pISSN 2005-8063
eISSN 2586-5854
2020. 9. 30.
Vol.12 No.3
pp. 25-32

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2020.12.3.025

An analysis of emotional English utterances using the prosodic distance between emotional and neutral utterances

So-Pae Yi*

Humphreys West Elementary School, Department of Defense Education Activity, Camp Humphreys, Pyeongtaek, Korea

Abstract

An analysis of emotional English utterances with 7 emotions (calm, happy, sad, angry, fearful, disgust, surprised) was conducted using the measurement of prosodic distance between 672 emotional and 48 neutral utterances. Applying the technique proposed in the automatic evaluation model of English pronunciation to the present study on emotional utterances, Euclidean distance measurement of 3 prosodic elements such as F0, intensity and duration extracted from emotional and neutral utterances was utilized. This paper, furthermore, extended the analytical methods to include Euclidean distance normalization, z-score and z-score normalization resulting in 4 groups of measurement schemes (sqrF0, sqrINT, sqrDUR; norsqrF0, norsqrINT, norsqrDUR; sqrzF0, sqrzINT, sqrzDUR; norsqrzF0, norsqrzINT, norsqrzDUR). All of the results from perceptual analysis and acoustical analysis of emotional utteances consistently indicated the greater effectiveness of norsqrF0, norsqrINT and norsqrDUR, among 4 groups of measurement schemes, which normalized the Euclidean measurement. The greatest acoustical change of prosodic information influenced by emotion was shown in the values of F0 followed by duration and intensity in descending order according to the effect size based on the estimation of distance between emotional utterances and neutral counterparts. Tukey Post Hoc test revealed 4 homogeneous subsets (calm<disgust, sad<happy, surprised<surprised, angry, fearful) statistically determined from the measurement of norsqrF0 and 3 homogeneous subsets (surprised, happy, fearful, sad, calm<calm, angry<angry, disgust) from norsqrDUR. Furthermore, the analysis of each of the 7 emotions showed that the present research outcome is in the same vein as the results of the previous study.

Keywords: emotional utterance, prosody, euclidean distance, arousal, valence, stance

1. 서론

운율(prosody)은 발화된 내용 자체의 언어적 정보(linguistic information) 뿐만 아니라 그 이상의 정보 즉, 화자의 감정이나 태도 등의 준언어적 정보(paralinguistic information)를 전달함으로써 발화를 좀 더 정확하게 해석하도록 도움을 주고 있다 (Kitayama & Ishii, 2002; Paulmann, 2016; Thompson & Balkwill, 2009). 이런 맥락에서 이루어진 연구들에서는 화자의 감정이나 태도의 전달에 있어서 피치 평균값, 피치 범위, 발화 속도(Cahn, 1990; Carlson et al., 1992; Kitahara & Tohkura, 1992; Mozziconacci,

* sopaeyi@pusan.ac.kr, Corresponding author

Received 30 July 2020; Revised 10 September 2020; Accepted 10 September 2020

1998; Vroomen et al., 1993), 억양 곡선의 유형(Bachorowski & Owren, 1995; Pell et al., 2009; Williams & Stevens, 1972) 등의 운율요소가 중요한 역할을 감당하는 것으로 보고 되었다. 더 나아가 정규화된 억양 곡선을 통해 자동으로 음향 자질들을 추출함으로써 감정음성들을 분석 및 분류하려는 시도(Yi, 2018)도 보고 되고 있다.

이러한 억양 곡선은 외국어 학습자의 발음평가에 사용되기도 한다. 이러한 연구에서는 원어민과 학습자의 발음이 얼마나 차이 나는가를 자동 평가하기 위해 억양 곡선뿐만 아니라 발화 문장의 강도 곡선 그리고 분절음의 길이 정보를 함께 사용해 원어민 발화와 학습자 발화에서 추출한 각 운율요소를 사용해 두 발화의 유클리디언 거리를 계산하였다. 이를 근거로 음의 높낮이(Hz), 음의 강도(dB), 음의 길이(sec)로 구성된 3차원 공간에서 원어민 발화와 학습자 발화의 위치를 계산하여 그 유사성 정도를 추정하였다(Yoon, 2009a). 그리고 학습자의 발음이 얼마나 좋은지를 사람이 판단하는 수동평가를 기계적인 운율계산을 통한 자동평가와 비교하기도 하였다(Yoon, 2013).

본 연구는 이러한 외국어 발음 자동평가 기법이 운율의 대표적 3요소인 음의 높낮이(Hz), 음의 강도(dB), 음의 길이(sec)에 기반을 두고 있는 점에 착안하여 음성으로 감정을 표현하는 데 있어서 운율의 3요소가 어떤 양상으로 역할을 감당하는지 살펴보기 위하여 감정표현과 운율요소들 간의 관계를 정량적으로 분석하였다. 그리고 사람이 평가한 수동평가와 감정발화와 감정중립 발화 간의 운율거리 값들과의 관계도 살펴보았다. 이를 위해, 기존연구에서 사용한 운율요소들의 유클리디언 거리계산 뿐만 아니라 유클리디언 거리계산 정규화 방법과 z-score 및 z-score 정규화 방법도 아울러 살펴보았다.

2. 연구방법

2.1. 감정발화 데이터베이스

본 연구는 감정발화 분석을 위해 The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)을 사용하였다. RAVDESS는 24명(남성 12명, 여성 12명)의 배우들이 어휘적으로 균형잡힌 2개의 문장을 calm(차분한), happy(행복한), sad(슬픈), angry(화난), fearful(두려운), disgust(혐오스런) 그리고 surprised(놀란)의 7개 감정으로 발화한 데이터베이스이다. 이들은 모두 북미 억양을 구사하는 사람들로서 각각의 감정을 2 수준의 감정표현 강도(1-보통, 2-강함)로 구분하여 발화하였고 동일한 문장의 감정중립 발화를 추가적으로 녹음 하였다. 그 두 문장은 다음과 같다:

Kids are talking by the door.
Dogs are sitting by the door.

또한, RAVDESS는 발화된 감정이 얼마나 잘 인지되는지를 검증하기 위해 247명의 청취자가 감정발화를 듣고 평가하였다. 그리고 이들 평가자들로부터 각각의 발화가 해당 감정을 얼마나 잘 표현하였는지를 나타내는 Goodness score(양호 점수)를 도출하였다(Livingstone & Russo, 2018).

2.2. 분석방법

본 연구에서는 2개의 문장을 24명의 배우가 7개의 감정과 2수준의 감정표현 강도로 발화한 672개(2×24×7×2)의 발화와 48개(2×24)의 감정중립 발화를 분석 대상으로 하였다. 먼저 Montreal Forced Aligner로 자동 레이블링을 한 후 수작업으로 검증하여 에러 난 경계들을 고치고 서로 다르게 레이블 된 부분들을 일관성 있게 통일하였다. 그리고 운율거리 계산을 위해 작성한 Praat(Boersma & Weenink, 2020) 스크립트를 사용하여 감정중립 발화와 각각의 감정발화들과의 운율거리를 계산하였다.

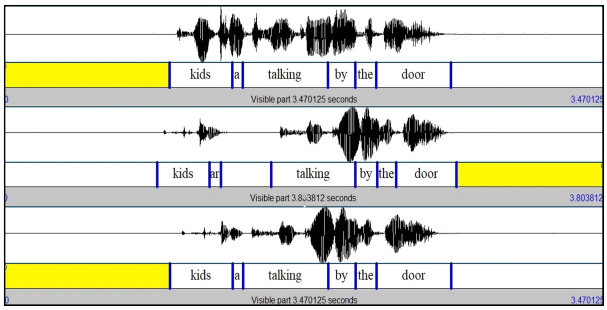

그림 1. 레이블링된 발화들(위에서부터 순서대로 중립 발화, 두려운 감정발화, 중립 발화의 단어길이를 복제한 두려운 감정발화)
Figure 1. Labeled utterances (from top to bottom: neutral, fearful, and fearful utterance cloned with durations of words from the neutral utterance)

한 화자가 한 문장을 발화한 감정중립 발화와 동일한 화자가 동일한 문장을 14가지 방식(7가지 감정×2수준 감정표현 강도)으로 발화한 감정발화와의 운율거리를 계산하기 위한 전처리 과정을 위해 praat 스크립트로 구현한 운율복제기법(Yoon, 2009b)을 이용해 14가지 발화들의 단어경계들이 해당 감정중립 발화의 단어경계들과 같아지도록 하였는데 기존의 운율복제기법에서는 휴지 구간의 추가 등으로 두 발화의 레이블 개수가 차이 나면 운율복제를 수행할 수 없었지만 본 연구에서는 두 발화의 레이블 개수가 서로 달라도 감정발화의 단어들 각각의 길이가 감정중립 발화의 해당 단어들의 길이와 서로 같아질 수 있도록 알고리즘을 수정하였다. 예를 들어, 그림 1에서처럼 6개 단어(Kids are talking by the door.)로 이루어진 문장의 감정중립 발화의 레이블 개수가 휴지구간을 포함해서 8개이고 두려운 감정발화의 레이블 개수가 9개일 때 분석 대상인 6개 단어들의 발화길이는 감정중립 발화의 단어길이를 두려운 감정발화에 복제하여 감정중립 발화와 감정발화 모두에서 같아질 수 있도록 했다. 이렇게 하여 길이가 같아진 각 단어별로 프레임당 운율자질(F0, Intensity) 값들을 구하였다. 그리고 음의 길이(Duration) 정보는 단어길이 복제 이전의 원래 파일들에서 구하였다. 그리하여 총 4가지 거리측정 방법으로 감정중립 발화와 개별 감정발화 간의 운율거리를 구하였고 그 수식들은 다음과 같다.

$$sqrF0 = \sqrt{\sum_{i=1}^{n}(F0i_{\neq ut} - F0i_{emo})^2} \quad (1)$$

$$sqrINT = \sqrt{\sum_{i=1}^{n}(INTi_{\neq ut} - INTi_{emo})^2} \quad (2)$$

$$sqrDUR = \sqrt{\sum_{j=1}^{m}(DURj_{\neq ut} - DURj_{emo})^2} \quad (3)$$

$$sqrzF0 = \sqrt{\sum_{i=1}^{n}(zF0i_{\neq ut} - zF0i_{emo})^2} \quad (4)$$

$$sqrzINT = \sqrt{\sum_{i=1}^{n}(zINTi_{\neq ut} - zINTi_{emo})^2} \quad (5)$$

$$sqrzDUR = \sqrt{\sum_{j=1}^{m}(zDURj_{\neq ut} - zDURj_{emo})^2} \quad (6)$$

식 (1)-(3)에서 F0는 음의 높낮이(Hz), INT는 음의 강도(dB), DUR은 각 단어들의 길이(sec)이다. 식 (4)-(6)에서 zF0, zINT, zDUR은 한 발화 문장에서 나온 각각의 프레임별 F0, INT, DUR 값들에서 그 발화 문장 전체의 F0, INT, DUR 평균을 빼고 그 발화 문장 전체의 F0, INT, DUR의 표준편차로 나눈 z score 값들로서 이는 각 데이터 값이 평균으로부터 표준편차의 몇 배 정도로 떨어져 있는지를 나타내는 값이다. 이 값은 0이 되면 정확히 평균에 해당한다.

식 (1), (2), (4), (5)의 F0*i*, INT*i*, zF0*i*, zINT*i*는 *i*번째 프레임의 F0, INT, zF0, zINT 값들이다. 식 (3), (6)에서 DURj, zDURj는 j번째 단어의 DUR, zDUR 값이다. 그리고 n, m은 각각 해당 발화 분석구간의 총 프레임 수와 총 단어 수이다. 아래 첨자 *neut*는 감정중립 발화, 아래 첨자 *emo*는 해당 감정중립 발화를 생성한 동일 화자가 발화한 각각의 감정발화를 나타낸다.

식 (1)-(3)에서 sqrF0, sqrINT, sqrDUR은 감정중립 발화와 감정발화, 두 발화에서 나온 F0, INT, DUR로 계산한 두 발화의 유클리디언 거릿값이고 식 (4)-(6)의 sqrzF0, sqrzINT, sqrzDUR은 zF0, zINT, zDUR로 계산한 두 발화의 유클리디언 거릿값이다.

$$norsqrF0 = \sqrt{\sum_{i=1}^{n}(F0i_{\neq ut} - F0i_{emo})^2/n} \quad (7)$$

$$norsqrINT = \sqrt{\sum_{i=1}^{n}(INTi_{\neq ut} - INTi_{emo})^2/n} \quad (8)$$

$$norsqrDUR = \sqrt{\sum_{j=1}^{m}(DURj_{\neq ut} - DURj_{emo})^2/m} \quad (9)$$

식 (7)-(9)의 norsqrF0, norsqrINT, norsqrDUR은 유클리디언 거릿값을 총 프레임수 n과 총 단어 수 m으로 정규화한 유클리디언 거릿값(normalized Euclidean distance) 이다.

$$norsqrzF0 = \sqrt{\sum_{i=1}^{n}(zF0i_{\neq ut} - zF0i_{emo})^2/n} \quad (10)$$

$$norsqrzINT = \sqrt{\sum_{i=1}^{n}(zINTi_{\neq ut} - zINTi_{emo})^2/n} \quad (11)$$

$$norsqrzDUR = \sqrt{\sum_{j=1}^{m}(zDURj_{\neq ut} - zDURj_{emo})^2/m} \quad (12)$$

한편, 식 (10)-(12)의 norsqrzF0, norsqrzINT, norsqrzDUR은 z score 값들로 구한 유클리디언 거릿값을 총 프레임 수와 총 단어 수로 정규화한 값들이다. 이러한 값들을 가지고 감정중립 발화로부터 특정 감정발화가 얼마나 운율적으로 떨어져 있는지를 계산하고 분석하였다.

3. 결과 및 분석

통계 분석은 IBM SPSS Statistics 26을 사용하였다. 먼저, 인지적 측면에서 감정인지의 Goodness score와 운율거리 측정값들의 관계를 분석하였고 음향음성학적 측면에서 감정들과 감정표현 강도가 운율거리 측정값들에 어떤 영향을 주었는지 살펴보았다.

3.1. 감정발화의 인지

각 감정발화가 해당 감정을 얼마나 잘 표현하여 인지되었는지를 나타내는 Goodness score를 종속변수로 두고 운율거리 측정값들(sqrF0, sqrINT, sqrDUR; norsqrF0, norsqrINT, norsqrDUR; sqrzF0, sqrzINT, sqrzDUR; norsqrzF0, norsqrzINT, norsqrzDUR)을 독립변수로 하는 다중선형회귀분석(Multiple Linear Regression)을 수행하였다. 단, 독립변수를 sqrF0, sqrINT, sqrDUR와 norsqrF0, norsqrINT, norsqrDUR와 sqrzF0, sqrzINT, sqrzDUR 그리고 norsqrzF0, norsqrzINT, norsqrzDUR와 같이 4개의 그룹으로 나누어 각각 따로 수행하였다.

표 1. Goodness 값과 운율거리의 다중선형회귀분석
Table 1. Multiple Linear Regression of goodness scores and prosodic distance

운율거리	B	β	*t*-value	*p*-value	VIF
(상수)	3.056		16.255	0.000	
norsqrF0***	0.013	0.341	8.989	0.000	1.097
norsqrINT	−0.005	−0.042	−1.122	0.262	1.081
norsqrDUR	3.295	0.065	1.714	0.087	1.088

*** $p<0.001$.

분석결과 각 운율거리 측정방법별로 sqrF0, sqrINT, sqrDUR의 경우, $F(3, 668)=29.431$, $p<0.001$, $R^2=0.117$이고 norsqrF0, norsqrINT, norsqrDUR의 경우, $F(3, 668)=31.811$, $p<0.001$, $R^2=0.125$이고 sqrzF0, sqrzINT, sqrzDUR의 경우, $F(3, 668)=10.602$, $p<0.001$, $R^2=0.045$이고 norsqrzF0, norsqrzINT, norsqrzDUR의 경우, $F(3, 668)=10.567$, $p<0.001$, $R^2=0.045$로 모든 선형회귀 모델들이 $p<0.001$ 수준에서 통계적으로 유의미한 것으로 나타났는데 Durbin-Watson 값들(각각, 1.418, 1.438, 1.300, 1.318)이 모두 1에서 3 사이였으므로 잔차의 독립성이 충족된 것으로 나타났다. 또한 VIF값들(표 1) 이 모두 10 미만이므로 너무 비슷한 변수가 독립변수에 포함되었는지

여부를 판단하는 다중공선성 문제도 없는 것으로 나타났다. 정리하면, 4개 그룹의 선형회귀 모델들 모두가 유효했는데 그 중 유클리디언 운율거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR에서 감정인지에 대한 모델의 설명력을 나타내는 R^2의 값이 가장 컸다. 그리고 표 1에서와 같이 음의 높낮이(Hz)만이 감정인지를 유의미하게 설명하고 있는 것으로 나타났다(p<0.001). 그리고 감정인지를 잘 설명하는 운율거리는 표준화 계수 β의 크기순으로 음의 높낮이가 가장 크고 그다음 음의 길이 그리고 음의 강도 순인 것으로 나타났다.

3.2. 감정발화의 운율거리

감정들(calm, happy, sad, angry, fearful, disgust, surprised)과 감정표현 강도(1-보통, 2-강함)를 모수요인으로 하고 운율거리 측정방법들(sqrF0, sqrINT, sqrDUR; norsqrF0, norsqrINT, norsqrDUR; sqrzF0, sqrzINT, sqrzDUR; norsqrzF0, norsqrzINT, norsqrzDUR)을 종속 변수로 하는 다변량분석(MANOVA)을 시행했다. 분석결과 음성에 나타난 감정들은 모든 운율거리 측정방법들에서 음의 높낮이(Hz)와 각 단어들의 길이(sec) 변화에 유의미한 영향을 주었지만 음의 강도(dB)에는 영향을 주지 않은 것으로 나타났다(표 2). 또한, 음성으로 표현된 감정이 운율거리 자질들에 미치는 효과크기(effect size)인 η_P^2 값은 유클리디언 운율거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR에서 다른 운율거리 측정방법들보다 더 크게 나타났다(표 2). 이러한 경향은 감정표현 강도에 따른 음향적 차이의 분석에도 동일하게 나타났다. 즉, 감정표현 강도에 따른 효과크기도 다른 측정방법들에서보다 norsqrF0, norsqrINT, norsqrDUR에서 더 크게 나타났고 norsqrF0, norsqrINT, norsqrDUR 모두에서 통계적으로 유의미한 차이를 보였다(표 2). 한편, 감정과 감정표현 강도 사이의 상호작용은 norsqrF0[F(6, 18.045), p<0.001, η_P^2=0.144], norsqrDUR[F(6, 3.360), p<0.01, η_P^2=0.030]에서 유의미한 차이를 보였지만 norsqrINT (p=0.822)에서는 차이가 없었는데 이것은 위에서 언급한 감정으로 인한 운율거리 측정값들의 변화 결과와 같은 경향이다.

앞 절에서 감정의 인지를 가장 잘 설명한 운율거리 측정방법이 유클리디언 운율거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR이었는데 표 2에서 나타난 바와 같이 감정이 운율거리에 미치는 효과크기의 값 그리고 감정표현 강도가 운율거리에 미치는 효과크기의 값이 가장 크게 나타난 측정방법도 norsqrF0, norsqrINT, norsqrDUR이었다. 그리고 효과크기의 순서도 음의 높낮이가 가장 크고 그다음 음의 길이 그리고 효과크기가 가장 작은 음의 강도 순이었는데 이것은 앞 절에서 표 1의 표준화 계수 β가 보여준 바와 같이 감정인지를 잘 설명하고 있는 운율순서와 마찬가지의 결과다. 따라서 본 연구에서는 norsqrF0, norsqrINT, norsqrDUR로 각각의 감정별로 감정표현 강도에 따른 운율거리 측정값들의 다변량 분석을 따로따로 수행하였다.

3.2.1. 차분한(calm) 감정발화의 운율거리

차분한 감정발화에서 감정표현 강도를 높인 결과 norsqrDUR [F(1, 15.489), p<0.001, η_P^2=0.144]에서만 감정중립 발화와 유의미한 차이가 나타났고 norsqrF0(p=0.854), norsqrINT(p=0.220)에서는 차이가 없는 것으로 나타났다(그림 2). 이것은 차분한 음성을 더 잘 표현하기 위해 화자가 음의 높낮이나 음의 강도보다는 단어들의 길이를 늘이는 책략을 사용했다는 것을 의미한다.

3.2.2. 행복한(happy) 감정발화의 운율거리

행복한 감정발화에서 감정표현 강도를 높이면 norsqrF0[F(1, 79.217), p<0.001, η_P^2=0.463]와 norsqrDUR[F(1, 41.792), p<0.001, η_P^2=0.312]에서 감정중립 발화와 유의미한 차이가 나타나고 norsqrINT(p=0.086)에서는 차이가 없는 것으로 나타났다. 이것은 행복한 감정을 더 강하게 표현하기 위해 화자가 음의 강도보다는 음의 높낮이를 높이고 단어들의 길이를 늘이는 책략을 사용했음을 보여준다. 또한, 음의 높낮이가 음의 길이보다는 더 효과가 크게 나타났다.

표 2. 운율거리 측정의 다변량 분석
Table 2. MANOVA of prosody distance measures

요인들	운율거리	*df*	*F*-value	*p*-value	η_P^2
감정	sqrF0***	6	43.057	0.000	0.286
	sqrINT	6	1.275	0.267	0.012
	sqrDUR***	6	17.123	0.000	0.138
	sqrzF0***	6	13.212	0.000	0.110
	sqrzINT	6	0.923	0.478	0.009
	sqrzDUR*	6	2.546	0.019	0.023
	norsqrF0**	6	42.835	0.000	***0.285***
	norsqrINT	6	1.928	0.074	***0.018***
	norsqrDUR**	6	17.590	0.000	***0.141***
	norsqrzF0***	6	16.008	0.000	0.130
	norsqrzINT	6	1.817	0.093	0.017
	norsqrzDUR*	6	2.562	0.018	0.023
감정 표현 강도	sqrF0***	1	284.976	0.000	0.307
	sqrINT***	1	14.552	0.000	0.022
	sqrDUR***	1	73.719	0.000	0.103
	sqrzF0***	1	38.430	0.000	0.056
	sqrzINT*	1	5.812	0.016	0.009
	sqrzDUR	1	0.252	0.616	0.000
	norsqrF0**	1	298.991	0.000	***0.317***
	norsqrINT**	1	18.165	0.000	***0.027***
	norsqrDUR**	1	75.067	0.000	***0.104***
	norsqrzF0***	1	48.450	0.000	0.070
	norsqrzINT**	1	8.430	0.004	0.013
	norsqrzDUR	1	0.259	0.611	0.000

*p<0.05, **p<0.01, ***p<0.001.

3.2.3. 슬픈(sad) 감정발화의 운율거리

슬픈 감정발화에서 감정표현 강도를 높인 결과 norsqrF0[F(1, 25.045), p<0.001, η_P^2=0.214]와 norsqrDUR[F(1, 7.516), p<0.01, η_P^2=0.076]에서 감정중립 발화와 유의미한 차이가 나타나고 norsqrINT(p=0.440)에서는 차이가 없는 것으로 나타났다. 이것은 행복한 감정에서와 마찬가지로 슬픈 감정을 더 강하게 표현하기 위해 화자가 음의 강도보다는 음을 높이고 단어의 길이를 늘이는 책략을 사용했음을 보여주는데 행복한 감정에서와 마찬가지로 음의 높낮이가 음의 길이보다 더 효과가 컸지만, 행복한 감정을 표현할 때보다는 그 변화의 폭이 훨씬 작았다(그림 2).

3.2.4. 화난(angry) 감정발화의 운율거리

화난 감정발화에서 감정표현 강도를 높이면 norsqrF0[F(1, 84.447), p<0.001, η_P^2=0.479], norsqrINT[F(1, 8.263), p<0.01, η_P^2= 0.082] 그리고 norsqrDUR[F(1, 14.738), p<0.001, η_P^2=0.138] 모두에서 감정중립 발화와 유의미한 차이가 나타났다. 이것은 화난 감정을 더 강하게 표현하기 위해 화자가 음의 높낮이, 강도, 길이 모두를 조정하는 방식을 택하였음을 보여주는데 효과크기는 음의 높낮이가 가장 크고 그다음 음의 길이 그리고 음의 강도순인 것으로 나타났다. 그러므로 화난 감정은 음의 강도보다는 단어의 길이 그리고 단어의 길이보다는 음의 높낮이를 더 변화시키는 경향이 크다고 말할 수 있다.

표 3. 감정별 운율거리 측정의 평균과 표준편차
Table 3. Mean and standard deviation of prosody distance per each emotion

운율거리	감정	강도	Mean	SD
dF0	Happy	1	26.395	27.279
		2	114.559	60.288
	Sad	1	0.828	25.858
		2	65.503	69.937
	Angry	1	27.731	37.719
		2	137.356	67.586
	Fearful	1	40.887	37.892
		2	137.582	64.479
	Disgust	1	1.263	27.800
		2	42.593	54.009
	Surprised	1	52.294	35.580
		2	90.115	41.059
dINT	Angry	1	9.261	8.229
		2	22.565	12.636
	Fearful	1	3.213	9.740
		2	16.302	10.579
dDUR	Calm	1	0.022	0.032
		2	0.055	0.044
	Happy	1	0.000	0.025
		2	0.034	0.038
	Sad	1	0.007	0.028
		2	0.031	0.038
	Angry	1	0.033	0.036
		2	0.054	0.049
	Disgust	1	0.034	0.037
		2	0.073	0.042

3.2.5. 두려운(fearful) 감정발화의 운율거리

두려운 감정발화에서 감정표현 강도를 높인 결과 norsqrF0 [F(1, 89.272), p<0.001, η_P^2=0.492], norsqrINT[F(1, 4.469), p<0.05, η_P^2=0.046]에서 감정중립 발화와 유의미한 차이가 나타나고 norsqrDUR(p=0.648)에서는 차이가 없는 것으로 나타났다. 효과크기는 음의 높낮이가 음의 강도보다 훨씬 컸다.

3.2.6. 혐오스러운(disgust) 감정발화의 운율거리

혐오스러운 감정발화에서 감정표현 강도를 높이면 norsqrF0 [F(1, 18.781), p<0.001, η_P^2=0.170]와 norsqrDUR[F(1, 15.324), p<0.001, η_P^2=0.143]에서 감정중립 발화와 유의미한 차이가 나타나고 norsqrINT(p=0.142)에서는 차이가 없는 것으로 나타났다. 효과크기는 음의 높낮이가 음의 길이보다 컸지만, 그 차이가 크지는 않았다. 이것으로 미루어 보면 음의 높낮이가 음의 강도나 음의 길이보다 훨씬 큰 영향을 주었던 다른 감정들과 비교할 때 혐오스러운 감정발화에서는 음의 길이가 상대적으로 중요한 역할을 했다는 설명을 가능하게 한다.

3.2.7. 놀란(surprised) 감정발화의 운율거리

놀란 감정발화에서 감정표현 강도를 높이면 norsqrF0[F(1, 23.468), p<0.001, η_P^2=0.203]만 유의미한 차이가 나타났다. 그리고 norsqrINT(p=0.261)와 norsqrDUR(p=0.084)에서는 차이가 없는 것으로 나타났다.

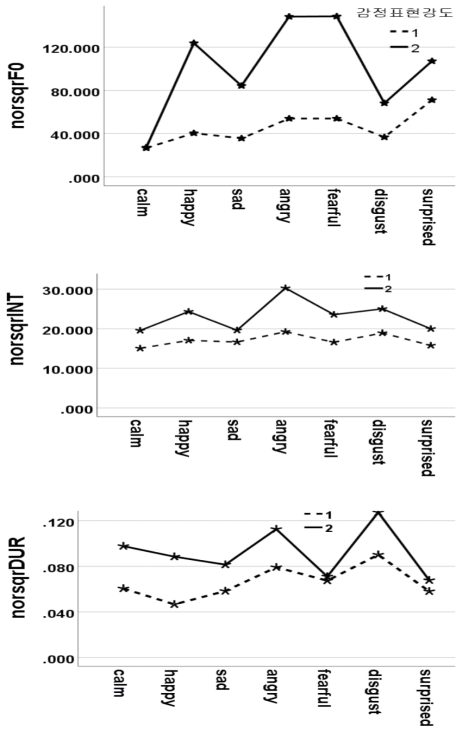

그림 2. 감정별 운율거리 측정
Figure 2. Prosody distance measure per each emotion

3.3. 감정표현 양상의 의미

본 연구의 운율거리 측정법으로 나타난 운율요소들의 감정표현 양상을 설명하기 위해 감정들을 3차원으로 스케일링하여 나타낸 연구(Breazeal, 2003)의 그림 4를 살펴보았다. 그림 4는 각성(arousal)의 정도를 나타내는 세로축과 정서가(valence)를 뜻하는 가로축 그리고 태도(stance)를 의미하는 높이축으로 구성되어 있다. 즉, 각성 값을 나타내는 세로축은 각성의 정도가 높은 감정(high arousal)에서 각성의 정도가 낮은 감정(low arousal)에 이르고 정서 값으로 이루어진 가로축은 정서가가 높은 긍정적인 감정(positive valence)에서 정서가가 낮은 부정적인 감정(negative valence)에 이르며 높이 축에서는 닫힌 태도(closed stance)에서 열린 태도(open stance)에 이르는 연속적인 3차원으로 구성되었다.

본 연구에서 분석한 7개의 감정들을 모두 고려할 때 표 3[1]에 나타난 바와 같이 차분한(calm) 감정을 제외하고 모든 감정에서 음을 높이는 책략으로 감정을 고조시켰음을 알 수 있다. 그리고 화난(angry) 감정과 두려운(fearful) 감정에서만 음의 강도(dB)가 유의미하게 증가했다. 이들은 각성(arousal)의 정도가 높고 부정적 정서가(negative valence)가 높은 감정(그림 4)들인데 부정적 감정에서 각성의 정도가 증가할 때 비로소 음의 강도를 높이는 책략이 사용되는 것으로 보여진다. 그리고 각성(arousal)의 정도가 높고 긍정적이지 않으며 닫힌 태도(closed stance)가 아닌 두려운(fearful) 감정과 놀란(surprised) 감정(그림 4)에서만 음의 길이의 효과가 유의미하지 않았음을 알 수 있는데 이것은 긍정적이지 않은 감정이 상황을 거부하는 방식으로 외부로부터의 영향을 최소화하지 못한 상태에서 각성의 정도가 증가하면 단어를 느리게 발화할 여유가 없어지기 때문으로 추정된다.

3.4. 감정음성의 사후검증

본 연구의 분석 대상인 7개 감정들에 대해 Tukey 사후검증을 시행한 결과, 감정들은 norsqrF0에 의해 calm<disgust, sad<happy, surprised<surprised, angry fearful의 4개 동일군으로 나누어졌다. 그리고 norsqrDUR에 의해 surprised, happy, fearful, sad, calm<calm, angry<angry, disgust의 3개 동일군으로 나누어졌다.

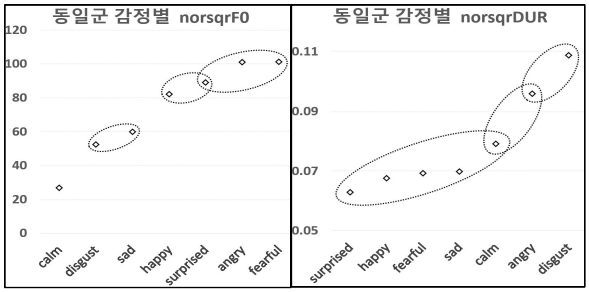

그림 3. 운율거리에 의한 동일군 감정그룹들(Alpha=0.05)
Figure 3. Emotion groups in homogeneous subsets by prosodic distance (Alpha=0.05)

그림 3과 그림 4에서 norsqrF0의 경우, 각성의 정도가 높지만 긍정적이지는 않은 3개의 감정 fear, angry, surprised가 한 동일군을 형성하고 있는데 이들 중 각성의 정도가 높고 부정적인 감정이 큰 fear, angry에서만 음의 강도(dB)가 유의미하게 증가했다. 그다음 동일군을 이룬 surprised와 happy(joy)는 각성의 정도가 높고 부정적이지 않다는 특징이 있고 그다음 sad(sorrow), disgust로 이루어진 동일군은 각성의 정도가 낮고 정서가가 낮은 부정적 감정이라는 공통점을 갖고 있다. 종합하면, 긍정적이지 않고 각성의 정도가 높을 때 음높이를 가장 극대화하고 그다음 경우가 부정적이지 않고 각성의 정도가 높을 때이고 가장 음높이를 소극적으로 활용하는 경우가 부정적 감정이면서 각성의 정도가 낮은 경우라 말할 수 있다.

norsqrDUR의 경우, 닫힌 태도를 취하면서 부정적 정서가가 높은 disgust, angry의 동일군 그룹과 calm을 중간에 두고 배치된 2개의 동일군 그룹들로 구성되어 있다. 닫힌 태도(closed stance)가 아닌 fearful과 surprised의 감정들이 상황을 거부하는 방식으로 외부로부터의 영향을 최소화하지 못해 느리게 발화할 여유가 없었던 것과는 대조적으로 본 연구의 7가지 감정들 중 유일하게 닫힌 태도(closed stance)를 보이는 disgust와 angry는 발화길이를 가장 잘 극대화해서 감정표현에 이용한 것으로 보인다.

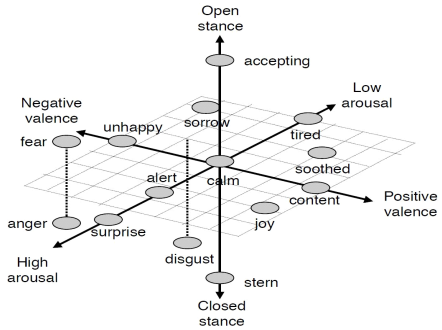

그림 4. 각성, 정서가, 태도 공간에 매핑된 감정 카테고리들(Breazeal, 2003)
Figure 4. Mapping of emotional categories to arousal, valence, and stance dimensions (Breazeal, 2003)

4. 결론

본 연구는 화자의 감정이나 태도의 전달에 중요한 역할을 담당하는 운율이 감정발화에 나타나는 양상을 운율거리 계산 기법을 가지고 분석하였다. 본 연구에 사용된 운율거리 측정방법들(sqrF0, sqrINT, sqrDUR; norsqrF0, norsqrINT, norsqrDUR; sqrzF0, sqrzINT, sqrzDUR; norsqrzF0, norsqrzINT, norsqrzDUR) 중 7가지 분석 대상인 감정들(calm, happy, sad, angry, fearful, disgust, surprised)이 얼마나 잘 표현되어 인지되었는지를 나타내는 Goodness score를 가장 잘 설명하고 감정들과 감정표현 강도에 가장 큰 효과크기로 반응한 측정방법은 유클리디언 운율거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR이었다.

유클리디언 거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR을 이용하여 각 감정별로 분석한 결과, 전반적으로(calm을 제외한 감정들에서) 감정에 따른 운율의 변화는 음의 높낮이(Hz)에서 가장 컸고 그다음 음의 길이(sec) 그리고 음의 강도(dB)에서

---

1 유클리디언 거리 측정법은 제곱을 하기 때문에 음양(+ −)의 정보가 사라지므로 *dF0, dINT, dDUR*을 사용하여 변화가 양의 방향인지 음의 방향인지 알 수 있도록 하였다. 그 수식들은 다음과 같다.

$$dF0 = \sum_{i=1}^{n}(F0i_{\neq ut} - F0i_{emo})/n,\ dINT = \sum_{i=1}^{n}(INTi_{\neq ut} - INTi_{emo})/n,\ dDUR = \sum_{j=1}^{m}(DURj_{\neq ut} - DURj_{emo})/m$$

가장 작았다. 본 연구의 결과를 기존연구(Breazeal, 2003)와 연결지어 살펴볼 때, 음의 강도가 angry와 fearful에서만 유의미하게 증가한 것은 부정적 감정에서 각성의 정도가 증가할 때 비로소 음의 강도를 높이는 책략이 사용되는 것으로 해석할 수 있다. 그리고 fearful과 surprised에서만 음의 길이의 효과가 유의미하지 않았는데 이것은 닫힌 태도(closed stance)의 방식으로 외부로부터의 영향을 상쇄하지 못한 상태에서 각성의 정도가 증가하면 단어를 느리게 발화할 여유가 없어지기 때문으로 추정된다. 그리고 감정음성의 사후분석으로 시행된 동일군 분석은 운율거리 계산에 의한 감정분류의 가능성을 보여주었는데 본 연구는 이 모든 것들이 그림 4에서 나타난 바와 같은 3차원 감정분류에 의해서 일관성있게 설명될 수 있음도 보여주었다.

본 연구는 감정발화를 해당 발화에서 추출한 음향 자질들로 분류하던 기존의 감정발화 분석과는 달리 외국어 발음 자동평가에 사용된 운율거리 측정기법을 적용하여 감정중립 발화와 감정발화 간의 운율적 거리를 계산함으로써 감정발화를 분석 및 분류할 수 있다는 가능성을 보여주었다. 그리고 어떤 운율거리 측정방법이 더 효율적으로 감정발화 분석에 쓰일 수 있는지를 논하였고 효과적인 운율거리 측정기법으로 구한 결과가 감정음성 분석의 기존연구와 맥락을 같이하고 있음을 보여주었다. 다만, 본 연구에 사용된 데이터베이스는 자연발화와는 다른 의도된 발화여서 해석에 한계가 있음을 밝힌다. 향후, 음성을 통해 감정을 인식하려는 응용 분야의 감정음성 인식률 향상을 위해, 감정음성의 억양 곡선에서 추출한 자질들로 감정분류를 시도한 기존연구(Yi, 2018)와 스펙트럼 거리를 이용한 감정음성 분류방법을 본 연구와 함께 적용할 필요가 있을 것으로 생각한다.

References

Bachorowski, J., & Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6*(4), 219-224.

Boersma, P., & Weenink, D. (2020). Praat: Doing phonetics by computer (version 6.1.16) [Computer program]. Retrieved from https://www.praat.org/

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*(1), 119-155.

Cahn, J. (1990). *Generating expression in synthesized speech* (Technical report). Boston, MA: MIT Media Lab.

Carlson, R., Granström, B., & Nord, L. (1992, October). Experiments with emotive speech: Acted utterances and synthesized replicas. *Proceedings of the International Conference on Spoken Language Processing (ICSLP-92)* (pp. 671-674). Banff, AB, Canada.

Kitahara, Y., & Tohkura, Y. (1992). Prosodic control to express emotions for man-machine interaction. *IEICE Transactions on Fundamentals of Electronics: Communications and Computer Sciences*, *75*(2), 155-163.

Kitayama, S., & Ishii, K. (2002). Word and voice: spontaneous attention to emotional utterances in two languages. *Cognition and Emotion, 16*(1), 29-59.

Livingstone, S. R., & Russo, F. A. (2018). The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS): A dynamic, multimodal set of facial and vocal expressions in North American english. *PLOS ONE, 13*(5), e0196391.

Mozziconacci, S. (1998). *Speech variability and emotion: Production and perception* (Doctoral dissertation). Technical University of Eindhoven, The Netherlands.

Paulmann, S. (2016). The neurocognition of prosody. In G. Hickok, S. Small (Eds.), *Neurobiology of language* (pp. 1109-1120). San Diego, CA: Elsevier.

Pell, M., Paulmann, M., Dara, S., Alasseri, A., & Kotzb, S. (2009). Factors in the recognition of vocally expressed emotions: A comparison of our languages. *Journal of Phonetics, 37*(4), 417-435.

Thompson, W. F., & Balkwill, L. L. (2009). Cross-cultural similarities and differences. In P. N. Juslin, J. A. Sloboda (Eds.), *Handbook of music and emotion: Theory, research, applications, 1st Edn,* 755-791. New York, NY: Oxford University Press.

Vroomen, J., Collier, R., & Mozziconacci, S. (1993)., September). Duration and intonation in emotional speech. *Proceedings of the 3rd European Conference on Speech Communication and Technology. Eurospeech-93*, 577-580. Berlin, Germany.

Williams, C. & Stevens, K. (1972). Emotions and speech: Some acoustical correlates. The *Journal of the Acoustical Society of America*, *52*(4B), 1238-1250.

Yi, S. P. (2018). Study on pitch contour extracted from Korean emotional speech using momel, *Journal of Language Sciences, 25*(3), 191-209.

Yoon, K. (2009a). Building a sentential model for automatic prosody evaluation. *Phonetics and Speech Sciences, 1*(4), 47-59.

Yoon, K. (2009b). Synthesis and evaluation of prosodically exaggerated utterances. *Phonetics and Speech Sciences, 1*(3), 73-85.

Yoon, K. (2013). A study on human evaluators using the evaluation model of english pronunciation. *Phonetics and Speech Sciences, 5*(4), 109-119.

• **이서배 (So Pae Yi)** 교신저자
Humphreys West Elementary School 한국어과 정교사
Department of Defense Education Activity
Camp Humphreys, Pyeongtaek
Tel: 031-8029-0444
Email: sopaeyi@pusan.ac.kr
관심분야: 음성학, 음운론, 음향음성학, 음성공학

영어 감정발화와 중립발화 간의 운율거리를 이용한 감정발화 분석

이 서 배

미국방성 교육부, 험프리스 웨스트 초등학교

국문초록

본 연구는 영어 발화에 나타난 7가지 감정들(calm, happy, sad, angry, fearful, disgust, surprised)을 분석하고자 감정발화(672개)와 감정중립 발화(48개)와의 운율적 거리를 측정하였다. 이를 위해 외국어 발음평가에 사용되었던 방법을 적용하여 음의 높낮이(Hz), 음의 강도(dB), 음의 길이(sec)와 같은 운율의 3요소를 유클리디언 거리로 계산하였는데 기존연구에서 더 나아가 유클리디언 거리계산 정규화 방법, z-score 방법 그리고 z-score 정규화 방법을 추가해 총 4가지 그룹(sqrF0, sqrINT, sqrDUR; norsqrF0, norsqrINT, norsqrDUR; sqrzF0, sqrzINT, sqrzDUR; norsqrzF0, norsqrzINT, norsqrzDUR)의 방법을 분석에 사용하였다. 그 결과 인지적 측면과 음향적 측면의 분석 모두에서 유클리디언 운율거리를 정규화한 norsqrF0, norsqrINT, norsqrDUR이 일관성 있게 가장 효과적인 측정방법으로 나타났다. 유클리디언 거리계산 정규화 방법으로 감정발화와 감정중립 발화를 비교했을 때, 전반적으로 감정에 따른 운율의 변화는 음의 높낮이(Hz)가 가장 크고 그다음 음의 길이(sec), 그리고 음의 강도(dB)가 가장 작게 나타났다. Tukey 사후검증 결과 norsqrF0의 경우 calm<disgust, sad<happy, surprised<surprised, angry, fearful의 4개 동일군으로 나누어졌고 norsqrDUR의 경우 surprised, happy, fearful, sad, calm<calm, angry<angry, disgust의 3개 동일군으로 나누어졌다. 그리고 각 감정별 세부분석은 본 연구의 결과가 기존연구와 맥락을 같이함을 보여주었다.

핵심어: 감정발화, 운율, 유클리디언 거리, 각성, 정서가, 태도

참고문헌

윤규철 (2013). 영어 발음 평가 모델을 활용한 수동 평가자 연구, *말소리와 음성과학, 5*(4), 109-119.

이서배 (2018). 한국어 감정 음성에서 모멜로 추출한 피치 곡선 연구, *언어과학*, *25*(3), 191-209.