

Momel을 이용한 한국어의 억양 패턴 분석¹⁾

김 선 희*, 유 현 지**

*서울대학교 인문정보연구소, ** 서울대학교 언어학과

An Analysis on Korean Intonation Patterns Using Momel

Sunhee Kim*, Hyunji Yoo**

*Center for Humanities and Information, Seoul National University

**Department of Linguistics, Seoul National University

E-mail : *sunhkim@snu.ac.kr, **yhjciel@hotmail.com

Abstract

This paper aims to propose an intonation labeling method using Momel and to present results of analyzing a speech corpus consisting of 80 passages pronounced by 4 speakers (2 male and 2 female) using the proposed method. The results show that Momel works well enough to derive meaningful pitch targets, which could be labeled with H and L tones. On the other hand, the results of the analysis of Korean speech corpus correspond to earlier work.

I. 서론

문장의 발화에 있어 운율 정보는 주요한 언어학적 이론의 연구 주제일 뿐만 아니라, 음성인식과 음성합성에도 절대적으로 필요로 하는 정보이다. 오랫동안 이에 대한 연구들이 수행되었는데, 그 가운데 초기에 영어 문장의 운율정보를 기술하기 위하여 제안되었던 ToBI(Tone and Break Indices)가 현재 세계적으로 억양 분석 방법론과 억양 모델링 분야에 있어서 거의 표준 시스템으로 인정되고 있다[1, 2]. ToBI는 음성학적 정보들을 이용한 합성음의 자연성 향상과 자연 발화에 대한 인식을 향상을 위한 연구로서 음성학자들이 음성

자료의 데이터베이스를 구축하는 공통적인 규약을 논의한 결과물이라고 할 수 있다. 그러나 ToBI에 대한 음성 데이터의 분석의 효율성에 대해서는 훈련 받은 전문가에 의해서만이 가능하고 객관적인 기준이 제시되지 않아서 전사자간의 불일치 문제 등의 논란이 있어왔다[1, 3].

본 논문은 프랑스에서 개발된 자동 피치 추출 알고리즘인 Momel[4, 5, 6]을 이용하여 한국어의 억양 패턴을 추출하는 방법을 제안하고, 제안한 방법에 의하여 한국어 낭독체 문장의 억양 패턴을 분석하여 보고자 한다. Momel은 자동으로 억양 변곡점을 추출하고 그 추출된 피치 타겟을 이용하여 그 패턴을 일관성 있게 레이블링 할 수 있어, ToBI 레이블링의 문제점으로 지적된 훈련 받은 전문가에 의한 작업의 필요성을 완화시키고 전사자간의 일치도를 높일 수 있다. 이러한 연구는 억양 레이블링 및 AP/IP 경계 추출의 자동화에 기여하여 할 것이다.

II. 연구 방법

2.1 Momel-Intsint

Momel은 [4]에서 제안한 기본주파수(F0)를 모델링하는 방법인데, 이후 [5]에서는 [4]에서 방법을 발전시켜 자동화하기에 이르렀다. 음성 파일을 입력으로 할 때 Momel에 의하여 기본주파수의 거시적인 부분을 정의하는 피치 타겟의 연쇄를 얻을 수 있다. 자동화한

1) 이 논문은 2007년도 정부재원(과학기술부 한·불 과학기술협력기금사업)으로 국제과학기술협력재단의 지원을 받아 연구되었음.

Momel의 성능은 [7]에서 영어, 프랑스어, 독일어, 스페인어, 불어 등의 언어들을 대상으로 하여 평가되었는데, 이 연구에 의하면 Momel은 평균 95.5%로서 높은 성능을 보인다고 한다.

Momel에 의하여 추출된 피치 타겟에 대하여 [5, 8]은 억양 레이블링 시스템인 Intsint (INternational Transcription System for INTonation)를 제안하였다. 이는 운율적으로 다른 특성을 보이는 두 언어인 프랑스어와 영어의 억양 패턴을 동시에 나타낼 수 있는 방법으로 고안된 것으로 억양의 IPA(International Phonetic Alphabet)에 해당한다고 한다. Intsint는 억양 패턴의 기술에 있어서 최소한의 대조 이루는 부분을 기준으로 기술한 것으로 Momel에 의하여 도출된 각각의 피치 타겟들에 대하여 절대값(Top, Bottom, Mid)과 상대값(Higher, Lower, Same, Upstepped, Downstepped)을 이용하여 그 레이블을 결정하게 된다. Intsint는 [8]에서 9개 언어에 적용하였고, 실제 음성 합성에 이용되는 경우에 매우 효과적이라고 보고되었다. 하지만, [8]에서도 지적한 바와 같이 음성학적인 레이블링 시스템으로서 억양 패턴을 기술하는 데에는 적합하지 않다.

2.2 음성데이터

본 연구에서 사용하는 음성데이터는 세계적인 표준 (Eurom1)[9]에 따른 한국어의 낭독체 음성 데이터로서, 서울 토박이 20대 남녀 각 5명으로 총 10명이 녹음에 참여하였다. 발성 목록은 Eurom1 텍스트(화자당 40문단 총 166문장)이고, 총 녹음 시간 약 2시간이다. 녹음은 Marantz PMD 670 (녹음기)와 Shure SM-58 (마이크)를 이용하여 무향실에서 이루어졌다.

본 논문에서는 전체 데이터 가운데 4명(남녀 각 2명)이 각각 20개의 문단을 발화한 총 80개 문단, 320개 문장을 그 분석 대상으로 하였다. 녹음된 음성 데이터는 MBORLA를 이용하여 SAMPA (Speech Assessment Methods Phonetic Alphabet)[10]로 음소 레이블링을 수행되었다.

2.3 Momel을 이용한 억양 레이블링

Momel 결과를 이용하여 억양 레이블링하는 방법은 다음과 같다.

- 1) 음성 파일의 Momel 분석
- 2) Momel로 검출된 피치 타겟 수정
- 3) AP, IP 경계 표시
- 4) 성조 레이블링: Momel에서 검출된 F0 값을 기준으로 H/L로 레이블링

4.1) AP 레이블링

- 가장 높은 타겟을 H로 하여 상대적으로 표시

4.2) 수평조의 판단: 두 타겟의 차이가 $\pm 10\text{ms}$ 인 경우

4.3) IP 경계 성조 레이블링

- 마지막 음절의 시작 지점을 확인해서 마지막 음절에 표시된 피치 타겟을 레이블링

III. 실험 결과

분석 대상이 되는 80개의 발화를 위에서 제시한 방법에 따라 AP와 IP경계와 성조 패턴이 레이블링한 결과, 최종적으로 1729개의 AP와 718개의 IP로 총 2447개의 운율 단위로 레이블링 하였다. 본 논문의 대상이 된 AP의 평균 지속시간은 485.04ms, 평균 음절수 3.40이었고, IP의 평균 지속시간은 764.23ms, 평균 음절수는 4.67이었다.

3.1 오류 유형

위 2.3에서 제시한 억양 레이블링 방법을 이용하여 레이블링 하는 경우에 있어서 전문가에 의한 수동 확인이 필요한 부분은 2) MOMEL로 검출된 피치 타겟을 수정과 4) 억양 레이블링의 부분이다. 3) AP/IP 경계 표시는 현재는 자동화하지 않고 모두 수동으로 작업하여 자동화의 문제는 지적할 수 없다.

MOMEL로 검출된 피치 타겟의 수정이 필요한 경우와 그 수정 내용은 다음과 같다.

- 1) 문장 경계 밖에서 target을 잡는다. → 경계 밖 타겟을 지우고 마지막 모음에 해당 타겟을 표시함.
- 2) 문장 시작과 끝 부분에 필요없는 타겟(M)을 하나씩 잡는다. → 지운다.
- 3) LH pattern이 반복 될 때, 반복되는 tone의 중간 값 하나로 잡는다. → 중간값의 위치를 이동한 다음 하나의 타겟을 추가함.
- 4) 연속한 둘 이상 음절이 같은 높이일 때, 하나만 잡아버려서, 앞뒤가 각각 rising, falling tone이 된다. → 두 음절 중간 위치의 타겟을 하나의 음절로 이동시키고 나머지 음절에 같은 높이의 타겟을 추가함.

이와 같은 방법으로 수정을 한 경우에 여성들은 IP 경계 성조를 제외하고 거의 문제가 없었는데, 남성 화자들의 경우는 좀 더 많은 수정이 필요하였다.

성조 레이블링에 있어서의 문제점은 Momel에 의하여 검출된 피치 타겟을 H와 L로 표기하는 단계에서 상대적인 F0 값에 의하여 레이블링 하는 경우에 수정이 필요한 경우에 해당한다. 그 구체적인 경우와 수정

한 예는 다음과 같다.

- 1) 하나의 피치 타겟만 나타나는 경우 음성을 확인하여 HH, LL, HL, LH로 표기함.
- 2) AP, IP의 경계 전후에 타겟이 밀린 경우 관찰되는 오류들을 수정함.
- 3) AP, IP 경계 내에서 피치 타겟이 나타나지 않는 경우 --> 최소한의 타겟을 추가함.

이와 같은 방법으로 수정하여 총 2447개의 운을 단위 가운데 296개를 수정하여 총 수정률 12.10%를 보였다.

3.2 AP 패턴

다음 표는 Momel을 이용하여 추출된 억양 패턴을 [11] 및 K-ToBI[12]와 비교한 것이다. K-ToBI에서 제안한 14개의 AP 패턴과 비교해 보면, 14개의 패턴 외에도 H, L, LHHL, LHHLH가 관찰되었다. 이 가운데, LHHL와 LHHLH는 각각 LHL, LHLH의 변이형으로서 빈도가 아주 낮음을 볼 수 있다. 거기에 비하여 Ha와 La는 꽤 높은 빈도를 보이며 나타났는데, 이들은 1음절어가 아닌 경우에도 그 AP 내부에서 특별한 변곡점이 추출되지 않고 하나의 톤으로 실현된 예들이라고 할 수 있다.

Lee[11]	Jun[12]	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

표 1. Momel을 이용하여 추출한 AP 억양 패턴

위 표에서 보는 바와 같이, 전체적으로 오름조(Rising, Fall-Rising, Rise-Fall-Rising: 62.69%)가 내림조(Falling, Rise-Falling: 22.80%)에 비하여 많이 나타나고, 수평성조 가운데에서는 H(11.45%)가 L(3.07%)에 비하여 많이 나타남을 볼 수 있었다.

3.3 IP 경계 성조

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

문장 중간 IP	빈도	%	문말 IP	빈도	%
H%	138	36.13	H%	52	15.48
HL%	176	46.07	HL%	198	58.93
L%	55	14.4	L%	78	23.21
LH%	6	1.57	LH%	3	0.89
LHL%	7	1.83	LHL%	5	1.49

표 2. Momel을 이용하여 추출한 IP 경계 성조

문장 중간이나 문말에서 모두 HL%이 가장 많이 나타나는 것을 볼 수 있었다. 문말 IP에서의 H%는 의문문에서 일반적으로 나타나는 것으로 문장 중간에서 나타나는 H%와는 다른 기능을 갖는데, H%는 문장 중간의 위치에서 문말에 비하여 두 배 정도 자주 나타나는 것을 볼 수 있다. 반대로 L%의 경우는 문말에 비하여 문장 중간에서는 절반 정도의 빈도로 나타나는 것을 볼 수 있다. LH%, LHL%의 경우는 문장에서의 위치에 관계없이 2%이하의 빈도를 보여 거의 나타나지 않는다고 할 수 있다.

IV. 토의 및 결론

본 논문은 자동 피치 추출 알고리즘인 Momel을 이용하여 한국어의 억양 패턴을 추출하는 방법을 제시하고 그 방법에 따라 표준 한국어 낭독체 음성 데이터를 분석하여 보았다.

Momel을 이용하여 레이블링 하는 방법은 1) 음성 파일의 Momel 분석, 2) MOMEL로 검출된 피치 타겟 수정, 3) AP, IP 경계 표시, 4) 성조 레이블링의 4단계로 진행되는데, 첫 번째 단계는 단순히 Momel을 실행하는 단계이고, 세 번째 단계는 전체 단계를 자동화하려고 하는 경우에 Momel의 이용과는 상관없이 따로 연구되어야 하는 부분이므로 본 논문에서는 논의에서 제외하였다.

본 논문에서는 4단계의 레이블링 방법 가운데 두 번

참고문헌

재와 네 번째 단계에서의 문제점을 고찰하여 그 타당성을 검토하였다. Momel로 검출된 피치 타겟의 수정이 필요한 경우는 크게 두 가지로 유형화할 수 있는데, 첫 번째의 경우(1)번과 2)번의 경우)는 모두 문장 경계에서 나타나는 오류들로서 Momel 알고리즘을 개선하여 그 성능을 향상시킬 수 있으며, 두 번째의 경우(3)과 4)의 경우)의 성능향상에 대해서는 좀 더 검토가 필요하다. 성조 레이블링에 있어서의 문제점은 Momel에 의하여 검출된 피치 타겟을 H와 L로 표기하는 단계에서 상대적인 F0 값에 의하여 레이블링 하는 경우에 수정이 필요한 경우로서 3가지가 지적되었는데, 이 가운데 Momel 알고리즘의 개선에 의하여 쉽게 성능향상을 가져올 수 있는 부분(2)번의 경우)과 수작업에 의해서만 작업이 가능한 경우(1)과 3)의 경우)로 나뉘었다. 즉, 본 논문에서 제시한 방법론에 의하여 억양 레이블링을 하는 경우에 Momel 알고리즘의 개선에 의하여 많은 부분이 수정될 수 있으나 수작업이 필요한 부분은 남게 되는데, 이 부분에 대한 중요도에 대한 평가는 다시 향후 다시 연구가 필요하다.

본 논문에서 제안한 방법론에 의하여 표준 한국어 낭독체 음성 데이터인 Eurom Korean 가운데 4명(남녀 각 2명)이 각각 20개의 문단을 발화한 총 80개의 발화를 그 분석한 결과, 1729개의 AP와 718개의 IP로 총 2447개의 운율 단위로 레이블링 되었다. 본 논문의 대상이 된 AP의 평균 지속시간은 485.04ms, 평균 음절수 3.40이었고, IP의 평균 지속시간은 764.23ms, 평균 음절수 4.67이었다. K-ToBI에서 제안한 14개의 AP 패턴과 비교해 보면, 여기에서는 14개의 패턴외에도 H, L, LHHL, LHHLH가 관찰되었다. 이 가운데, LHHL와 LHHLH는 아주 낮은 빈도를 보이는데 이들은 각각 LHL, LHLH의 변이형이라고 할 수 있다. 반면에 H와 L은 꽤 높은 빈도를 보이며 나타났는데, 이들은 AP 내부에서 특별한 변곡점이 추출되지 않고 하나의 수평성조로 실현된 예라고 할 수 있다. IP 경계 성조로는 [12]와는 달리 H%, HL%, L%, LH%, LHL%의 다섯 가지만 관찰되었고, 그 가운데 LH%, LHL%의 경우는 문장에서의 위치에 관계없이 2%이하의 낮은 빈도를 보였다. 이러한 결과는 기존의 K-ToBI에 의한 한국어 분석 결과와 거의 일치하는 것으로 [1, 12, 13], Momel을 이용하여 자동 추출된 피치 타겟을 이용하여 음성 데이터를 분석하는 경우에 매우 유용함을 보인다고 할 수 있다. 이와 같이 본 연구는 Momel을 이용한 억양 레이블링의 자동화, AP/IP 경계 추출의 자동화의 필요성과 그 유효성을 보였다.

- [1] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. 1992. ToBI: a standard for labelling English prosody. *Proc. of ICSLP*. Banff: Canada. 867-870.
- [2] Jun, Sun-Ah (Ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press, 2005.
- [3] Wightman, C., ToBI Or Not ToBI?. *Proc. of the International Conference on Speech Prosody*. Aix-en-Provence: France. 25-29, 2002.
- [4] Hirst, D.J., Espesser, R., Automatic modelling of fundamental frequency curves using a quadratic spline function. *Travaux de l'Institut de Phonétique d'Aix* 15. 71-85, 1993.
- [5] Hirst, D.J., Di Cristo, A., Espesser, R., Levels of representation and levels of analysis for the description of intonation systems. In: Horne, M. (Ed.) *Prosody: Theory and Experiment*. Dordrecht: Kluwer. 51-87, 2000.
- [6] <http://aune.lpl.univ-aix.fr/~auran/english/ressources.html>
- [7] Campione, E., Etiquetage semi-automatique de l'intonation dans les corpus oraux : algorithmes et methodology. Thèse de doctorat. Aix-en-Provence: Universit de Provence, 2001.
- [8] Hirst, D. & Di Cristo, A., (eds), *Intonation Systems : A Survey of Twenty Languages*, Cambridge University Press, Cambridge, 1998.
- [9] <http://www.phon.ucl.ac.uk/shop/eurom1.php>
- [10] <http://www.phon.ucl.ac.uk/home/sampa/>
- [11] Lee, H.Y., An acoustic phonetic study of Korean nuclear tones (in Korean). *Malsori* 38. The Phonetic Society of Korea. 25-39, 1999.
- [12] Jun, S.-A. 2000. K-ToBI (Korean ToBI) labeling conventions: Version 3.1. *UCLA Working Papers in Phonetics* 99. 149-173.
- [13] 이숙향, 김종진, 대용량 데이터베이스를 이용한 한국어 운율 특성에 관한 연구, *한국음향학회지* 제24권 2호, 한국음향학회, 117-126, 2005.