

## 억양의 근접복사 유형화를 이용한 감정음성의 음향분석

### An acoustical analysis of emotional speech using close-copy stylization of intonation curve

이 서 배<sup>1)</sup>

Yi, So Pae

#### ABSTRACT

A close-copy stylization of intonation curve was used for an acoustical analysis of emotional speech. For the analysis, 408 utterances of five emotions (happiness, anger, fear, neutral and sadness) were processed to extract acoustical feature values. The results show that certain pitch point features (pitch point movement time and pitch point distance within a sentence) and sentence level features (pitch range of a final pitch point, pitch range of a sentence and pitch slope of a sentence) are affected by emotions. Pitch point movement time, pitch point distance within a sentence and pitch slope of a sentence show no significant difference between male and female participants. The emotions with high arousal (happiness and anger) are consistently distinguished from the emotion with low arousal (sadness) in terms of these acoustical features. Emotions with higher arousal show steeper pitch slope of a sentence. They have steeper pitch slope at the end of a sentence. They also show wider pitch range of a sentence. The acoustical analysis in this study implies the possibility that the measurement of these acoustical features can be used to cluster and identify emotions of speech.

**Keywords:** emotional speech, pitch contour, pitch range, pitch slope, pitch distance, intonation curve

#### 1. 서론

억양의 특성을 통한 감정음성의 음향적 분석은 최근까지 활발하게 이루어져 왔다(Williams et al., 1972; Bachorowski et al., 1995; Pell et al., 2009). 특히, 사람의 감정이 피치의 절대적인 값과 같은 피치레벨보다는 억양곡선의 모양에 더 잘 나타난다는 연구(Bänziger et al., 2005; Rodero, 2011)와 피치의 움직임 정도가 다른 음향 자질들보다 화자의 감정과 관계가 깊다는 연구(Mozziconacci, 2002)는 주목할 만하다. 그러나 이러한 연구는 청취실험을 통해 억양곡선이 어떠한 감정으로 느껴지는지를 알아본 인지적 관점의 연구이거나 억양곡선의 음향적 연구이더라도 적은 양의 발화자료를 가지고 사람의 주관적 판단에 기초해 가변적으로 억양곡선의 모양을 단순화해서 음향값들을 뽑았기 때문에 일관성과 정확성에서 한계를 가질 수밖에 없었다.

억양곡선의 모양과 감정과의 관계를 음향적으로 연구하기 위해서는 억양의 모양을 표현할 수 있어야 하는데 국내에 소개된 이러한 시도에는 억양곡선의 모양에 따라 억양을 기술하려는 영국식 억양 이론(Lee, 1990)이나 억양의 상대적인 음높이 수준에 주목해서 억양을 기술한 억양음운론의 방법(Jun, 1993) 등이 있다. 그러나 이러한 방법들은 궁극적으로 주관적 판단에 기초하고 있어서 억양을 판단하는 기준이 연구자들마다 다를 수 있으므로 물리적으로 동일한 억양일지라도 다르게 기술 될 수 있다는 점에서 공통적인 한계를 지니고 있다. 그리고 이러한 표현들만 가지고 실제 음성신호의 억양을 복원해 내는 것에는 많은 제약이 따른다.

이런 맥락에서 ‘추상적, 상대적, 가변적’이라는 한계를 극복하기 위해 억양을 정규화하려는 최근의 시도는 주목할 만하다. 이러한 시도의 일환으로 억양의 근접복사 유형화(close-copy stylization)를 이용한 연구는(오재혁, 2014a; 오재혁, 2014b) 억양의 물리적인 실체에 많이 접근하였고 한국인 화자들이 자연스럽게 대화한 발화 자료에 적용한 결과 남녀 간, 화자 간 생리적인 차이(성대의 크기, 성도의 길이, 방패연골의 각도)에 기인한 기본 주파수와 피치변위의 변화에 관계없이 억양의 정규화가 가능할 수 있다는 것을 보여주었다.

1) 창원대학교, 영어영문학과 sopacyi@pusan.ac.kr

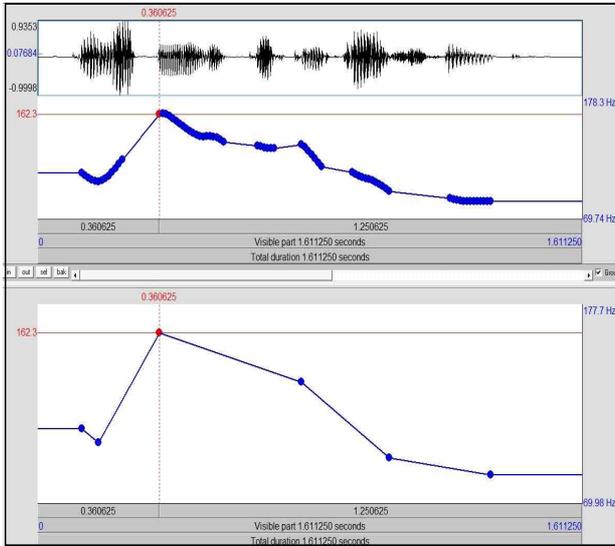


그림 1. 억양 근접복사 유형화  
Figure 1. Close-copy stylization of an intonation curve

<그림 1>은 본 연구에서 Praat(Boersma, 2001)을 사용해 구현한 억양 근접복사 유형화의 한 예이다. <그림 1> 상단에는 음성파형이 나타나 있고 이 음성에서 추출한 피치곡선이 바로 그 밑에 있으며 맨 아래에 억양 근접복사 유형화로 단순화된 피치 포인트들이 나타나 있다(여기서는 6개의 피치 포인트로 표시되었다). 이것은 지각적 등가성(perceptual equality)을 이용하여 피치의 미시적인 변동을 청각적 관점에서 단순화한 것이다. 원 억양곡선에서 최소한의 변곡점을 찾으면 많은 피치 값들을 대표성을 가지는 몇 개의 값으로 단순화 할 수 있게 된다. 즉, 위의 예에서 6개의 피치포인트만 가지고 음성합성을 통해 원래 음성으로 복원해도 피치곡선 단순화 이전의 원음성과 비교해서 억양의 차이가 느껴지지 않는 것이다.

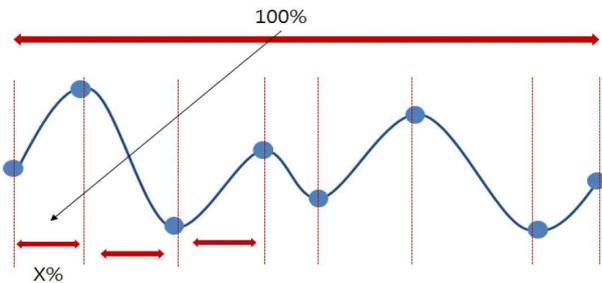


그림 2. 피치포인트 이동시간(X%)의 측정(오재혁, 2014a)  
Figure 2. Measurements of pitch point movement time (X%) (Oh, 2014a)

단순화된 억양의 유형으로부터 남녀 화자들 간의 차이를 정규화한 기존 연구(오재혁, 2014a; 오재혁, 2014b)에서는 <그림 2>에서와 같이 각 발화문장의 전체 발화시간을 100%로 했을 때 하나의 피치포인트가 다음 피치포인트의 위치까지 이동

한 시간인 피치포인트 이동시간(을 X%로 환산하고 <그림 3>에서와 같이 각 발화문장별 최소 피치포인트 값에서 최대 피치포인트 값까지의 변위폭을 100%로 했을 때 피치포인트들 간의 피치 값의 차이를 Y%로 환산하여 계산하였다.

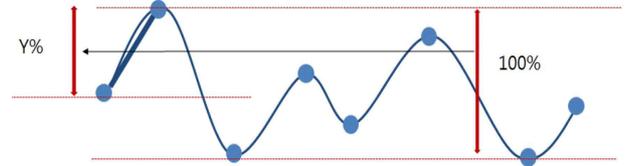


그림 3. 피치포인트 변위(Y%)의 측정(오재혁, 2014a)  
Figure 3. Measurements of pitch point range (Y%) (Oh, 2014a)

그리고 백분율로 환산된 값을 대상으로 두 피치포인트 간의 기울기를 측정한 것이 <그림 4>이다. 또한 환산된 값을 대상으로 <그림 5>와 같이 피치포인트 이동거리를 구하였다. 이렇게 추출한 억양곡선 측정방법을 가지고 자연스런 한국어 대화에서 얻은 발화들을 연구한 결과 남녀의 차이와 화자들 간의 차이 그리고 발화 맥락상의 여러 가지 가변적 요소들이 있었음에도 불구하고 정규화에 필요한 억양 자질들의 항상성이 발견되었다(오재혁, 2014a; 오재혁, 2014b).

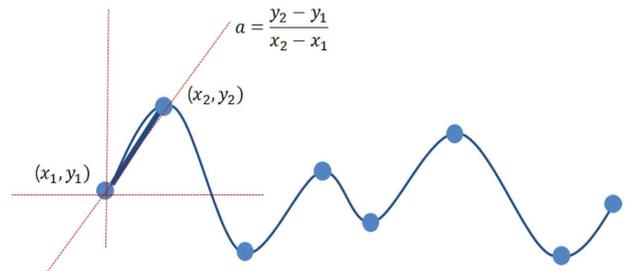


그림 4. 피치포인트들 간의 기울기 측정(오재혁, 2014a)  
Figure 4. Measurement of the slope between pitch points (Oh, 2014a)

본 연구는 이러한 억양곡선 정규화기법을 감정음성의 음향적 분석에 적합하도록 보완해서 적용하고 분석하였다. 감정음성 말뭉치(Corpus)자료를 대상으로 억양근접복사 유형화 기법(Close-copy stylization)을 사용해서 일관성 있게 자동으로 값을 추출했다는 점에서 적은 양의 감정음성에서 주관적 판단에 기반한 수작업으로 뽑은 측정값을 사용한 기존연구(Bänziger et al., 2005)와 차별성을 보인다.

2) 원문에서 ‘음높이 움직임의 이동시간’이라고 표현한 것을 본 연구에서는 편의상 피치포인트 이동시간으로 바꾸어 사용하였다.

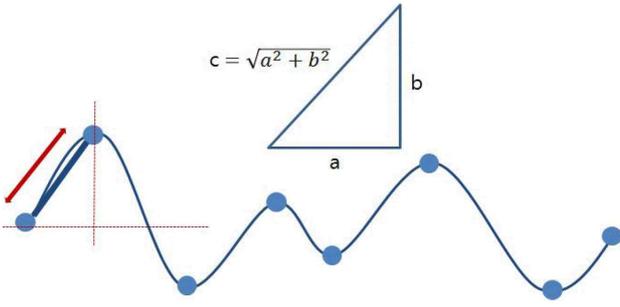


그림 5. 피치포인트 이동거리 측정(오재혁, 2014a)

그림 5. Measurement of the distance between pitch points (Oh, 2014a)

## 2. 실험

### 2.1 감정음성 데이터베이스

본 연구에서 분석한 독일어 감정음성 데이터베이스(A Database of German Emotional Speech)는 Burkhardt등에 의해 2005년에 만들어졌다. 이를 위해 5명의 여성 배우와 5명의 남성 배우들이 7가지 감정(anger, fear, happiness, neutral, sadness, disgust, boredom)을 연기하면서 발화했고 무향실에서 Sennheiser MKH 40P 48 마이크와 Tascam DA-P1 DAT 녹음기로 총 800개의 발화문장을 녹음했다. 발화에 사용된 문장은 감정중립적인 문장으로 총 10문장이었다.

이 발화자료는 청취 평가자들 중 60%이상이 80%이상의 점수를 준 발화를 엄선했으므로 높은 신뢰도를 가진 것으로 알려져 있다. 더 나아가 음소와 음절단위 뿐만 아니라 음절과 음성의 자질까지도 레이블링 되어 있어서 연구가치가 높은 자료이다(Burkhardt et al., 2005). 본 연구에서는 7가지 감정 중 감정음성 연구에서 주로 다루고 있는 5가지 감정 즉, 분노(angry), 두려움(fearful), 행복(happiness), 중립(neutral), 슬픔(sadness)의 음성을 대상으로 분노에서 127문장, 두려움에서 69문장, 행복에서 71문장, 중립에서 79문장, 슬픔에서 62문장을 뽑아 분석대상으로 삼았다.

### 2.2 분석 방법

먼저 피치측정 값의 물리적인 단위로 기존연구(오재혁, 2014a; 오재혁, 2014b)에서 사용한 Hz대신 청각적 스케일 단위인 mel을 사용하였다<sup>3)</sup>. 감정음성의 인지는 청각으로 처리된 신호를 대상으로 인간이 판단하는 작업이므로 물리적인 단위인 Hz보다는 청각적 특성을 반영한 mel 스케일을 사용해서 피치 값을 구하는 것이 보다 감정음성의 음향분석에 합리적일 것이다. 또한 자동음성인식 시스템에서도 성능향상을 이유로

mel 스케일로 필터간격을 디자인하고 있으므로 청각적 스케일로의 변환은 공학시스템과의 연계적 가치를 위해서도 필요한 일이라 생각한다.

본 연구는 억양을 근접복사 유형화시킨 후 기존연구에서 살펴본 발화문장 내 억양자질들(피치포인트 이동시간, 피치포인트 이동거리, 피치포인트 피치변위, 피치포인트 기울기)뿐만 아니라 각각의 발화문장 전체의 피치기울기와 피치변위 그리고 발화문미의 피치변위 등의 억양자질들을 추가하여 기존연구(오재혁, 2014a; 오재혁, 2014b)를 보완하였다.

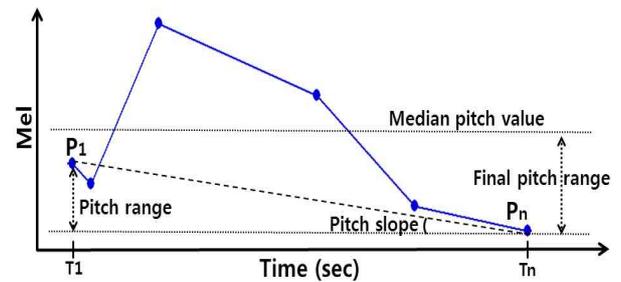


그림 6. 발화문장 전체 억양자질

Figure 6. Intonation features from a sentence

본 연구에 추가 억양자질로 사용된 발화문장 피치기울기(Pitch slope of a sentence)는 <그림 6>에서와 같이 마지막 피치포인트(Pn)와 첫 번째 피치포인트(P1)간의 피치차이를 이들 간의 시간(Tn - T1)으로 나눈 값이다. 발화문장 피치변위(Pitch range of a sentence)는 첫 번째 피치포인트와 마지막 피치포인트의 피치차이로 측정하였다. 그리고 발화문미 피치변위(Final pitch range)는 한 발화문장에서 발견되는 모든 피치포인트들의 피치 중앙값(Median pitch value)과 마지막 피치포인트의 피치와의 차이로 계산하였다. 중앙값은 일반적으로 사용되는 평균값보다 피치의 분석에서와 같이 오류와 정확도의 문제로 인해 생기기 쉬운 극단적 값들(outliers)의 영향을 덜 받기 때문에 정규화가 필요한 음성분석의 연구에 쓰여지고 있다(Forsell 2007, 이서배 et al., 2011). 또한 본 연구에서 억양 특성을 산출하기 위해 작성한 Praat 스크립트에서 피치분석의 정확도를 높이기 위해 측정대상 지점의 전후 20ms 구간의 평균과 표준편차를 구한 뒤 평균값에 3표준편차(3\*표준편차) 값을 더하거나 빼어서 피치측정 범위의 한계를 정한 후 구한 값이 이 범위 안에 들어있을 때는 그 값을 피치 값으로 지정하고 그렇지 않으면 그 지점 전후 20ms구간의 평균값을 대입했다. 이 범위 밖으로 나가는 경우 성대의 떨림이 이 정도로 급작스럽게 변하지 않을 것이라는 전제하에 해당 20ms구간의 평균값을 취함으로써 피치에러를 제거한 것이다(양병근, 2000).

3) 변환식  $mel = 1127.01048 \times \ln(1 + f/700)$  을 사용함. 여기서 f0는 기본주파수이고 mel은 청각스케일의 음높이 단위이다.

3. 결과 및 분석

3.1 발화문장 내 피치포인트 간 억양자질

두 화자 집단(여성, 남성)과 다섯 가지 감정음성(분노, 두려움, 행복, 중립, 슬픔)을 모수요인으로 하고 억양 근접복사 유형화로부터 얻은 발화 내 자질들(피치포인트 이동시간, 피치포인트 변위, 피치포인트 기울기, 피치포인트 이동거리)을 종속변수로 하는 MANOVA(다변량 분산분석)를 실시한 결과 남녀 두 화자 집단 간에는 어느 변수에서도 유의미한 차이가 없었고 두 화자 집단과 감정음성 간에는 어떠한 상호작용도 없었다. 이것은 기존의 억양곡선 정규화 연구(오재혁, 2014a; 오재혁, 2014b)와 부합하는 결과이다. 또한 측정된 억양 자질들이 화자들의 생리적인 특성에 영향을 받지 않는다는 점에서 정규화된 자질로서 감정음성의 분석에 바람직하다는 것을 뜻하기도 한다.

그리고 감정음성은 피치포인트 이동시간[F(4, 31.590),  $p < 0.001$ ,  $\eta_p^2 = 0.026$ ]과 피치포인트 이동거리[F(4, 21.901),  $p < 0.001$ ,  $\eta_p^2 = 0.018$ ]에서 유의미한 차이를 보였다(<그림 7>, <그림 8>, <표 1> 참조). 그러나 피치포인트 변위와 피치포인트 기울기에서는 유의미한 변화가 없었다.

피치포인트의 기울기에서 감정음성 간에 유의미한 차이가 없었던 것은 억양곡선의 모양과 감정음성의 관계를 살펴본 연구(Bänziger et al., 2005)와 맥락을 같이하고 있다. 이 연구에서도 감정발화에서 피치변위의 영향을 통제했을 때 수작업으로 설정한 피치포인트 간의 기울기가 감정에 의해 받는 영향이 유의하지 않은 것으로 나타났다.

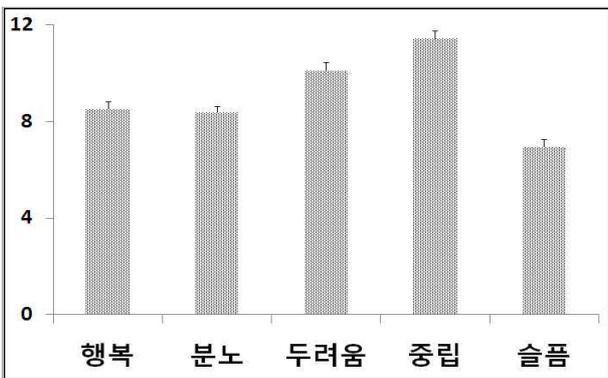


그림 7. 감정음성에서 측정된 피치포인트 이동시간 평균(%)  
Figure 7. Pitch point movement time (%) measured from emotional utterances

효과크기에서 보면, 감정음성의 효과가 피치포인트 이동거리( $\eta_p^2 = 0.018$ )보다 피치포인트 이동시간( $\eta_p^2 = 0.026$ )에서 더 크게 나타났다. 이것은 피치포인트가 얼마나 움직이는지에 대한 거리 정보보다도 얼마의 시간간격으로 피치곡선에 있어서 중요한 변화가 생기느냐(피치곡선에 변곡점이 생길 때마다 피

치포인트가 발생함)가 화자의 감정을 더 크게 반영하고 있다는 것을 뜻한다.

유의미한 차이를 보인 두 종속변수(피치포인트 이동시간과 피치포인트 이동거리)별로 어느 감정들 간에 차이가 있는지를 살펴보기 위해 사후검정(Tukey HSD)을 실시했다. 그 결과, 피치포인트 이동시간의 경우, 다음과 같은 감정음성들 간에 차이가 났다: 분노 vs. 두려움 ( $p < 0.01$ ), 분노 vs. 중립 ( $p < 0.001$ ), 분노 vs. 슬픔 ( $p < 0.001$ ), 두려움 vs. 행복 ( $p < 0.05$ ), 두려움 vs. 중립 ( $p < 0.05$ ), 두려움 vs. 슬픔 ( $p < 0.001$ ), 행복 vs. 중립 ( $p < 0.001$ ), 행복 vs. 슬픔 ( $p < 0.001$ ), 중립 vs. 슬픔 ( $p < 0.001$ ). 그리고 피치포인트 이동거리의 경우는 다음과 같은 감정음성들 간에 차이가 났다: 분노 vs. 두려움 ( $p < 0.005$ ), 분노 vs. 슬픔 ( $p < 0.001$ ), 두려움 vs. 행복 ( $p < 0.001$ ), 두려움 vs. 슬픔 ( $p < 0.001$ ), 행복 vs. 중립 ( $p < 0.05$ ), 행복 vs. 슬픔 ( $p < 0.001$ ), 중립 vs. 슬픔 ( $p < 0.001$ ). 정리하면, 피치포인트 이동시간의 경우 행복과 분노 간에는 차이가 없었지만 나머지 감정들은 서로 간에 유의한 차이를 보였고 피치포인트 이동거리의 경우, 행복과 분노 간에 역시 차이가 없었고 분노와 중립 그리고 두려움과 중립 간에도 차이가 없었다.

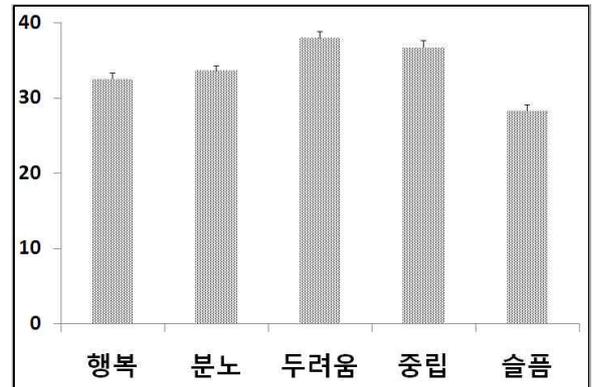


그림 8. 감정음성에서 측정된 피치포인트 이동거리  
Figure 8. Pitch point distance measured from emotional utterances

표 1. 감정음성별 피치포인트 이동시간(%)과 피치포인트들간 이동거리의 평균과 표준오차

Table 1. Mean and standard error of pitch point movement time (%) and distance between pitch points

	분노	행복	두려움	중립	슬픔
이동시간 평균	8.52	8.39	10.12	11.43	6.96
표준오차	0.30	0.22	0.33	0.32	0.29
이동거리 평균	32.57	33.68	38.04	36.78	28.31
표준오차	0.80	0.58	0.87	0.86	0.78

동일집단군분석에서 피치포인트 이동시간의 경우, 행복과

분노는 동일집단으로 분류되고 두려움, 중립, 슬픔은 각각 따로 따로 구별되는 것으로 나타났다<sup>4)</sup>. 피치포인트 이동거리의 경우에도 행복과 분노가 동일집단으로 분류되지만 분노와 중립도 한 묶음 그리고 두려움과 중립이 한 묶음 될 수 있는 것을 보여주고 있다<sup>5)</sup>.

두 억양자질의 분석에 있어서 공통되는 것은 행복과 분노가 음향적으로 서로 차이가 없고 다른 감정들과도 구별이 되고 슬픔도 다른 감정들과 확연히 구별된다는 것이다. 즉, 발화문장 내 억양의 음향적 관점에서 보면 행복과 분노의 감정음성은 음향적 특징이 서로 비슷하고 슬픔의 감정음성과 마찬가지로 다른 감정음성과 음향적으로 구별될 수 있는 특징을 가진다고 말할 수 있다.

기존 연구(Fontaine et al., 2007, Goudbeek et al., 2010)에서 감정음성은 Arousal(각성)의 크고 작음의 정도에 따라 또한 Potency/Control(힘/통제)에 따라 그리고 Valence(긍정적, 부정적 감정)에 따라 <표 2>과 같이 나누어 질 수 있다. 이러한 맥락에서, Arousal과 Potency/Control 둘 다가 크다는 점에서 행복과 분노의 음성은 음향적 특징이 다른 감정음성들과 비교해서 일관성 있게 구별되었다는 추정이 가능하다(행복과 분노는 Arousal이 크다는 점에서 슬픔과 구분되고 Potency/Control이 높다는 점에서 두려움과 구분됨).

표 2. 세가지 차원으로 구분된 감정들:

Potency/Control(힘/통제)이 높은 감정은 이태릭체의 굵은 폰트로 나타내었음(Goudbeek et al., 2010)

Table 2. Three-dimensional representation of emotions: Emotions high in potency/control are shown in italic boldface (Goudbeek et al., 2010)

Arousal(각성)	Valence(긍정적, 부정적)	
	긍정적	부정적
높음	<b>행복(기쁨)</b>	<b>분노</b>
	재미	두려움(공포)
	<b>자극(교만)</b>	절망
낮음	안도	근심(걱정)
	<b>관심</b>	슬픔(우울)

또한 슬픔의 감정음성이 다른 감정음성들과 구별되는 이유도 슬픔이 Arousal이 작고 Potency/Control이 낮고 Valence가 부정적인 것과 관련이 있을 가능성이 크다. 이와 같은 결과는 인지적이고 주관적인 감정평가를 물리적인 음향자질들(피치포인트 이동시간, 피치포인트 이동거리)의 측정값과 연관 지어

- 4) 예를 들면, 집단 1은 행복과 분노, 집단 2는 두려움, 집단 3은 중립 그리고 집단 4는 슬픔으로 분류될 수 있다.
- 5) 예를 들면, 집단 1은 행복과 분노, 집단 2는 분노와 중립, 집단 3은 두려움과 중립 그리고 집단 4는 슬픔으로 분류될 수 있다.

생각할 수 있는 가능성을 제시하고 있다(<그림 7>, <그림 8>, <표 1> 참조).

### 3.2 발화문장 전체 억양자질

두 화자 집단(여성, 남성)과 다섯 가지 감정음성(분노, 두려움, 행복, 중립, 슬픔)을 모수요인으로 하고 억양 근접복사 유형화로 억양곡선 전체 즉, 각각의 발화문장 전체로부터 얻은 억양 자질들(발화문미 피치변위, 발화문장 피치변위, 발화문장 피치기울기)을 종속변수로 하는 MANOVA(다변량 분산분석)를 실시한 결과 남녀 두 화자 집단 간에는 발화문미 피치변위 [F(1, 5.194), p<0.05,  $\eta_p^2$ = 0.013]와 발화문장 피치변위[F(1, 4.016), p<0.05,  $\eta_p^2$ = 0.010]에서 유의미한 차이가 있었다. 반면에 문장전체의 피치기울기는 남녀 간 차이가 없는 것으로 나타났다. <그림 9>에서 보면 전반적으로 남성보다 여성의 피치변위가 큰 것을 볼 수 있다. 이러한 경향은 기존연구(오재혁, 2014a; 오재혁, 2014b)에서도 보고된 바 있다. 그리고 앞 결과 마찬가지로 화자집단과 감정음성 간에는 어떠한 상호작용도 나타나지 않았다.

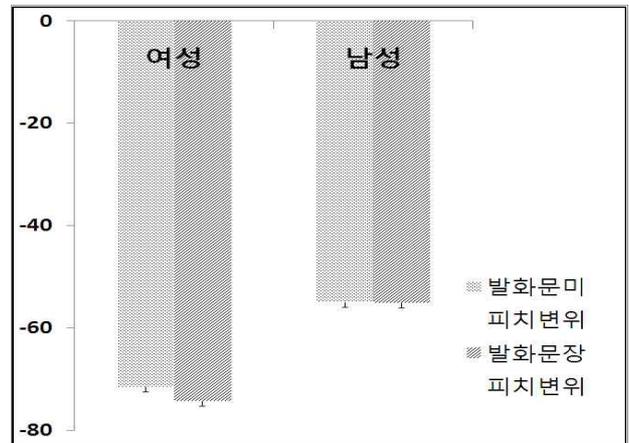


그림 9. 남녀 간 피치변위 평균값(mel)

Figure 9. Average pitch range between men and women

감정음성과 억양음향 자질들의 관계에서는 발화문미 피치변위[F(4, 28.614), p<0.001,  $\eta_p^2$ = 0.223]와 발화문장 피치변위 [F(4, 4.626), p<0.005,  $\eta_p^2$ = 0.044] 그리고 발화문장 피치기울기[F(4, 4.616), p<0.005,  $\eta_p^2$ = 0.044] 모두에서 유의미한 차이가 나타났다. 효과크기에서 보면, 감정음성의 효과가 발화문장 피치변위( $\eta_p^2$ = 0.044)와 발화문장 피치기울기( $\eta_p^2$ = 0.044)보다 발화문미의 피치변위( $\eta_p^2$ = 0.223)에서 컸다. 이것은 감정음성의 효과가 발화문장 피치기울기에서는 약한 반면 발화문미의 피치변위에서는 잘 드러난다는 보고(Bänziger et al., 2005)와 맥락을 같이 하는 결과이다.

유의미한 차이를 보인 세 억양자질(발화문미 피치변위, 발화문장 피치변위, 발화문장 피치기울기)별로 어느 감정들 간

에 차이가 있는지 살펴보기 위해 사후검정(Tukey HSD)을 실행했다. 사후검정결과, 발화문미 피치변위의 경우 다음과 같은 감정음성들 간에 차이가 났다: 분노 vs. 두려움 ( $p < 0.001$ ), 분노 vs. 중립 ( $p < 0.001$ ), 분노 vs. 슬픔 ( $p < 0.001$ ), 두려움 vs. 행복 ( $p < 0.001$ ), 두려움 vs. 슬픔 ( $p < 0.05$ ), 행복 vs. 중립 ( $p < 0.001$ ), 행복 vs. 슬픔 ( $p < 0.001$ ). 발화문장 피치변위의 경우 다음과 같은 감정음성들 간에 차이가 났다: 두려움 vs. 행복 ( $p < 0.05$ ), 행복 vs. 중립 ( $p < 0.005$ ), 행복 vs. 슬픔 ( $p < 0.05$ ). 그리고 발화문장 피치기울기의 경우 다음과 같은 감정음성들 간에 차이가 났다: 행복 vs. 중립 ( $p < 0.005$ ), 행복 vs. 슬픔 ( $p < 0.005$ ).

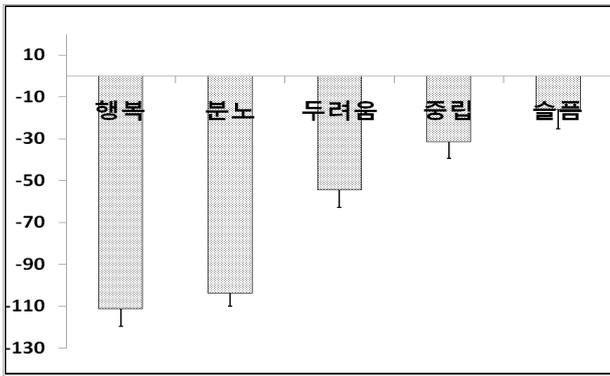


그림 10. 그림 10. 감정음성에서 측정된 발화문미 피치변위 평균(mel)  
Figure 10. Mean of sentence final pitch range measured from emotional utterances (mel)

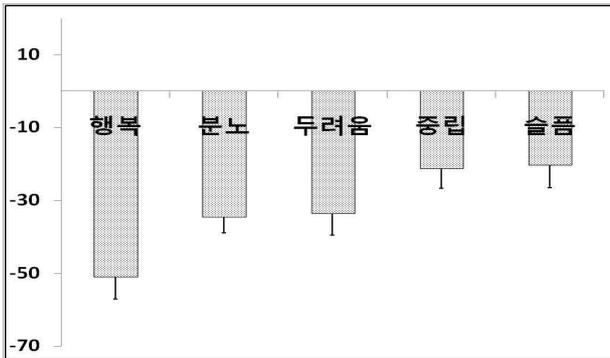


그림 11. 감정음성에서 측정된 발화문장 피치기울기 평균값  
Figure 11. Mean of sentence pitch slope measured from emotional utterances

동일집단분석에서는 발화문미 피치변위의 경우 행복과 분노가 한 집단으로 묶이고 두려움과 중립이 한 집단으로 묶였으며 중립과 슬픔이 하나로 묶였다. 발화문장 피치변위의 경우 행복과 분노가 한 집단으로 묶이고 분노, 두려움, 중립, 슬픔이 하나로 묶였다. 발화문장 피치기울기의 경우 행복, 분노, 두려움이 한 집단을 형성했고 분노, 두려움, 중립, 슬픔이 동일 집단에 소속되었다.

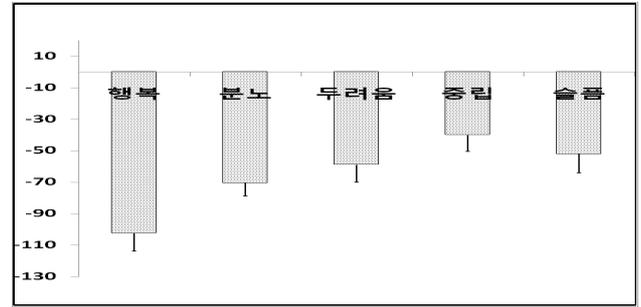


그림 12. 감정음성에서 측정된 발화문장 피치변위 평균(mel)  
Figure 12. Mean of sentence pitch range measured from emotional utterances (mel)

이상의 결과를 보면 발화문장 내 억양자질과 마찬가지로 발화문장 전체의 억양자질에서도 행복과 분노가 일관성 있게 동일그룹에 속했다. 즉, 행복과 분노는 발화문장 수준에서도 억양 음향적으로 서로 차이가 없고 다른 감정음성들과도 구분된다는 것이다. 이러한 패턴은 <표 2>에 나타난 바와 같이 감정의 Arousal이 높은 분노와 행복 그 다음의 두려움과 중립 그리고 Arousal의 정도가 가장 낮은 슬픔으로 분류되는 것과 관련이 있는 것 같다.

<그림 10>에서 보면 감정의 Arousal이 클수록 발화문미 피치변위 평균값은 더 큰 음의 값을 가지는 것을 볼 수 있다. 발화문미 피치변위가 더 큰 음의 값을 가졌다는 것은 마지막 피치포인트의 피치값과 발화문장의 중앙값과의 차이가 다른 감정의 경우보다 컸다는 것을 뜻한다. 즉, Arousal이 클수록 마지막 피치포인트가 더 많이 내려간다는 해석이 가능하다. 이것은 Arousal이 클수록 문미피치곡선이 더 큰 기울기를 가지고 내려간다는 보고(Bänziger et al., 2005)와도 부합하는 내용이다.

표 3. 발화문장 전체를 대상으로 측정된 억양자질들의 평균과 표준오차

Table 3. Mean and standard error of the acoustic features measured at sentence level

	행복	분노	두려움	중립	슬픔
발화문미 피치변위 평균(mel)	-111.09	-103.59	-54.22	-31.45	-15.99
표준오차	8.58	6.24	8.46	7.9	9.09
발화문장 피치변위 평균(mel)	-102.307	-70.47	-58.82	-40.03	-52.03
표준오차	11.33	8.24	11.17	10.43	12
발화문 기울기 평균	-51.12	-34.64	-33.73	-21.31	-20.32
표준오차	5.87	4.27	5.79	5.41	6.22

발화문장 피치기울기의 경우도 발화전체의 피치기울기가 급하게 내려갈수록 감정의 Arousal 정도가 커진 것을 반영했다고 말할 수 있을 것이다(<그림 11> 참조). 이것은 발화문장의 마지막 피치포인트가 첫 번째 피치포인트보다 피치가 더 떨어졌고(낮았고) 전체 발화의 길이가 더 짧았다는 것을 의미한다.

바꾸어 말하면, <그림 10>, <그림 11>, <그림 12>에서 나타나는 Arousal이 제일 작은 슬픔의 감정음성이 Arousal이 큰 행복, 분노, 두려움의 감정음성보다 음향적 절대값이 일관성 있게 작게 나타났다고도 말할 수 있다. 이는 슬픔의 감정음성이 Arousal과 Potency/Control이 낮고 Valence가 부정적인 것과 관련이 있을 것으로 추정된다. 한편 Arousal이 크고 Potency/Control이 높은 행복과 분노의 감정음성은 다른 감정음성들보다 음향적 절대치가 더 컸다. 이상과 같이, 앞 절의 결과와 마찬가지로 발화문장 전체와 관련한 억양자질들도 주관적인 감정음성의 평가를 물리적인 음향자질들의 측정값과 연관 지을 수 있는 가능성을 보여준다.

#### 4. 결론

지금까지 억양의 근접복사 유형화(close-copy stylization)를 이용한 감정음성의 음향적 특징을 살펴보았다. 발화문장 내 피치포인트 간 억양자질들로 감정음성을 분석했을 때 남녀 간에 유의한 차이가 없는 것으로 나타나 기존연구(오재혁, 2014a; 오재혁, 2014b)에서와 마찬가지로 정규화에 필요한 억양자질들의 항상성이 재확인 되었다.

발화문장 내 피치포인트 간 억양자질들 중 피치포인트 이동시간과 피치포인트 이동거리는 분노와 행복을 일관성 있게 하나로 묶어 주어 다른 감정음성들과 음향적으로 차별됨을 보여주었다. 분노와 행복은 Arousal(각성)과 Potency/Control(힘/통제)의 정도가 크다는 공통점이 있다. 한편, 슬픔의 감정은 Arousal이 작고 Potency/Control이 낮고 Valence가 부정적이데 음향적으로도 다른 감정들과 일관성 있게 구별되었다.

발화문장 전체를 대상으로한 억양자질들 중 발화문장 피치변위와 발화문미 피치변위의 측정에서 여성과 남성사이에 의미 있는 음향적 차이가 있는 것으로 나타났고 남성보다 여성의 피치변위가 크다는 연구(오재혁, 2014a; 오재혁, 2014b)와 부합하는 결과가 나왔다. 그리고 Arousal이 작은 슬픔보다 Arousal이 큰 행복과 분노가 발화문미 피치변위와 발화문장 피치변위 그리고 발화문장 피치기울기의 절대값이 더 큰 것으로 보아 문장전체 기울기와 문미 기울기가 더 급하게 내려 갈수록 그리고 피치변위의 폭이 더 클수록 감정적으로 더 큰 Arousal을 반영한다고 볼 수 있다.

이러한 결과로 볼 때, 본 연구는 음성으로 표현되어 주관적으로 인지되는 감정의 정도를 물리적인 음향자질들의 값으로

추정할 수 있는 가능성을 보여준다고 생각한다. 향후 감정음성의 억양과 관련한 초분절적 자질들을 감정별로 차이가 나는 모음의 포트먼트값(이서배, 2011)과 같은 분절적 자질들과 연계해 더 많은 음향적 자질들로 감정음성을 분류하는 방안을 모색할 필요가 있을 것이다.

#### 참고문헌

- [1] Bachorowski, J.& Owren, M. (1995) Vocal expression of emotion: acoustic properties of speech are associated with emotional intensity and context. *Psychological Science*, 6, 219-224.
- [2] Bänziger, T. & Scherer, K.R. (2005). The role of intonation in emotional expressions. *Speech Communication*. 46, 252-267.
- [3] Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott International*. 5:9/10, 341-345.
- [4] Burkhardt, F., Paeschke, A., Rolfes, M., Sendlmeier, W. and Weiss, B. (2005). A database of German emotional speech. *Proceedings of Interspeech 2005*, 1517-1520.
- [5] Fontaine, J., Scherer, K., Roesch, E., and Ellsworth, P. (2007). The world of emotions is not two-dimensional. *Psychological Science*, 18(12),1050-1057.
- [6] Forsell, M. (2007). *Acoustic correlates of perceived emotions in speech*. MS Thesis, KTH, Royal Institute of Technology, Stockholm, Sweden.
- [7] Goudbeek, M. & Scherer, K. (2010). Beyond arousal: valence and potency/control cues in the vocal expression of emotion. *Journal of the Acoustical Society of America*, 128(3),1322-36.
- [8] Jun, Sun-Ah. (1993). *The phonetics and phonology of Korean prosody*. Ph. D. Dissertation, The Ohio State University.
- [9] Lee, Ho-Young. (1990). *The structure of Korean prosody*. Ph. D. Dissertation, University College London. (Published by Hanshin Publishing.)
- [10] Mozziconacci, S. (2002). Prosody and emotions. *Proceedings of Speech Prosody*, Aix-en-Provence, 1-9.
- [11] Oh, J. H. (2014a). A study of methods of standardization for Korean intonation curve. *Korean Linguistics*. 62, 395-420. (오재혁 (2014a). 한국어 억양 곡선의 정규화 방안에 대한 연구. *한국어학*, 62, 395-420.)
- [12] Oh, J. H. (2014b). A study of intonation curve slopes in Korean spontaneous speech. *Phonetics and Speech Sciences*. 6(1), 21-30. (오재혁 (2014b). 자유 발화 자료에서 나타나는 한국어 억양 곡선의 기울기 특성에 대한 연구. *말소리와 음성과학*, 6(1), 21- 30.)

- [13] Pell, M., Paulmann, M., Dara, S., Allasseri, A. and Kotzb, S. (2009) Factors in the recognition of vocally expressed emotions: a comparison of our languages. *Journal of Phonetics*, 37, 417-435.
- [14] Rodero, E. (2011). Intonation and emotion: influence of pitch levels and contour type on creating emotions. *Journal of Voice*. 25(1), 25-34.
- [15] Yang, B. G. (2000). An acoustical analysis method of numeric sounds by Praat. *Speech Science*. 7(2), 127-137.  
(양병곤 (2000). Praat를 이용한 숫자음의 음향적 분석법, 음성과학, 7(2), 127-137.)
- [16] Yi, S., and Kim, S. (2011). A study on low pitch accent produced in different locations in English sentences. *Phonetics and Speech Sciences*, 3(4), 63-70.  
(이서배, 김수정 (2011). 영어 문장 내 상이한 위치에 나타난 저성조 피치 액센트 연구, 말소리와 음성과학, 3(4), 63-70.)
- [17] Yi, S. (2011). An analysis of formants extracted from emotional speech and acoustical implications for the emotion recognition system and speech recognition system. *Phonetics and Speech Sciences*, 3(1), 45-50.  
(이서배 (2011). 독일어 감정음성에서 추출한 포먼트의 분석 및 감정인식 시스템과 음성인식 시스템에 대한 음향적 의미, 말소리와 음성과학, 3(1), 45-50.)
- [18] Williams, C. & Stevens, K. (1972) Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America*, 52, 1238-1250.

• 이서배 (Yi, So Pae)

창원대학교 영어영문학과  
(641-773) 경남 창원시 의창구 사림동  
Tel.: 055-540-5466 Fax: 055-540-5465  
Cel.: 010-5555-6305  
Email: sopaeyi@pusan.ac.kr