

발화 말 억양구 경계 성조의 기울기 정보를 이용한 음성의 감정 정보 분류

박 미 경
고려대학교 국어국문학과

Acoustical Analysis of Emotional Speech based on the Declinations of F0 Contours from Utterance Final IP Boundary Tones

Mikyoung Park
Department of Korean Literature and Linguistics, Korea University
E-mail: miky@korea.ac.kr

Abstract

The purpose of this paper is to testify the possibility of using declinations of F0 contours from utterance final IP boundary tones in order to classify emotions such as normal, happiness, sadness and anger. The ranges of declination are widen when emotional speech is realized. Level boundary tone is realized when the declination values are between -5 and 5. In case of more than ± 5 value, boundary tone is rising or falling. Depending on the last declination value, happy utterance can be distinguished with other emotional statuses. When speakers uttered in sad mood, the range of declination is wider than normal mood, and there can be the influence of phonation types such as breathy or creaky.

I. 서론

본 논문은 억양구말 경계 성조에 나타나는 기본주파수 값에서 기울기를 측정하고, 이 기울기를 이용하여 음성에 나타난 감정 정보를 효과적으로 분류할 수 있는 가능성을 타진하는 데 있다. 경계 성조, 특히 발화 말에 위치한 억양구 경계 성조는 다양한 기능을 수행한다. 문장 유형을 판단할 수 있는 기준이 될 뿐만 아니라, 화자의 감정이나 태도를 표현하고, 화용론적 의미 기능을 전달하기도 한다.[1]

특히 감정 전달에 있어 발화 말 억양구 경계 성조는 매우 큰 역할을 한다.[2][3][4] 특히 [2]에서는 감정 발화에서 전체 발화와 발화 마지막 강세구의 기본주파수 분포도를 비교한 결과를 제시하였다. 이 연구에서는 전체 발화의 기본주파수 분포도와 발화 마지막 강세구의 기본주파수 분포는 그 양상이 매우 비슷함을 확인할 수 있었다.

본고에서는 기존의 감정 정보 분류에 있어 중요한 역할을 담당할 것으로 기대되는 발화 말 억양구 경계 성조가 어떻게 음성학적으로 실현되는지를 살펴볼 것이다. 각 경계 성조의 실현이 해당 성조 내 기본주파수 값을 어느 정도의 시간 내에 실현하였는지를 밝혀 경계 성조의 음향음성학적 분석 근거에 대한 가능성을 제시해보고자 한다. 이를 위해 기본주파수 값을 시간 변화에 따라 얼마나 변화하였는지에 대해 기울기 추출 방법을 사용하여 나타내고자 한다. 기울기 정보는 시간 변화에 따른 기본주파수의 변화량을 나타낼 수 있는 값이므로 경계 성조 실현의 객관적인 자료를 제시할 수 있을 것으로 기대한다. 그리고 이 방법이 감정 정보 분류에 도움이 될 가능성이 있는지에 대해 검토할 것이다.

II. 감정 음성 DB

감정 발화의 자연스러운 감정 표현을 위해 실험 문장은 모두 대화체 문장으로 구성하였다. 대화가 이루어지는 문맥과 상황을 한정하기 위해 대화 진행 배경과 화자의 관계 등을 구체적으로 제시하였다. 보조 지문의 경우 감정을 잘 표현하고, 문맥 해석을 일정한

방향으로 할 수 있도록 적절히 수정하여 제시하였다. 자료 측정을 용이하게 하기 위해 모든 문장의 종결어미는 '-요'로 통일하였다. 대상 문장 유형은 평서문 10문장으로 한정하였다.

표 1. 실험 참가 성우 목록

파일명	화자명	성별	나이
f01	KSH	여	22
f02	ACH	여	31
f03	KIJ	여	27

본 실험에 참가한 성우들은 성우 전문 녹음실인 '보이스위즈(voicewiz)'를 통해 경력이 3년 이상인 전문 성우들로 구성되었으며, 모두 표준어를 구사였다. 녹음 시 실험 문장들은 평상, 기쁨, 슬픔, 화남 등의 네 가지 감정으로 각 2회씩 반복하여 녹음하였다. 이 때 설정 감정이 극명하게 표현하도록 주지하였고, 해당 감정이 끝까지 일정하게 유지되도록 주의시켰다.

녹음에 사용된 장비는 믹서 MACKIE 1604-VLPRO, 오디오 인터페이스 ESI WaMi Rack 192X, 컴프레서 JOEMEEK VC3Q, 마이크 RODE NT2가 사용되었으며, CoolEdit Pro 2.1에서 22050 byte, 16bit로 디지털화되었다.

이렇게 수집된 발화의 수는 총 240개이다.(화자 3명×4감정×평서문 10문장×반복 2회) 이들 중에서 말소리 외에 녹음된 웃음소리나 울음소리가 나온 발화, 기식성이 강하게 나타난 음절, 약하게 발음된 음절은 기본주파수가 제대로 추출되지 않아 이들 발화 13개를 제외하고, 본 실험에서는 총 227개 발화를 대상으로 하였다.

III. 경계 성조의 기울기 추출

본고에서는 억양구 경계 성조의 기본주파수 추출을 위해 Wavesurfer 1.8.5를 이용하였다. Wavesurfer를 이용하여 나타낸 기본주파수 곡선을 보면 다음의 그림 1과 같다.

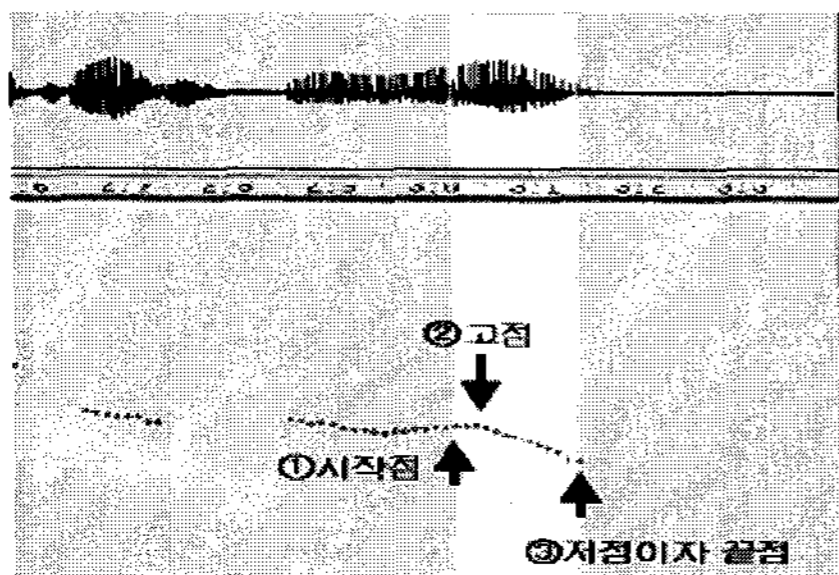


그림 1. 화자 f01의 평서문 02-1 중에서 '해줬어요'의 파형(상)과 기본주파수 곡선(하)

그림 1에서 발화 말 경계 억양이 실현된 '-요'는 두 실선 사이에 해당된다. 이 곡선에서 나타난 기본주파수 값은 Wavesurfer에서 매 10ms마다 값을 저장하여 *.f0라는 파일에 저장된다. 이 값을 가진 f0 파일에서 '-요'의 시작점 ①의 시간과 마지막 부분의 점 ③까지의 시간을 입력하면 시작점①~고점②와 고점②~저점③에 해당하는 부분까지의 기울기를 추출하였다. 여기서 말하는 기울기란 기본주파수 값의 변화량을 시간 정보의 변화량으로 나눈 값을 말한다. 이를 공식으로 나타내면 다음과 같다.

$$\text{기울기} = \frac{\frac{\text{고점/저점의 } F_0 - \text{시작점의 } F_0}{\text{고점/저점시간} - \text{시작점시간}}}{\frac{\text{다음고점/저점의 } F_0 - \text{이전고점/저점의 } F_0}{\text{다음고점/저점의시간} - \text{이전고점/저점의시간}}}$$

위 그림에서는 기울기가 두 개 나온다. ①시작점~②고점 사이의 기울기와 ②고점~③끝점 사이의 기울기가 그것이다. 이것을 표로 정리하면 다음과 같다.

표 2. 화자f01의 평서문02-1의 발화 말 경계 억양의 기울기(시간은 10ms단위)

시작점 ①시간	끝점 ③시간	고점② 저점③ 시간	고점② 저점③ F0값	④F0값 차 (②-①) 또는 ③-②)	⑤시간 차 (②-①) 또는 ③-②)	기울기 (=④/⑤)
302	313	303	168.24 5	2.213	1	2.213
302	313	313	142.05 8	-26.186	10	-2.618

그림 1과 표 2에서 보면 ①시작점~②고점 사이의 기울기는 상승하고 있기 때문에 +값을 가지고, ②고점~③저점이자 끝점 사이에는 기울기가 하강하고 있으므로 -값을 가진다. K-ToBI에 의해 이 부분을 기술하면 L%값으로 기술된다. 하지만 여기서의 L%는 단순한 하강이 아닌, 약간의 상승을 동반하는 하강이고, 이 하강의 정도는 시간 변화에 따라 기본주파수가 얼마나 변하였는지를 수치화하여 나타낼 수 있다.²⁾ 여기서 예측되지 않은 기본주파수의 오류값은 모두 개별적으로 검토하여 기울기를 구할 때 반영되도록 하였다.

기울기 정보는 평서문에 자주 나타나는 L% 혹은 HL% 내에서도 다양한 값들을 가질 수 있다. 이는 상

2) 물론 이 기울기들이 모두 청각적 인상에 영향을 주는 것은 아니다. 하지만 본고는 기존의 경계 성조에 더해 기본주파수 곡선에서의 기울기가 감정 정보 분류에 얼마나 효율적인가를 논하는 것이 주목적이므로, 어느 정도의 기본주파수 변화가 높낮이의 변화에 영향을 주는지에 대한 논의는 여기서 다루지 않을 것이다.

대적인 높낮이를 바탕으로 하여 경계 성조를 주는 기존의 운율 기술 방법에 더해, 실제 자료에서 나타나는 다양한 기본주파수의 변화를 더 상세히 설명할 수 있는 보완 장치가 될 것이다. 또한 이러한 수치들을 활용한다면 감정 표현에서 나타나는 발화 말 경계 성조의 미묘한 변화도 더욱 다양하게 음향음성학적으로 표현할 수 있을 것이다.

IV. 실험 결과 및 논의

3장에서 제안한 방법에 따라 화자 f01의 발화 중 평서문02-1번의 발화 말 억양구 경계 성조에서 추출한 기본주파수 곡선의 기울기를 도식화하면 다음과 같다.

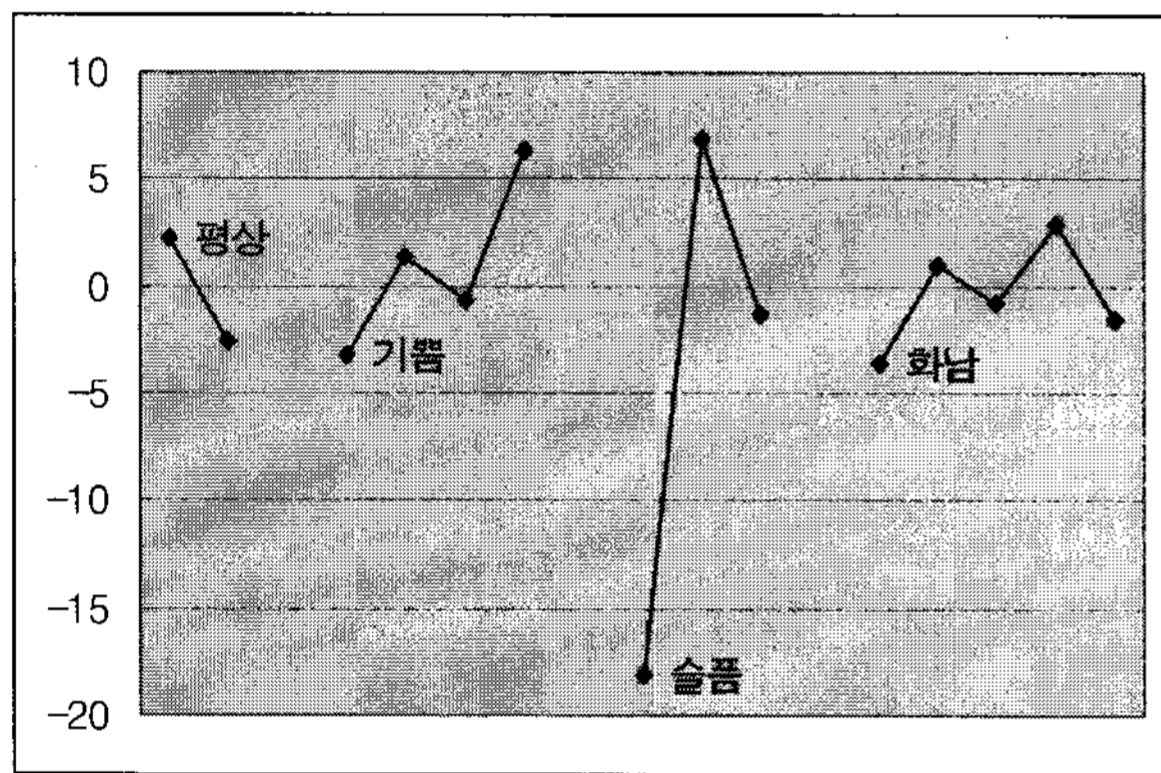


그림 2. 화자f01의 평서문02-1 발화 말 억양구 경계 성조 내 기울기(세로축은 기울기, 가로축은 감정)

정상 발화에서 기울기 분포를 보면 대부분 -값을 가진 기울기로 끝나는 것을 확인할 수 있다. 이것은 L%로 실현되는 경계 성조를 나타내는 것과 일맥상통하는 것이다. 이들 중 +값을 가지는 기울기로 끝나는 경우도 있으나 이러한 경우는 대개 0에 가까운 값을 가진다. 0에 가까울수록 기울기의 차이가 많이 나지 않는다는 뜻이며, 이것은 [1]의 수평조에 해당하는 부분이다. 이 때 실현된 기본주파수 값은 $\pm 150\text{Hz}$ 이므로 이것은 낮은 수평조에 해당될 것이다. 수평조를 나타내는 기울기들은 대략 -5~5 사이의 값을 갖는다. 즉, 수평조는 기울기가 -5~5사이, 상승이나 하강조는 그 이상의 범위를 갖는 것으로 풀이된다.

화남의 경우는 예상과 달리 평상과 그 범위가 많이 차이하지 않았다. 이는 f01이 화남을 표현할 때 낮은 기본주파수로 실현한다는 의미가 아니다. 오히려 기본주파수 값은 200Hz대의 범위에서 많이 나타났다. 하지만 기울기가 크지 않은 까닭은 평상과 마찬가지로 수평조를 유지하는 경우가 많이 때문이다. 그러므로 기울기로만 본다면 평상과 큰 차이가 없어 보이지만 이 경우 실현된 기본주파수 값에서 차이가 나타나므로

이들의 구분은 큰 어려움이 없을 것으로 보인다. 또한 \pm 값의 교차가 평상에 비해 슬픔에서 더 많이 나타난다. 물론 그 범위가 -5~5 사이이긴 하지만 기본주파수의 미세한 변화가 평상에 비해 더 많이 나타난다는 것을 알 수 있다.

기쁨의 경우는 다른 감정과 달리 기울기가 +값을 가지는 기울기로 끝나는 경우를 많이 볼 수 있다. 이는 기쁨의 발화 말 경계 성조가 LH%로 실현되는 경향이 많은 것과 관계가 있는 것으로 보인다.³⁾ 즉, 상승조로 끝난 경우는 기본주파수 실현도 또한 평상과 비교하여 그 범위가 넓어진다. 평서문의 경우 보통은 L%나 HL%로 끝나는 것이 보통이지만 기쁨의 경우 상승조를 동반하는 경우를 많이 볼 수 있었다.

슬픔의 경우는 울음소리가 섞이거나 기본주파수가 추출되지 않아 측정이 곤란한 경우가 많이 나타났다. 슬픔을 표현하는 음성의 경우 기식음이 많이 섞이거나 울음소리가 많이 섞이는 경우가 대부분이었고, 그렇지 않은 경우는 목소리의 떨림이 나타나는 경우가 있었다. 기본주파수 기울기의 절대값이 커지는 이유는 바로 이 떨림 때문으로 보인다. 불안정한 떨림이나 기식음 등의 영향은 발생유형과 관계있다는 것을 알 수 있다.

모든 화자의 발화 말 경계 성조에서 나타난 기울기의 범위를 정리해보면 다음과 같다.

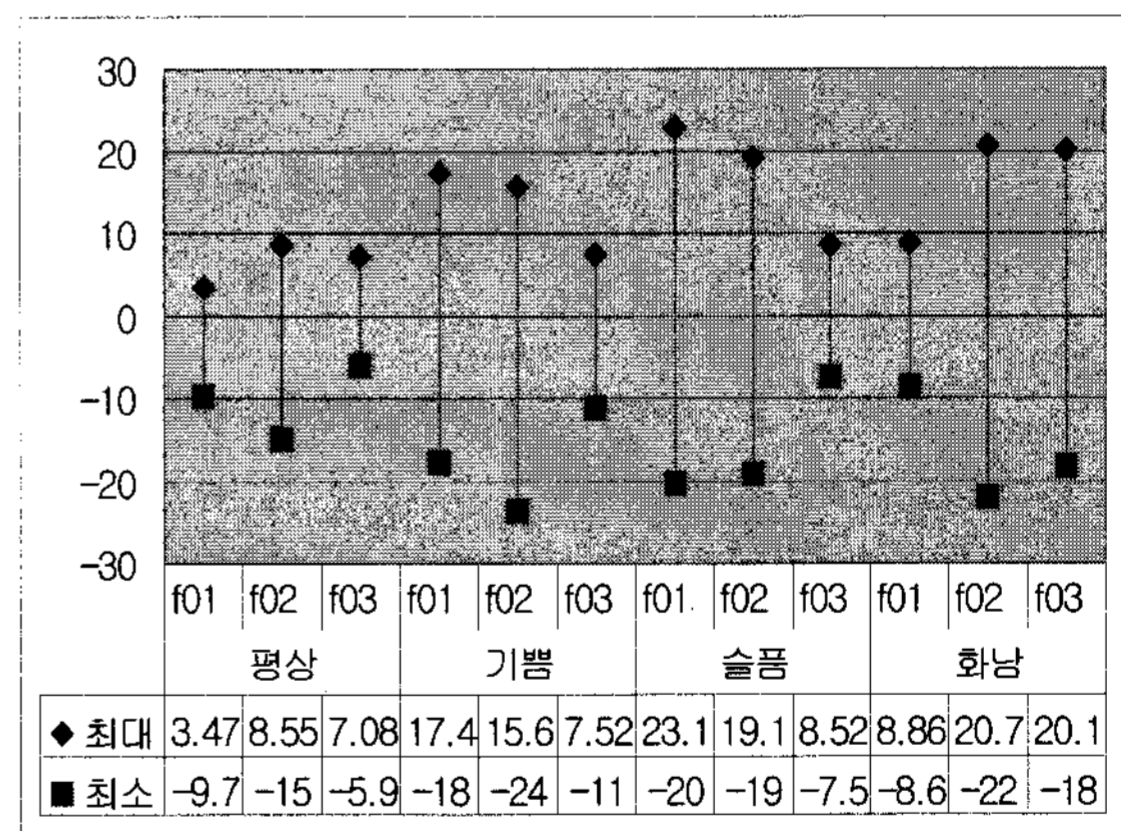


그림 3. 감정별 기울기의 최대, 최소(가로축은 기울기, 세로축은 감정)

모든 화자에서 공통적으로 나타나는 점은 평상에 비해 기쁨, 슬픔, 화남에서 기울기의 범위가 넓어진다는 점이다. 이것은 감정이 실린 발화의 기본주파수 변화가 더 많이 일어난다는 점이다. 슬픔의 경우 기본주파수의 변화가 일어나는 요인이 음의 높낮이에만 국한되지 않고 발생유형에 영향을 받는다는 점을 앞에서 지적한 바 있다. 그러므로 감정 발화에 대한 연구에서

3) 화자 f01은 기쁨 발화 20회 중 12회가, 화자 f02는 기쁨 발화 20회 중 14회가 LH%로 실현되었다.

앞으로는 발성유형에 대한 연구가 구체적으로 이루어져야 할 것이다.

화자에 따라 기본주파수의 기울기가 감정별로 달리 나타남을 알 수 있다. f01의 경우 기쁨과 슬픔의 기울기 범위가 넓은데 비해 화남의 기울기 범위는 평상과 비교해 큰 차이는 보이지 않는다. 하지만 f03의 경우는 이와 반대로 기쁨과 슬픔의 기울기보다 화남의 기울기 변화 범위가 더 넓다. f02의 경우는 다른 두 화자에 비해 평상의 기울기 범위도 넓은 뿐만 아니라 모든 감정에 있어 그 변화의 폭이 넓다. 이것은 개인별로 감정 표현에 있어 음의 높낮이 변화에 차이가 있음을 보여준다. 화자에 따라 사용하는 기본주파수 대역에 차이가 있으므로 절대적인 값의 비교는 어려울 것으로 생각한다. 이를 극복하기 위해 기본주파수의 정규화에 대한 연구가 더욱 필요하리라 생각한다.

또한 단순한 기울기의 범위만 고려해서는 감정 정보 분류에 판단이 어려울 것으로 생각한다. 기울기가 어떤 값으로 끝나고 있는지, 발화 말 경계 성조가 실현된 기본주파수 값이 어느 범위에 있는지 등을 종합적으로 고려하여야 한다는 것이다. 하지만 기울기의 범위가 중요한 것은 기울기의 범위가 평상에 비해 넓어진다는 점이고, 이것은 감정 표현이 실린 문장에서의 기본주파수가 좀 더 동적으로 움직인다는 것을 나타낸다. 단순히 L% 혹은 HL%의 기술에서 벗어나 어느 정도의 값으로 얼마나 실현되는가에 대한 연구가 앞으로 충분히 이루어져야 할 것이다.

V. 정리 및 남은 문제

이상으로 기울기가 감정 정보 분류에 활용될 가능성이 있는지 검토해보았다. 기울기는 감정 정보가 실린 발화에서 그 범위가 모두 넓어진다는 것을 모든 화자에게서 발견할 수 있었다. 즉, 기본주파수의 역동적인 변화가 감정 발화에서 나타난다는 점을 알 수 있었다.

기울기 실현 중 특이할 만한 것은 -5~5 사이의 값을 가지는 경우 수평조로 실현된다는 점이었고, 이 수평조가 어느 주파수대에서 실현되는지에 따라 높은수평조 혹은 낮은수평조가 결정된다고 볼 수 있다.

슬픔의 경우, 평상보다 더 넓은 범위에서 기울기가 실현되고 있다는 점이 특이했다. 이것은 발성유형의 변화가 기본주파수에 영향을 미치기 때문이라고 볼 수 있다. 그러므로 앞으로 음성에 실린 감정 정보의 연구에 있어 반드시 발성 유형에 대한 고려가 필요할 것으로 본다.

발화 말 억양구 경계 성조에서 나타난 기본주파수의 기울기로 감정 정보를 분류하는데 가능성이 있음을 알 수 있다. 하지만 단순히 기울기 범위만을 활용할

것이 아니라, 맨 마지막에 실현된 기울기의 \pm 값, \pm 의 교차 횟수, 그리고 기본주파수의 실현이 어느 정도의 값에서 이루어지는지를 모두 면밀히 살펴보아야 할 것이다.

또한 지각에 영향을 미치는 기울기의 범위를 정확히 찾아내는 것이 필요하다. 이 점을 보완한다면 기울기로 발화 말 경계 성조의 움직임에 대해 객관적으로 수치화된 값을 활용하여 어느 정도의 기본주파수 변화가 일어났는지를 명쾌히 정리할 수 있을 것이다. 그리고 이러한 변화가 의문문이나 명령/청유문에서도 어떻게 실현되는지, 남자 화자의 경우는 어떻게 실현되는지에 대해서도 앞으로 더 연구해야 할 것이다.

참고문헌

- [1] 이호영, "국어 해석양의 음향음성학적 연구", 말소리 38호, pp.25-39, 1999.
- [2] 박미영, 박미경, "감정에 따른 음성의 기본주파수 실현 연구", 2005년 대한음성학회 가을 학술대회 발표논문집, pp.15-23, 2005.
- [3] 장인창, 이태승, 박미경, 김태수, 장동식, "발화 내 감정의 정밀한 인식을 위한 한국어 문미억양의 활용", 한국통신학회논문지, 30호, 2005.
- [4] T. Lee, M. Park, T. Kim, "Toward More Reliable Emotion Recognition of Vocal Sentences by Emphasizing Information of Korean Ending Boundary Tones", The 10th International Conference of Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, pp.304-313, 2005.
- [5] A. W. Black, "Predicting the Intonation of Discourse Segments from Examples in Dialogue Speech", In Y. Sagisaka, N. Campbell, N. Higuchi (Eds), Computing Prosody: Computational Models for Processing Spontaneous Speech, pp.117-128, Springer, 1996.