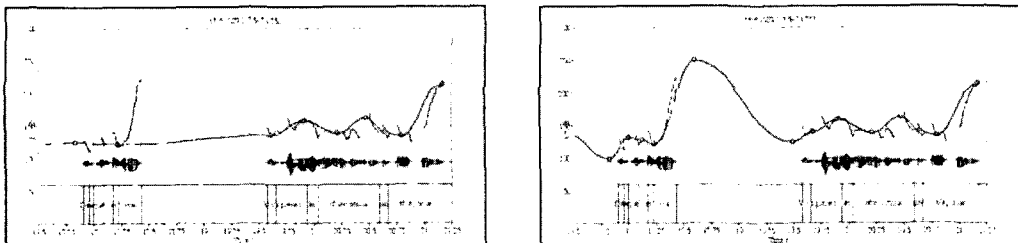
3.2.1. 수동 검토

음성 데이터는 2.2.절에서 제시한 방법에 따라 4단계로 억양 레이블링을 하였는데, 위 K-ToBI 데이터에서와 마찬가지로 수동 확인이 필요한 부분은 2) Momel로 검출된 억양 목표점의 수정 단계와 4) 성조 레이블링 부분이다. 본 데이터에서 AP/IP 경계 표시의 문제는 현재 Momel과는 무관한 부분으로 모두 2명의 훈련된 전사자에 의하여 수동으로 작업하였다.

Momel로 검출된 억양 목표점의 수정이 필요한 경우와 그 수정 내용은 다음과 같다.

- 1) 문장 시작과 끝 부분에 필요 없는 목표점을 하나씩 잡은 경우 → 불필요한 목표점을 지웠다.
- 2) LH pattern이 반복될 때, 반복되는 tone의 중간 값 하나만 목표점으로 잡은 경우 → 중간값의 위치를 L 위치로 이동하고 H와 연결될 목표점을 하나 추가하였다.
- 3) 연속한 둘 이상 음절이 수평조로 실현되는데 하나의 억양 목표점으로 잡은 경우 → 두 음절 중간 위치의 목표점을 하나의 음절로 이동시키고 나머지 음절에 같은 높이의 목표점을 추가하였다.

1)의 경우는 여러 개의 문장으로 구성된 하나의 문단이 하나의 음성 파일로 발화되거나, 연속하는 억양구 사이에 긴 휴지(pause)가 있는 경우에 흔히 야기되는 오류로서 250ms 이상의 휴지 구간을 검출한 방법을 추가하여 개선한 최근 Momel 알고리즘을 이용하여 해결하였다[19]. 아래 <그림 5>에서 왼쪽은 휴지 앞에서 원래의 F0 값과 Momel에서 도출된 값이 차이가 나는 것을 볼 수 있는데, 오른쪽에 서는 두 값이 거의 일치하는 것을 볼 수 있다.



<그림 5> 기존의 Momel에 의하여 추출된 억양 곡선(왼쪽)과 개선된 Momel에 의하여 추출된 억양 곡선(오른쪽) (비연속선: 원래의 F0 값, 연속선: Momel에 의하여 추출된 억양 궤적)

2)의 경우는 이미 3.1.절의 K-ToBI 데이터 분석에서도 지적인 바와 같이 목표점 추가가 필요한 경우로 원음을 들으면서 필요한 목표점을 추가하였다. K-ToBI 데이터의 경우는 분석 대상이 되는 음성 파일의 수가 많지 않아서 이러한 목표점 추가에 대하여 남녀 성별 차이나 화자의 차이를 볼 수 없었는데, EUROM 데이터의 경우는 이러한 목표점 추가가 필요한 경우가 피치 대역폭이 큰 여성 화자보다 피치 대역폭이 좁은 남성 화자에서 훨씬 빈번하게 관찰되는 것을 볼 수 있었다. 이는 이미 위에서 언급한 대로 인접하는 두 개의 목표점의 FO 값이 일정한 임계값 이상의 차이가 있는 경우에만 최종적인 목표점으로 선택되어 거시적인 억양 곡선을 추출하도록 되어 있는 Momel 알고리즘 자체에 기인한 것으로 남성의 경우나 피치 대역폭이 적은 화자의 경우에 특히 빈번하게 관찰될 수 있는 오류이다. 이와 같은 방법으로 수정을 한 경우에 여성들은 IP 경계 성조를 제외하고 거의 문제가 없었는데, 남성 화자들의 경우는 2)에 해당하는 경우가 많아서 보다 많은 수정이 필요하였다.

이 밖에도 문장 마지막 음절의 성조가 경계 밖으로 밀린 경우가 있었는데 이때에는 경계 밖의 목표점을 경계 안의 마지막 모음으로 옮기면 되고, 이러한 수정은 억양 패턴의 차이를 야기하지 않는다.

다음은 억양 목표점을 수정한 다음 추출된 억양 목표점에 성조를 부여하여 억양 레이블링을 할 때 관찰되는 오류와 그 수정한 예이다.

- 1) 하나의 억양 목표점만 나타나는 경우 → 음성을 확인하여 HH, LL, HL, LH 로 표기하였다.
- 2) AP, IP의 경계 전후에 목표점이 밀린 경우 → 경계 내부로 목표점을 이동하였다.
- 3) AP, IP 경계에서 억양 목표점이 나타나지 않는 경우 → 최소한의 목표점을 추가하였다.

하나의 억양 목표점만이 나타난 경우는 1음절어인 경우만이 아니라 2음절 이상의 경우도 해당되어 음성을 확인하여 각각 HH, LL, HL, LH 등으로 표기할 필요가 있었다. AP, IP 경계 전후에서 목표점이 밀린 경우도 음성을 확인하여 목표점을 이동하였다. AP, IP 경계에서 억양 목표점이 나타나지 않는 경우는 위에서 언급한 대로 Momel 알고리즘에 기인하여 의미 있는 억양 목표점으로 최종 도출되지 않은 것으로 음성을 기준으로 최소한의 목표점을 추가하였다.

수정된 억양 목표점을 바탕으로 성조 레이블링을 수행한 결과 총 2447개의 억양 단위 가운데 296개를 수정하여 총 수정률 12.10%를 보였다.

3.2.2. AP 패턴

[5, 20]은 억양의 형태 중심 모델(혹은 pitch movement model)을 기반으로 한 연구로서 운율 단위(억양 및 리듬 단위)를 말토막과 말마디로 구분하고, [21, 22]는 억양 목표점(pitch target) 혹은 피치 레벨(pitch level: 음고 층위)로서 억양을 나타내는 모델을 기반으로 한 연구로서 운율 단위로는 AP와 IP를 상정한다. 비록 이론은 다르지만 [2, 22]의 억양구와 강세구는 [20]의 말마디, 말토막과 거의 일치하는 단위이다[5]. 따라서, [22]의 AP 억양 패턴은 각각에 상응하는 [20]의 억양 패턴으로 나타낼 수 있으며, 같은 방법으로 Momel을 이용하여 추출한 억양 패턴도 [20], [22]와 함께 나타낼 수 있다. 다음 <표 1>은 Momel을 이용하여 추출된 억양 목표점을 이용하여 레이블링한 억양 패턴을 [20] 및 [22]와 비교한 것이다.

<표 1> Momel을 이용하여 추출한 AP 억양 패턴

이호영[20]	Jun[22]	Momel을 이용한 결과		
		AP	빈도	%
Level	H Ha	H/	109	6.3
		HH/	89	5.15
	L La	L/	37	2.14
		LL/	16	0.93
Rising	L Ha	LH/	766	44.3
	L+H Ha	LHH/	98	5.67
	L L+Ha	LLH/	5	0.29
Falling	H La	HL/	166	9.6
	H+H La	HHL/	14	0.81
	H L+La	HLL/	5	0.29
	H+H L+La	HHLL/	1	0.06
Fall-Rising	H L+Ha	HLH/	64	3.7
	H+H L+Ha	HHLH/	12	0.69
Rise-Falling	L+H La	LHL/	204	11.8
	L+H L+La	LHLL/	2	0.12
		LHHL/	2	0.12
Rise-Fall-Rising	L+H L+Ha	LHLH/	138	7.98
		LHHLH/	1	0.06
Total			1729	100

Momel을 이용하여 추출한 AP의 억양 패턴을 K-ToBI에서 제안한 14개의 AP

패턴과 비교해 보면, 14개의 패턴 외에도 H, L, LHHL, LHHLH가 관찰되었음을 볼 수 있다. 이 가운데, LHHL와 LHHLH는 각각 LHL, LHLH의 변이형으로 간주되는데, 이들의 빈도는 아주 낮았다. 거기에 비하여 H와 L은 꽤 높은 빈도를 보이며 나타났는데, 이들은 대부분 1음절어가 아닌 경우에도 그 AP 내부에서 특별한 억양 목표점이 추출되지 않고 하나의 성조로 실현된 예들이었다.

3.2.3. IP 경계 성조

IP 경계 성조로는 H%, L%, HL%, LH%, LHL%의 5개가 각각 문장의 중간과 문장의 끝에 나타났다.

<표 2> Momel을 이용하여 추출한 IP 경계 성조

이호영[5, 20]	Jun[22]	Momel을 이용한 결과				
		패턴	문장 중간		문장 끝	
			빈도	%	빈도	%
높은 수평조 (High Level): H%	H%	H%	138	36.13	52	15.48
가운데 수평조 (Mid Level): M%						
낮은 수평조 (Low Level): L%	L%	L%	55	14.4	78	23.21
높내림조 (High Fall): HL%	HL%	HL%	176	46.07	198	58.93
낮내림조 (Low Fall): ML%						
온오름조 (Full Rise): LH%	LH%	LH%	6	1.57	3	0.89
낮오름조 (Low Rise): LM%						
오르내림조 (Rise Fall): LHL%	LHL%	LHL%	7	1.83	5	1.49
내리오름조 (Fall Rise): HLH%	HLH%					
	LHLH%					
	HLHL%					
	LHLHL%					

음성 코퍼스 분석 결과 IP 경계 성조로는 문장 중간이나 문장 끝에서 모두 HL%이 가장 많이 나타나는 것을 볼 수 있었는데, 이는 해당 음성 코퍼스가 낭독체 평서문으로 구성되어 있기 때문이라고 볼 수 있다. 문장 끝 IP에서의 H%는 의문문에서 일반적으로 나타나는 것으로 문장 중간에서 나타나는 H%와는 다른 기능을 갖는데, H%는 문장 중간의 위치에서 문장 끝에 비하여 두 배 정도 자주 나타나는 것을 볼 수 있다. 반대로 L%의 경우는 문장 끝에 비하여 문장 중간에서는 절반 정도의 빈도로 나타나는 것을 볼 수 있다. LH%, LHL%의 경우는 문장에서의 위치에 관계없이 2% 이하의 낮은 빈도를 보였다.

<표 1>과 <표 2>에서 보듯이 Momel을 이용하여 자동 추출된 억양 목표점을 이용하여 음성 데이터를 분석할 때 도출되는 억양 패턴은 기존의 K-ToBI를 기반으로 한 한국어에 대한 분석 결과[2, 5, 6, 20, 22]와 거의 일치함을 볼 수 있었다.

4. 결 론

본 논문은 자동 억양 궤적 추출 알고리즘인 Momel을 이용하여 한국어의 억양 패턴을 추출하는 방법을 제안하고, 제안한 방법론에 따라 음성 데이터를 분석한 다음 그 결과를 기존의 연구[2, 5, 6, 20, 22]와 비교하였다.

두 가지 음성 코퍼스를 대상으로 분석한 결과 Momel에 의하여 추출된 억양 목표점을 기준으로 성조를 레이블링할 경우에는 억양 목표점이 부족한 경우가 문제가 되었으나, 억양 목표점을 추가하여 성조를 레이블링한 경우에는 기존의 연구와 비교할 때 얻어진 AP 패턴과 IP 경계 성조 패턴은 거의 유사하였다. 일단 음성 파일을 기준으로 억양 목표점의 오류가 수정된 이후에 AP, IP의 경계를 표시한 다음, AP와 IP의 내부의 성조를 레이블링할 때 주로 하나의 억양 목표점만 나타나서 음성을 확인하여 레이블링하는 것이 필요한 경우가 대부분으로 이 때의 오류율은 EUROM 데이터의 분석에서 12.10%를 보였다.

따라서 본 논문에서는 Momel을 이용하여 억양 패턴을 추출하는 경우에 수동 작업에 의한 의존도를 줄여서 억양 패턴을 보다 일관성 있게 추출할 수 있게 되어, ToBI 레이블링의 문제점으로 지적된 훈련 받은 전문가에 의한 작업의 필요성을 경감시키고 전사자 간에 불일치가 야기되었던 부분이 적어짐에 따라 결과적으로 보다 일관성 있는 데이터 분석이 가능함을 보였다. 뿐만 아니라, 이러한 결과는 Momel에 의하여 추출된 억양 목표점을 기준으로 억양 레이블링을 자동화할 경우 우수한 성능을 기대할 수 있음을 보여준다고 할 수 있다.

참 고 문 헌

- [1] K. Silverman, M. Beckman, J. Pitrelli, M. Ostendorf, C. Wightman, P. Price, J. Pierrehumbert, J. Hirschberg, "ToBI: a standard for labelling English prosody", *Proc. ICSLP*, Banff, Canada, pp. 867-870, 1992.
- [2] S. -A. Jun, "K-ToBI (Korean ToBI) labeling conventions: Version 3.1.", *UCLA Working Papers in Phonetics*, Vol. 99, pp. 149-173, 2000.
- [3] 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, pp. 71-85, 1993.
- [4] D. J. Hirst, A. Di Cristo, R. Espesser, "Levels of representation and levels of analysis for the description of intonation systems", In: M. Horne(Ed.), *Prosody: Theory and Experiment*, Dordrecht: Kluwer, pp. 51-87, 2000.
- [5] 이호영, "국어 핵억양의 음향음성학적 연구", *말소리*, 38권, pp. 25-39, 1999.
- [6] 이숙향, 김종진, "대용량 데이터베이스를 이용한 한국어 운율 특성에 관한 연구", *한국 음향학회지*, 24권, 2호, pp. 117-126, 2005.
- [7] W. Wang, S. Narayanan, "Piecewise linear stylization of pitch via wavelet analysis", *Proc. Eurospeech 2005*, Lisbon, Portugal, 2005.
- [8] P. A. Taylor, "The rise/fall/connection model of intonation", *Speech Communication*, Vol. 15, pp. 169-186, 1995.
- [9] 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.
- [10] J. 't Hart, R. Collier, "Integrating different levels of intonation analysis", *Journal of Phonetics*, Vol. 3, pp. 235 - 255, 1975.
- [11] H. Fujisaki, "Dynamic characteristics of voiced fundamental frequency in speech and singing", In *The production of speech*, Berlin: Springer, 1983.
- [12] <http://aune.lpl.univ-aix.fr/~auran/english/ressources.html>
- [13] E. Campione, "Etiquetage semi-automatique de l'intonation dans les corpus oraux: algorithmes et methodologie", Thèse de doctorat, Aix-en-Provence, Université de Provence, 2001.
- [14] D. J. Hirst, A. Di Cristo (Ed.), *Intonation Systems: A Survey of Twenty Languages*, Cambridge: Cambridge University Press, 1998.
- [15] <http://www.phon.ucl.ac.uk/shop/eurom1.php>
- [16] <http://tcts.fpms.ac.be/synthesis/mbrola.html>
- [17] <http://www.phon.ucl.ac.uk/home/sampa/>
- [18] 이호영, 손남호, "한국어 말토막 억양 패턴의 인지", *한글*, 277권, 2007(출판중).
- [19] 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", *Interspeech 2007*.
- [20] 이호영, "한국어의 억양체계", *언어학*, 13권, pp. 129-151, 1990.
- [21] S. -A. Jun, *The Phonetics and Phonology of Korean Prosody*, PhD. dissertation, Ohio State University, 1993.
- [22] S. -A. Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing*, Oxford:

Oxford University Press, 2005.

접수일자: 2007년 8월 15일

게재결정: 2007년 9월 14일

▶ 김선희(Sunhee Kim) : 교신저자

주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 인문관 5동 313호 인문정보연구소

소속: 서울대학교 인문학연구원

전화: 02) 880-7735

E-mail: sunhkim@snu.ac.kr

▶ 유현지(Hyunji Yoo)

주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 언어학과

소속: 서울대학교 언어학과

전화: 02) 880-6164

E-mail: yhjciel@hotmail.com

▶ 홍혜진(Hyejin Hong)

주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 언어학과

소속: 서울대학교 언어학과

전화: 02) 880-9039

E-mail: souble1@snu.ac.kr

▶ 이호영(Ho-Young Lee)

주소: 151-745 서울특별시 관악구 관악로 599 서울대학교 언어학과

소속: 서울대학교 언어학과

전화: 02) 880-6166

E-mail: hylee@snu.ac.kr