

미니언즈 음성 분석을 통한 감정인식 연구

윤 광 열¹⁾, 전 인 서²⁾, 이 인 구³⁾, 김 윤 정⁴⁾, 신 지 영⁵⁾
고려대학교 국어국문학과¹⁾²⁾³⁾⁵⁾, 루이지애나 주립대학교 의사소통 과학 및 장애학과⁴⁾

A Study of Emotion Recognition through Analyzing the Sounds of Minionese

GwangYeol Yun¹⁾, InSeo Jeon²⁾, InGu Yi³⁾, YunJung Kim⁴⁾, JiYoung Shin⁵⁾
Korean Language and Literature, Korea University¹⁾²⁾³⁾⁵⁾
Department of communication sciences and disorders, Louisiana State University⁴⁾
yoon853618@naver.com¹⁾, jisjsjs@korea.ac.kr²⁾, demiust@gmail.com³⁾,
ykim6@lsu.edu⁴⁾, shinjy@korea.ac.kr⁵⁾

Abstract

The purpose of this study was to understand the relationship between listeners' emotion recognition and prosodic characteristics of speech. Fifty listeners were asked to judge the degree of valence and arousal of fifteen meaningless sounds of Minions. Results indicated that the arousal was correlated with pitch, while the valence was with other factors such as vocal quality.

1. 서론

영화 <미니언즈>에 등장하는 가상 언어 '미니언어'는 불어, 영어, 스페인어, 러시아어, 한국어 등 다양한 언어에서 우스꽝스럽고 헛설수설하는 소리들을 차용하여 만든 비체계어이다. 영화 <미니언즈>가 미니언즈에 자막을 사용하지 않았음에도 불구하고, 영화 관객들은 이 무의미한 음성의 연쇄를 듣고 감정을 인식하는 데 어려움을 겪지 않는다. 본 연구는 이것이 의미의 영향을 배제하고 소리 자체의 특성에 주목할 수 있는 예시라고 생각했다. 이에 미니언즈 말소리의 초분절적 요소와 청자의 감정인식 간의 상관관계를 살펴보고자 한다.

현재 감정인식 분야에서 많이 사용되고 있는 감정도넬은 크게 두 가지로, 첫 번째는 인간의 감정을 2차원 영역에 표현한 Valence-arousal Model이고, 두 번째는 몇 가지의 대표 감정을 선정하는 방법이다. 대부분의 음성기반 감정인식 분야에서는 표현의 명확성과 분류의 용이성을 위하여 후자를 취하고 있는데[1], 이 방법은 감정 선정에 대한 명확한 기준이 부족하고, 감정의

정도성 및 다양성을 포괄하지 못한다는 한계가 있다. 이에 본 연구는 전자를 택하여, 광범위한 감정을 두개의 축을 이용해 효율적으로 나타낼 수 있는 Russel의 이론[2]을 참조하였다.

2. 본론

본 연구의 감정인식 실험은 다음과 같이 진행하였다. 20대 성인 50명(남성, 여성 각 25 명)으로 피험자 집단을 구성하고, 소음을 통제된 상황에서 15개의 미니언즈 음성을 두 번씩 들려주었다. 그리고 음성을 듣고 인식한 감정을 10x10 크기의 좌표(x축: 유쾌-불쾌, y축: 각성-비각성)에 직관적으로 표시하도록 하였다. 수집한 실험 자료를 바탕으로, 음성 샘플별로 하나의 좌표평면에 50개의 응답을 모아 집약한 후 이를 분면별로 분석하였다.

다음으로, 분석할 초분절적 요소로 말속도(speed), 음도(pitch - mean, min, max)를 설정하고 praat을 사용하여 실험에 사용된 미니언즈 음성을 분석하였다. (말속도는 초당 음절수(syl/sec)로 나타내었고, 발화 단위 내 1.5초 이하의 휴지 및 머뭇거림, 또는 비유창성의 시간은 포함하였다.) 분석 결과 전체 미니언즈의 말속도는 평균 5.04syl/sec로 한국어인 표준어 화자의 평균 말속도 4.13syl/sec[3]에 비해 빠른 편이며, 음도 또한 평균 318Hz로 높은 수치를 보이고 있다. 이러한 빠른 말속도와 높은 음도는 일반적으로 분노(각성-불쾌) 또는 기쁨(각성-유쾌)의 감정에서 드러나는 특징이다.[2][4]

그림 1은 감정인식 실험 결과의 한 예시로, 피험자들이 4번 음성을 듣고 유쾌-불쾌, 각성-비각성의 정도를 표시한 결과이다.

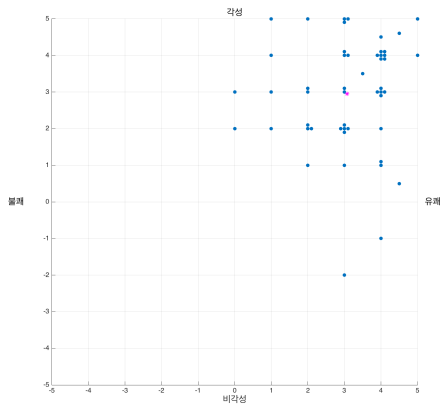


그림 1. 4번 음성의 감정인식 실험 결과

평균값은 x축 3.1, y축 2.97로, 대부분의 피험자들이 4번 음성을 유쾌-각성으로 인식했다는 사실을 알 수 있다.

표 1. 전체 음성의 음도와 감정인식 실험 결과

번호	pitch			감정인식	
	mean	min	max	유쾌도(x축)	각성도(y축)
1	314.1	126.8	488.3	-0.15	2.3
2	315.8	78.2	520.7	1.99	3.11
3	313.3	90.6	514.1	-3.34	2.77
4	339.2	225.2	523.2	3.1	2.97
5	304.4	77.5	519.9	2.14	1.98
6	157.3	70.5	421.1	3.06	0.22
7	127	75	172.9	-2.18	-0.48
8	312.4	119.8	529.4	-1.99	3.24
9	280	98.7	514.6	-1.59	2.37
10	317.6	266.3	494.1	-1.69	2.29
11	273.8	81.9	470.3	-2.33	1.32
12	362.8	189.9	528.8	0.34	2.84
13	287.5	81	429.6	0.79	-2.18
14	344.3	217.5	523.1	-0.23	3.3
15	303.5	74.7	474	-2.4	-0.3
평균	290.2	124.9	474.9	-0.3	1.72

표 1은 전체 15개 음성의 평균, 최소, 최대 음도와 피험자들의 감정인식 실험 결과를 보인 것이다. 실험 분석 결과, 청자들은 음성의 음도 최소값이 81Hz보다 작고 최대값이 474Hz보다 작을 때(6, 7, 13, 15번) 비각성으로 인식하는 경향을 보였다. 또한 음도의 순간적인 변화(기울기 200Hz/msec 이상)가 나타날 때 각성으로 인식하였다.

그림 2의 구간(1)에는 0.91초 동안 동일 음절 [케]가 네 번 반복되는 말더듬이, 구간(2)에는 한 어절 내에서 급격한 음도의 변화가 드러나고 있다.

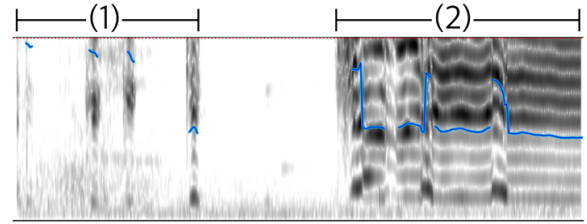


그림 2. 10번 음성의 스펙트로그램

이렇듯 짜내기 발성(creaky voice), 비유창성으로 이어지는 동일 음절의 반복인 말더듬, 짧은 시간 내 급격한 음도 변화를 반복적으로 동반하는 떨림, 과장된 들숨 등은 청자가 불쾌를 인지하게 하는 중요한 표지로서 작용한다.

3. 결론

본 연구는 미니언즈 음성을 연구 자료로 선정함으로써 의미를 배제하고 온전히 소리의 특성에 집중했다는 점에 의의가 있다. 또한 유쾌-불쾌, 각성-비각성의 축에 따라 피험자가 감정을 수치화하여, 청자가 인식한 감정과 음성의 초분절적 요소 사이의 연관성을 더욱 객관적으로 파악할 수 있었다.

연구 결과 음도와 각성도 간의 뚜렷한 상관관계를 발견할 수 있었다. 그리고 말속도, 음도로 구분되지 않는 유쾌도는 발성유형의 변화, 말더듬, 떨림, 과장된 들숨과 같은 표지들에 의해 구분된다는 사실 또한 발견하였다.

참고문헌

- [1] 방재훈, 이승룡. (2014). 감성기반 서비스를 위한 통화 음성 감정인식 기법, 정보과학회논문지, 41, 3, 210.
- [2] Russell, J. A. (1989). Measures of emotion. In R. Plutchik & H. Kellerman (Eds.), Emotion: Theory, research, and experience, 4, 83-111.
- [3] 심홍인. (2004). 한국 표준어 화자의 유창성과 말속도에 관한 연구, 음성과학, 11, 3, 198.
- [4] Yildirim, S., Bulut, M., Lee, C., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., and Narayanan, S. (2004). An acoustic study of emotions expressed in speech, INTERSPEECH 2004 - ICSLP, 2193-2196.