

Development and validation of a Korean Affective Voice Database

Yeji Kim · Hyesun Song · Yesol Jeon · Yoorim Oh · Youngmee Lee*

Department of Communication Disorders, Ewha Womans University, Seoul, Korea

Abstract

In this study, we reported the validation results of the Korean Affective Voice Database (KAV DB), an affective voice database available for scientific and clinical use, comprising a total of 113 validated affective voice stimuli. The KAV DB includes audio-recordings of two actors (one male and one female), each uttering 10 semantically neutral sentences with the intention to convey six different affective states (happiness, anger, fear, sadness, surprise, and neutral). The database was organized into three separate voice stimulus sets in order to validate the KAV DB. Participants rated the stimuli on six rating scales corresponding to the six targeted affective states by using a 100 horizontal visual analog scale. The KAV DB showed high internal consistency for voice stimuli (Cronbach's $\alpha=.847$). The database had high sensitivity (mean=82.8%) and specificity (mean=83.8%). The KAV DB is expected to be useful for both academic research and clinical purposes in the field of communication disorders. The KAV DB is available for download at <https://kav-db.notion.site/KAV-DB-7539a36abe2e414cbf4a50d80436b41a>.

Keywords: prosody, affective voices, emotion recognition

1. 서론

인간은 음성으로 감정을 정확하게 표현하고 대화 상대방의 감정 음성(affective voice)을 정확하게 이해하는 능력을 통해, 사회적 상호작용을 촉진하고 친밀한 대인 관계를 형성한다(Carstensen et al., 1999; Kang et al., 2014). 인간은 상대방의 감정을 이해하거나 표현할 때 언어적 단서(linguistic cues)와 비언어적 단서(nonlinguistic cues) 정보를 활용한다. 언어적 단서란 화자의 메시지가 글자로 기록되어 그 의미가 정확하게 전달될 수 있는 단

서이며, 비언어적 단서는 글자로 기록될 수 없는 표정, 제스처, 음도, 운율, 억양 등을 말한다(Planalp, 1996). 이 중에서, 운율(prosody)은 음도, 강도, 말속도 등을 통해 파악할 수 있는 요소로, 상대방의 감정을 인식하는 데 큰 역할을 한다고 알려져 있다(Coutinho & Dikken, 2013; Planalp, 1996; Trainor et al., 2000). 즉, 청자(listener)가 화자(speaker)의 감정을 이해한다는 것은 단어와 문장 등의 언어적 단서와 결합된 준언어적 요소인 운율을 파악하는 것을 의미한다. 그런데 명시적인 언어적 단서가 없어도 비언어적 단서인 운율만으로도 감정이 전달될 수 있다

* youngmee@ewha.ac.kr, Corresponding author

Received 27 July 2022; Revised 15 August 2022; Accepted 15 August 2022

© Copyright 2022 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

(Trainor et al., 2000). 특히, 표정, 제스처와 같은 시각적인 정보가 없는 상황에서는 화자 발화의 언어적 의미와 운율정보에 초점을 맞추어야 하므로, 화자의 발화에서 운율을 정확하게 파악하는 것은 의사소통에 중요하다.

발화의 운율적 특성을 감지하고 이를 바탕으로 감정을 인식하는 능력은 영유아기 때부터 발달된다(Trainor et al., 2000). 학령전기 아동은 상황 맥락과 내용을 토대로 감정을 판단하지만, 학령기 아동은 상황 맥락, 문장 의미, 운율정보를 통합적으로 파악하여 화자의 감정을 판단한다(Kang et al., 2014; Park, 2010). 의사소통에서 운율의 중요성이 부각되면서, 언어병리학, 특수교육, 의학 등의 분야에서 다양한 의사소통장애군을 대상으로 운율 인식 연구가 진행되었다. 그 결과, 자폐범주성장애군(Filipe et al., 2014; Oerlemans et al., 2014), 뇌혈관 사고나 교통사고로 우뇌손상을 경험하는 환자군(Sim, 2007)에서 운율 단서 이해 및 표현에 어려움을 겪는 것을 확인하였다. 난청 아동의 운율 인식 및 표현에 관한 선행연구에서도 보청기나 인공와우 착용 후에도 문법적인 운율과 감정적 운율 모두에서 인식과 표현에 어려움을 보인다고 보고하였다(Choi & Lee, 2009; Peng et al., 2008). 국내에서도 운율인식에 관한 연구가 진행되었으나(Lee & Kim, 2013; Shin et al., 2015; Sim, 2007), 다양한 연령대의 일반 집단이나 의사소통장애 집단을 대상으로 한 운율 관련 연구가 충분히 이루어지지 않고 있다. 일부 연구에서 특정 장애군의 운율 특성을 다루고 있을 뿐 평가나 중재 연구는 지극히 드문 편이다(Park et al., 2016). 운율 인식, 운율 인식 능력, 운율 인식 중재 연구를 위해서는 다양한 감정이 잘 표현된 음성 데이터베이스(database, DB) 구축이 필요하다. 또한, 운율 인식 과제의 형태나 제시 감정 유형에 따라 연구자 간 다른 결과를 보고한다는 점(Kim et al., 2013)을 고려하였을 때, 운율 관련 능력을 정확하게 평가할 수 있는 신뢰롭고 타당한 한국형 감정 음성 데이터베이스(Korean Affective Voice Database, KAV DB)가 개발될 필요가 있다.

국내에서는 감정 음성 DB 구축에 관한 연구를 찾기 어려우나, 국외에서는 관련 연구와 DB 구축이 진행된 바 있다. 국외 데이터 중에서, 'The Montreal Affective Voices(MAV)'는 인지신경과학 연구에 많이 사용되는 Ekman faces(Ekman & Friesen, 1978)의 분노, 역겨움, 공포, 행복, 슬픔, 놀람, 중립의 감정을 불러 모음인 ah(/ɑ/)로 녹음하여 제작한 음성 DB이다(Belin et al., 2008). 그리고 국외에서는 의미적으로 중립적인 문장에 감정을 담아 녹음하거나(Imaizumi et al., 1997; Kotz et al., 2003; Laukka, 2005), 의미가 없는 비단어(Banse & Scherer, 1996; Grandjean et al., 2005)를 이용하여 감정 음성 DB가 개발되기도 하였다. 사회, 문화, 언어에 따라 감정을 표현하고 인식하는 것에는 문화 간에 차이가 있으므로, 한국인을 대상으로 국외 감정 음성 DB를 사용하여 감정 및 운율 인식 연구를 진행하기에는 무리가 있다(Laukka & Elfenbein, 2021; Laukka et al., 2016). 문화에 따른 감정 인식 차이를 살펴본 Kitayama & Ishii(2002)는 영국인은 음성에서 발화의 감정과 관련된 운율정보보다 내용에 집중하는 반면에, 일본인은 발화의 내용보다는 운율정보에 더 집중한다는

것을 확인하였다. Tanaka et al.(2010)은 네덜란드인보다 일본인이 운율정보에 더욱 민감하다고 하였다.

이처럼 문화와 언어에 따라 감정 인식에 대한 시각적 정보와 청각적 정보의 처리에 차이가 나타난다는 것을 고려할 때(Ekman et al., 1987; Elfenbein & Ambady, 2002; Scherer et al., 2001), 한국인의 감정, 문화가 잘 반영된 국내 감정 음성 DB 개발이 필요하다고 볼 수 있다. 이에 본 연구에서는 한국인을 대상으로 한 감정 음성 DB를 개발하고, 해당 DB의 타당도를 신뢰도, 민감도(sensitivity) 및 특이도(specificity) 차원에서 검증하고자 하였다. 이를 통해, 음성 속 감정 인식 능력 향상을 위한 임상 콘텐츠의 기초자료를 제공하고, 다양한 의사소통장애군을 대상으로 한 음성 속 감정 인식 능력 연구에 활용할 수 있는 자료를 제공하고자 한다.

2. 연구방법

2.1. 발화 문장 제작

본 연구에서 사용된 발화 문장은 의미적인 중립성, 한국어 문법, 어휘의 영향을 최소화하기 위해서 다음과 같은 기준에 따라 문장을 제작하였다. 발화 문장은 (1) 3-4어절의 단문으로, (2) 어절당 음절 수는 2-5음절로 제한하였다. 문장에 사용된 어휘는 Seo(2021)의 '2021년 국어 기초 어휘 선정 및 어휘 등급화 연구'에 수록된 1등급 어휘 목록의 어휘만으로 선정하였다. Seo(2021)는 일상 언어생활 속에 필요한 기초 어휘 중, 가장 필수적이고 많이 쓰이는 어휘를 1등급으로 간주하였으며, 4-6세의 아동이 사용하는 어휘를 1등급 어휘로 제안하였다. 이러한 기준에 따라서 총 40개의 문장을 제작하였으며, 해당 문장이 의미적으로 감정이 표현되지 않은 중립적인 문장인지 알아보기 위해서 언어병리학 석·박사과정생 6인과 언어병리학 교수 1인을 대상으로 5점 리커트 척도(1점: 의미적으로 매우 부정적이다, 2점: 의미적으로 부정적이다, 3점: 의미적으로 중립적이다, 4점: 의미적으로 긍정적이다, 5점: 의미적으로 매우 긍정적이다)로 타당도 검정을 실시하였다. 이때, '의미적으로 중립적이다'를 나타내는 3점을 기준으로 편차값을 구하여 의미적으로 중립에 가깝다고 평가된 상위 문장 10개를 최종적으로 선정하였으며, 선정된 문장은 표 1과 같다.

표 1. 녹음에 사용된 발화 문장 목록
Table 1. List of sentences used for recording

No.	Sentence
1	작은 방에 들어갔어
2	비누는 화장실에 있어
3	문 앞에 있어
4	미끄럼틀을 타고 있어
5	그건 할아버지 물건이야
6	방에 불을 켜어
7	연필은 왼쪽 서랍에 있어
8	파란 버튼을 눌러봐
9	밭에서 고추를 땀어
10	컴퓨터는 저쪽 방에 있어

2.2. 녹음

연구자는 배우(남 1명, 여 1명)에게 10개의 문장을 6가지 감정(행복, 분노, 공포, 슬픔, 놀람, 중립)으로 표현되도록 발화해달라고 요청하였다. 배우는 모두 뮤지컬 동호회 소속으로 연기 경력은 평균 5년이었으며, 연령은 각각 27세와 29세로 평균 28세였다. 녹음은 소음이 통제된 방음 부스에서 디지털 레코더(Roland EDIROL R 05HR)와 핀 마이크(ECM CS10)를 이용하였다. 이때, 마이크는 배우의 입과 20 cm의 거리를 두어 녹음하였다. 녹음 과정에서 연구자가 실시간으로 배우의 발화를 모니터링하면서 배우가 목표 감정을 제대로 표현하지 못하였다고 판단되었을 경우, 추가적으로 2-3회 다시 발화해달라고 요청하였다. 이러한 녹음 작업을 통해, 총 246개의 음성 파일이 생성되었다. 감정별 음성 샘플은 그림 1과 같다.

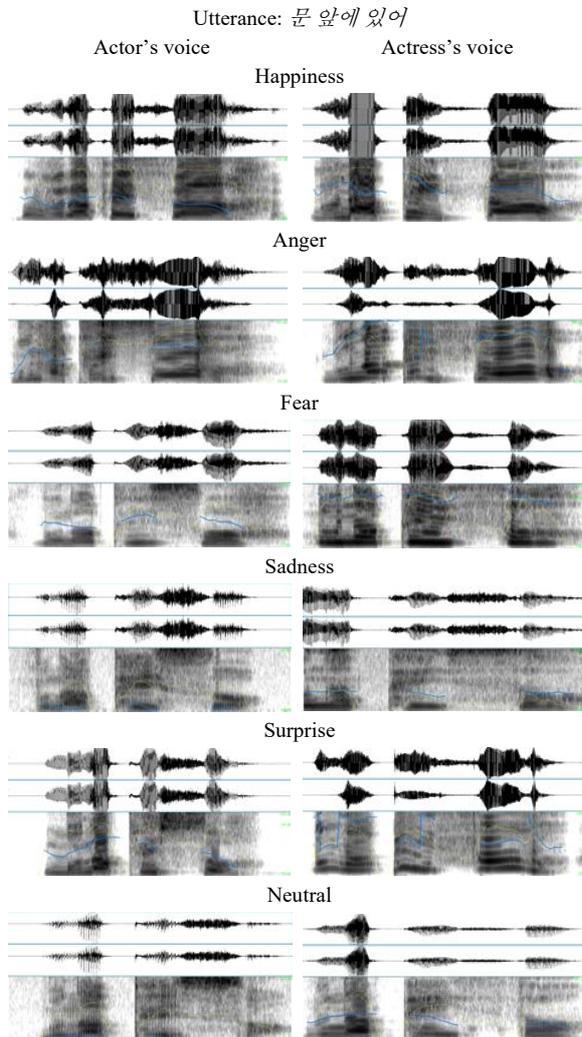


그림 1. 감정별 음성샘플 예시

Figure 1. Example of speech sample for 6 affective states

2.3. 음성 감정 발화 선정

녹음된 음성이 목표로 한 6가지 감정을 잘 표현하였는지를 언어병리학 석사과정생 4명이 평가하여, 각 음성에 표현된 감

정에 대해 모두 일치된 견해를 보이는 음성을 선정하였다. 그 결과, 행복 음성 18개, 분노 음성 20개, 공포 음성 18개, 슬픔 음성 18개, 놀람 음성 19개, 중립 음성 20개, 총 113개의 음성이 타당도 평가를 위한 음성으로 선정되었으며, 감정별 음성 파일 개수는 표 2와 같다.

표 2. 감정별 발화 문장 수

Table 2. The number of sentences per emotion

Affective states	Actor's voice	Actress's voice	Total
Happiness	9	9	18
Anger	10	10	20
Fear	9	9	18
Sadness	10	8	18
Surprise	9	10	19
Neutral	10	10	20

2.4. 타당도 평가

2.4.1. 대상자

타당도 평가에 참여한 대상자는 한국어를 사용하는 20대 성인 31명이었으며, 대상자의 평균 연령과 성비는 표 3과 같다. 대상자 선정 기준은 (1) 정상 청력이며, (2) 한국어가 모국어이고, (3) 최종 학력이 고등학교 졸업 이상으로, (4) 과제를 수행하는 데 필요한 시각에 문제가 없으며, (5) 언어 및 인지적 병력이 보고되지 않았으며, (6) 정신·신경학적 병력이 없는 자를 대상으로 하였다.

표 3. 대상자 정보

Table 3. Participants information

Participants		
Age (years)	Mean	26.9
	SD	1.6
	Ranges	24-29
Gender	Male	15
	Female	16

2.4.2. 절차

2.4.2.1. 타당도 반응 자료 제작

녹음자료는 Adobe Audition(ver. 22.4)을 이용하여 편집하였다. 순서효과와 학습효과를 배제하기 위해 배우의 성별과 감정 유형을 고려하여, 청자에게 들려준 음성 파일을 무선회(randomization)하여 3개의 블록에 각각 38개, 38개, 37개씩 배치하였다(A블록: 38개, B블록: 38개, C블록: 37개). 그리고 해당 블록의 제시순서를 역균형화(counterbalancing)하여, A-B-C, B-C-A, C-A-B 블록으로 구성된 총 3개의 음성 파일 세트를 구성하였다. 이때, 신뢰도 평가를 위한 음성 파일을 각 블록에 4개씩 총 12개를 배치하였다.

2.4.2.2. 감정 평가

연구자는 소음이 차단된 조용한 공간에서 노트북(Gram 17ZD990, LG Inc., Seoul, Korea)을 통해 대상자에게 음성 파일을 들려주었다. 청자는 시각 아날로그 평정(visual analogue scale)을 이용하여, 본인이 지각한 감정을 그림 2와 같이 각 감정 척도에 0에서 100점으로 표시하였다. 대상자의 평안한 청취를 위해, 들려주는 음성 강도 조절은 스스로 조정하도록 하였다. 이때, 대상자에게 휴식이 필요한 경우, 언제든지 휴식을 취할 수 있도록 하였다.

한국 감정 음성 데이터베이스 구축 설문 (1)

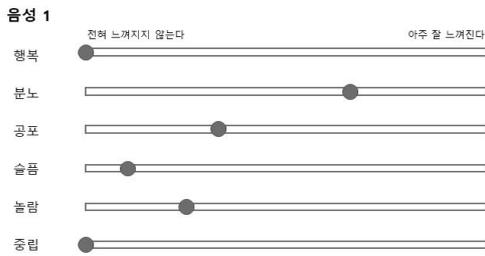


그림 2. 시각 아날로그 평정 예시
Figure 2. Example of visual analog scale

2.5. 신뢰도 분석

신뢰로운 감정 음성 DB 구축을 위해서는 청자가 감정 음성 평정 과제에 집중하여 잘 참여하는 것이 중요하므로, 과제 참여의 신뢰성을 확인하기 위한 평가 음성을 12개 삽입하였다. 평가 음성은 6개의 감정별로 2개씩, 총 12개의 문장으로 구성되어 있으며, 의미와 운율이 일치되고 간투사와 감탄사를 추가한 문장을 사용하였다(부록 1). 본 연구에서는 해당 평가 음성에서 정반응률이 80% 미만인 대상자의 데이터는 신뢰롭지 않다고 판단하여, DB의 신뢰도, 민감도, 특이도 분석에서 제외하였다. 그리고 본 연구에서는 다수의 청자가 감정 음성을 평정하였기 때문에, 내적 일관성 신뢰도(internal consistency reliability)는 Cronbach's α 계수로 확인하였다.

2.6. 타당도 분석

본 연구에서 구축한 감정 음성 DB의 타당도를 분석을 위해 Belin et al.(2008)의 방법을 참고하여 민감도와 특이도를 계산하였다. 해당 개념을 감정 음성 DB의 타당도 검증에 적용할 경우, 민감도는 청자가 감정 음성에 표현된 감정을 정확하게 인식하는 지표로 사용된다. 즉, 청자가 들려주는 감정 음성을 듣고, 실제로 목표로 한 감정을 인식한 경우를 적중(hit)으로 간주하고, 적중률(hit rate)을 계산하여 민감도를 측정할 수 있다. 민감도가 높은 감정 음성은 화자가 표현하려고 한 감정이 잘 인식되는 음성으로 간주할 수 있다. 예를 들면, 청자에게 제시한 음성의 감정이 행복인데, 실제로 청자가 가장 높게 평정한 점수의 감정이 행복인 경우에 적중했다고 판단하며, 전체 청자 중에서 같은 음

성에 대해서 적중한 청자의 백분율(%)을 산출하여 민감도를 구하는 것이다. 감정별 민감도 산출을 위한 적중률 계산식은 아래와 같다.

$$hitrate(\%) = \frac{\text{the number of hit}}{\text{the number of listeners}} \times 100$$

특이도는 평정 척도와 관련된 것으로, 청자가 특정 감정에 대해서 가장 높은 점수를 준 음성이 실제로 의도된 감정과 일치하는 경우를 정기각(correct rejection)으로 간주하고, 정기각률(correct rejection rate)을 계산하여 특이도를 측정하였다. 예를 들면, 청자가 시각적 아날로그 평정 척도에서 행복을 가장 높게 평정한 음성이 실제로 배우가 행복을 표현한 음성인 경우를 정기각으로 판단하며, 전체 청자 중에서 같은 감정에 대해서 정기각한 청자의 백분율(%)을 산출하여 특이도를 구하는 것이다. 감정별 특이도 산출을 위한 정기각률 계산식은 아래와 같다.

$$correct\ rejection\ rate(\%) = \frac{\text{the number of correct rejection}}{\text{the number of listeners}} \times 100$$

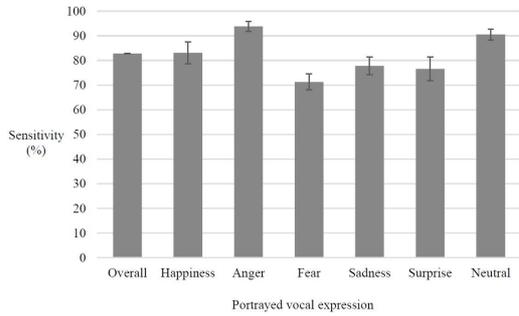
3. 연구 결과

3.1. 신뢰도

본 연구에서 음성에 대한 감정 평가 과제 참여의 신뢰도를 확인한 결과, 31명의 청자 중에서 감정 음성에 대한 정반응률이 80% 미만인 청자가 1명이었다. 청자 30명의 평정 결과로, 내적 일관성 신뢰도는 Cronbach α 가 .847로 높은 수준으로 나타났다.

3.2. 민감도 및 특이도

본 DB의 감정별 평정값과 민감도, 및 특이도는 표 4에 제시하였다. 각 음성 파일의 민감도와 특이도를 살펴보기 위해 각각 적중률과 정기각률을 산출하였다. 청자의 감정 인식에 대한 적중률로 계산된 전체 민감도는 82.8%였다. 감정 음성별로 살펴보면, 그림 3과 같이 행복 83.1%, 분노 93.8%, 공포 71.3%, 슬픔 77.8%, 놀람 76.6%, 중립 90.5%로 산출되었다. 감정 평정에서의 감정에 따른 정기각률로 계산된 전체 특이도는 83.8%였으며, 세부적으로는 그림 4와 같이 행복 88.7%, 분노 93.0%, 공포 80.2%, 슬픔 78.9%, 놀람 74.9%, 중립 86.2%로 산출되었다.



KAV DB, Korean Affective Voice Database.

그림 3. DB 전체 민감도 및 감정별 민감도
Figure 3. Overall and each sensitivity by emotion category in KAV DB (Korean affective voice database)



KAV DB, Korean Affective Voice Database.

그림 4. DB 전체 특이도 및 감정별 특이도
Figure 4. Overall and each specificity by emotion category in KAV DB (Korean affective voice database)

4. 논의 및 결론

본 연구는 사람들 간의 의사소통에서 운율정보를 통해 상대

방의 정서를 인식하는 능력이 중요하다는 것에 착안하여, 다양한 연령대의 일반인과 의사소통장애인을 대상으로 한 운율 기반의 정서 인식 능력 측정에 활용될 수 있는 국내 감정 음성 DB를 구축하고자 하였다. 이를 위하여, 남녀 배우 각 1인에게 의미적으로 중립적인 문장 10개를 행복, 분노, 공포, 슬픔, 놀람, 중립의 6개 정서로 표현해보도록 하여 음성 파일을 녹음하였고, 일반 성인 31명을 대상으로 감정 음성 DB의 신뢰도, 민감도, 특이도를 측정하여, DB에 대한 타당도를 점검하였다. 본 연구 결과에 따른 논의 및 결론은 아래와 같다.

본 연구에서 감정 음성 DB의 신뢰도를 확인한 결과, 내적 일관성 신뢰도 Cronbach's α 계수가 .847로, 신뢰도가 매우 높은 수준이었다. 이러한 결과는 시각 및 청각에 문제가 없는 20대 일반 성인이 '행복, 분노, 공포, 슬픔, 놀람, 중립'이라는 정서 어휘를 들었을 때 대표적으로 떠올리는 음성과 운율을 본 DB가 매우 잘 반영하고 있음을 의미한다. 이처럼 높은 수준의 신뢰도 결과는 감정 음성 DB 구축 과정이 매우 체계적이었기 때문에 가능했던 것으로 생각된다. 첫째, 발화 문장 제작 시 어절 및 음절 수와 난이도를 통제하여 대상자의 인지적 부담을 최소화하였다. 둘째, 언어병리학 석·박사과정생 6명과 언어병리학 교수 1인이 문장마다 의미적으로 중립적인지를 평가하고 선별함으로써, 대상자가 의미보다 운율에 더욱 집중하여 감정을 판단할 수 있도록 하였다. 셋째, 녹음에 참여한 배우는 뮤지컬 배우로 감정 음성 표현에 훈련된 화자였다. 또한, 녹음 과정에서 연구자가 실시간 모니터링을 하면서 배우가 감정을 제대로 표현할 때까지 2-3회 반복적으로 표현을 요청하였다. 넷째, 음성 파일 세트를 구성하기 전에, 모든 연구자가 각 음성에 어떤 감정이 표현되었는지를 평가하여 만장일치의 의견을 보인 음성 파일을 선별한 후에 세트를 구성하였다. 다섯째, 국외 감정 음성 DB와 달리, 많은 음성 파일을 듣고 평가하는 대상자의 과제 참여 신뢰성 확인을 위해 음성 파일 세트마다 평가 음성을 삽입하였다는 점 역시 높은 수준의 신뢰도 산출에 기여한 것으로 보인다.

표 4. 감정별 평정값, 민감도 및 특이도

Table 4. Ratings, sensitivity (hit rates), and specificity (correct rejection rates) per emotion

		Portrayed vocal expression						Correct rejection rate (%)
		Happiness	Anger	Fear	Sadness	Surprise	Neutral	
Intensity rating scale	Happiness	61.18* (17.82)	0.08 (0.42)	0.47 (1.68)	0.00 (0.01)	2.10 (3.15)	0.82 (1.30)	88.70 (3.71)
	Anger	0.85 (1.73)	78.68* (18.03)	3.05 (4.11)	5.18 (4.69)	6.64 (6.25)	1.76 (2.80)	93.00 (2.81)
	Fear	0.25 (0.73)	2.09 (3.48)	62.00* (20.26)	12.41 (9.03)	8.91 (8.10)	0.54 (1.26)	80.20 (3.19)
	Sadness	0.30 (0.72)	3.02 (6.31)	16.53 (11.28)	61.18* (18.85)	1.51 (3.99)	4.50 (5.67)	78.90 (3.48)
	Surprise	2.43 (2.83)	6.97 (9.48)	21.53 (15.16)	3.47 (4.96)	62.91* (19.05)	1.00 (1.60)	74.90 (5.14)
	Neutral	12.98 (12.00)	2.27 (3.86)	1.66 (4.29)	5.46 (5.64)	6.01 (6.02)	71.01* (20.98)	86.20 (3.08)
Hit rates (%)		83.10 (4.39)	93.80 (2.03)	71.30 (3.20)	77.80 (3.56)	76.70 (4.88)	90.50 (2.18)	

Values are presented as mean (SD). Cells indicate ratings (0-100) averaged across all actors and respondents for each emotion and rating scale.

*Asterisk indicates maximum average rating.

본 감정 음성 DB의 전체 민감도는 82.8%로 나타났다. 이러한 결과는 청자가 특정 감정이 표현된 음성을 듣고, 화자가 의도한 정서대로 인식할 확률이 82.8%임을 의미한다. 본 연구 결과의 민감도와 MAV(Belin et al., 2008) DB를 비교하여 살펴보면, 본 DB의 전체 민감도는 물론 행복, 분노, 공포 음성의 민감도가 국외 DB보다 높은 수치를 보인 반면, 슬픔과 놀람은 더 낮은 민감도를 보였다. 청자의 민감도 평정 결과를 세부적으로 살펴보면, 상반된 정서(valence)를 가진 정서에 대해서는 적게 인식하고 있었다. 예를 들면, 청자는 행복 음성에서는 공포 감정을, 슬픔 음성에서는 행복 감정을 가장 적게 인식하는 것으로 나타났다. 한국인의 감정별 음성 특징을 알아본 Kwon et al.(2012)의 연구 결과에 따르면, 상반된 정서를 지닌 감정을 표현할 때는 jitter, shimmer, 평균 기본주파수, LTAS(long term average spectral slope), H1-A1(첫 번째 포먼트 대역폭 B1의 크기), H1-A3(스펙트럼 기울기), 배음 대 소음 비율, SPI(soft phonation index) 변수에서 정반대의 경향을 보인다. 이러한 음성 감정의 특성으로 인하여, 본 연구에서는 상반된 감정에 대해서는 청자들이 가장 적게 인식한 것으로 생각된다.

본 감정 음성 DB의 전체 특이도는 83.8%였다. 이는 청자가 높게 평정한 감정이 실제로 배우가 표현한 감정일 확률이 83.8%임을 의미한다. 본 감정 음성 특이도를 MAV(Belin et al., 2008) DB와 비교하면, 본 DB의 전체 특이도는 물론 행복, 분노, 공포, 놀람이 더 높은 수치를 보였으나, 슬픔에서는 낮은 특이도를 보였다. 본 감정 음성의 DB는 국외 DB보다 전체 민감도와 전체 특이도가 높았는데, 이는 서양인(영국인, 네덜란드)보다 동양인(일본인)이 말의 내용보다 운율정보에 더 집중하고 민감하다는 경향(Kitayama & Ishii, 2002; Tanaka et al., 2010)이 반영된 것으로 보인다. 하지만 국외 DB에 비하여, 슬픔과 놀람에 대해서는 낮은 민감도와 특이도를 나타냈다. 이는 독일, 스위스, 영국, 네덜란드, 미국, 이탈리아, 프랑스, 스페인을 포함한 서양 국가와 동양국가 중 인도네시아를 비교한 Scherer et al.(2001)의 연구에서 인도네시아 집단이 슬픔에서 가장 낮은 감정 인식 정확도를 보였으며, 캐나다 집단보다 일본인 집단이 놀람에서 유의하게 높은 점수를 주었다는 Koeda et al.(2013)의 연구 결과를 지지한다고 볼 수 있다.

본 감정 음성 DB의 신뢰도, 민감도 및 특이도 결과를 통해, 각 감정 음성에는 연구자가 의도한 감정이 문화적으로 잘 반영되었다는 것을 확인하였다. 이는 본 DB가 운율에 기반하여 감정을 인식하는 능력을 측정하기 위한 도구의 자료로 사용되기에 적합함을 시사한다. 감정 인식 능력은 사회적 상호작용을 촉진한다는 점에서 매우 중요하지만, 일부 의사소통 장애를 경험하는 사람들이 유의하게 낮은 정서 인식 능력을 가지고 있다는 사실이 여러 선행연구(Choi & Lee, 2009; Filipe et al., 2014; Oerlemans et al., 2014; Peng et al., 2008; Sim, 2007)에서 확인되었다. 따라서 의사소통장애군의 운율 특성을 살펴보는 연구에 그치지 않고, 운율 인식 향상을 위한 중재 연구 및 임상 콘텐츠 제작이 필요하며, 이때 타당성이 갖춰진 본 DB가 적극적으로 활용될 수 있을 것이다.

본 연구에서 구축된 감정 음성 DB가 신뢰도와 타당도가 높은 것으로 확인되었음에도 제한점 역시 존재한다. 첫째, 인간이 인식하는 행복과 분노, 공포, 슬픔의 감정은 정서가, 각성(arousal)에 따라서 하부적인 감정으로 세분화될 수 있는데 (Willcox, 1982), 본 연구에서는 청자가 인식하는 감정 음성의 유형을 6개로 단순화하여 감정 음성 DB를 구축하고 타당도를 확인하였다. 둘째, 본 연구에서 구축한 음성 DB는 비언어적 단서 중 말속도, 음도, 강도와 같은 운율적 요소로만 감정이 표현되도록 하였다. 이는 청자가 의사소통 상황에서 화자의 감정을 이해할 수 있는 다양한 비언어적 단서(예: 음질, 음색, 표정, 제스처 등)에 대한 정보를 활용할 수 있다는 점을 반영하지 못하였다. 마지막으로, 본 DB에서는 비교적 명료하고 단순하게 특정 정서를 표현한 문장만을 포함하고 있으므로, 일상회화에 얽힌 보다 복잡하고 미묘한 여러 개의 정서를 잘 인식하는 능력은 본 DB를 통해 측정되지 않을 수 있다는 것이다.

후속연구를 위한 제언점은 다음과 같다. 첫째, 사투리 사용자의 감정 음성도 DB에 포함될 필요가 있다. 본 연구에 참여한 청자와 화자는 모두 서울말 구사자였기 때문에, 각 정서 및 운율에 대해 모종의 공통적이고 제한적인 이해를 가지고 있었다. 하지만 억양, 음조, 강세, 템포, 리듬 등 운율을 조성하는 가장 중요한 기술적인 요소들이 사투리에서 두드러진다는 점을 감안하였을 때(Ma, 2012), 청자 혹은 화자에 사투리 발화자를 포함시킨다면 감정 음성 DB의 타당도 결과는 달라질 수도 있다. 따라서 후속연구에서는 다양한 방언으로 정서 발화를 수집하여 더욱 다양하고 광범위한 데이터를 구축할 필요가 있다. 둘째, 정서 처리는 문화적 범위뿐만 아니라 개인적 범위에서도 차이가 드러난다. 예를 들어, 놀람의 정서를 떠올렸을 때 갑작 선물을 받고 기뻐하는 긍정적인 놀람을 느끼는 사람이 있는 반면, 어떤 무서운 것이 갑작스럽게 나타나 느끼게 되는 부정적인 놀람을 느끼는 사람 역시 존재한다. 따라서 더 많은 청자와 화자를 대상으로 타당도 검증을 진행한다면 개인적 정서 처리의 차이를 극복할 수 있는 범용성이 큰 데이터로서의 가치를 입증할 수 있을 것이다.

References

- Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614-636.
- Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40*(2), 531-539.
- Carstensen, L. L., Isaacowitz, D. M., & Charles, S. T. (1999). Taking time seriously: A theory of socioemotional selectivity. *American Psychologist, 54*(3), 165-181.
- Choi, Y. G., & Lee, J. Y. (2009). A study of the emotional recognition of children with cochlear implants. *Journal of Speech-Language*

- and *Hearing Disorders*, 18(3), 137-150.
- Coutinho, E., & Dikken, N. (2013). Psychoacoustic cues to emotion in speech prosody and music. *Cognition and Emotion*, 27(4), 658-684.
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Palo Alto, CA: Consulting Psychologists Press.
- Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., Krause, R., ... Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*, 53(4), 712-717.
- Elfenbein, H. A., & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, 128(2), 203-235.
- Filipe, M. G., Frota, S., Castro, S. L., & Vicente, S. G. (2014). Atypical prosody in Asperger syndrome: Perceptual and acoustic measurements. *Journal of Autism and Developmental Disorders*, 44(8), 1972-1981.
- Grandjean, D., Sander, D., Pourtois, G., Schwartz, S., Seghier, M. L., Scherer, K. R., & Vuilleumier, P. (2005). The voices of wrath: Brain responses to angry prosody in meaningless speech. *Nature Neuroscience*, 8(2), 145-146.
- Imaizumi, S., Mori, K., Kiritani, S., Kawashima, R., Sugiura, M., Fukuda, H., Itoh, K., ... Nakamura, K. (1997). Vocal identification of speaker and emotion activates different brain regions. *NeuroReport*, 8(12), 2809-2812.
- Kang, E. J., Hwang, M., & Jeong, M. (2014). Emotional recognizing ability from the prosodies of children with high-functioning autism. *Journal of Emotional and Behavioral Disorders*, 30(3), 79-94.
- Kim, C. H., Kim, Y. T., & Lee, S. J. (2013). Effect of context and affective prosody on emotional perception in children with high-functioning autism. *Communication Sciences and Disorders*, 18(1), 24-34.
- Kitayama, S., & Ishii, K. (2002). Word and voice: Spontaneous attention to emotional utterances in two languages. *Cognition and Emotion*, 16(1), 29-59.
- Koeda, M., Belin, P., Hama, T., Masuda, T., Matsuura, M., & Okubo, Y. (2013). Cross-cultural differences in the processing of non-verbal affective vocalizations by Japanese and Canadian listeners. *Frontiers in Psychology*, 4, 105.
- Kotz, S. A., Meyer, M., Alter, K., Besson, M., von Cramon, D. Y., & Friederici, A. D. (2003). On the lateralization of emotional prosody: An event-related functional MR investigation. *Brain and Language*, 86(3), 366-376.
- Kwon, C., Song, S., Kim, J., Kim, K., & Jang, J. (2012). Extraction of speech features for emotion recognition. *Phonetics and Speech Sciences*, 4(2), 73-78.
- Laukka, P. (2005). Categorical perception of vocal emotion expressions. *Emotion*, 5(3), 277-295.
- Laukka, P., & Elfenbein, H. A. (2021). Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, 13(1), 3-11.
- Laukka, P., Elfenbein, H. A., Thingujam, N. S., Rockstuhl, T., Iraki, F. K., Chui, W., & Althoff, J. (2016). The expression and recognition of emotions in the voice across five nations: A lens model analysis based on acoustic features. *Journal of Personality and Social Psychology*, 111(5), 686-705.
- Lee, S. J., & Kim, Y. T. (2013). Review of prosodic ability in children and adolescents with autism spectrum disorder. *Journal of the Korean Association for Persons with Autism*, 13(1), 47-71.
- Ma, X. (2012). *A study on the effective acquisition method of rhythm for the dialect performance: Based on the cases of the command guidance of hamgyeong-do dialect for the characters* (Master's thesis). Korea National University of Arts, Seoul, Korea.
- Oerlemans, A. M., van der Meer, J. M. J., van Steijn, D. J., de Ruiter, S. W., de Bruijn, Y. G. E., de Sonnevile, L. M. J., Buitelaar, J. K., ... Rommelse, N. N. J. (2014). Recognition of facial emotion and affective prosody in children with ASD (+ADHD) and their unaffected siblings. *European Child and Adolescent Psychiatry*, 23(5), 257-271.
- Park, C. O. (2010). *The development of emotion reading ability in prosody of language* (Master's thesis). Chungbuk National University, Cheongju, Korea.
- Park, S. H., Kim, J. Y., & Park, S. Y. (2016). A literature review of prosody in people with communication disorders. *Journal of Speech-Language and Hearing Disorders*, 25(3), 155-171.
- Peng, S. C., Tomblin, J. B., & Turner, C. W. (2008). Production and perception of speech intonation in pediatric cochlear implant recipients and individuals with normal hearing. *Ear and Hearing*, 29(3), 336-351.
- Planalp, S. (1996). Varieties of cues to emotion in naturally occurring situations. *Cognition and Emotion*, 10(2), 137-154.
- Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76-92.
- Seo, S. (2021). *2021 Version of basic research for lexical grading and selection of basic vocabulary in Korean*. Seoul, Korea: National Institute of Korean Language.
- Shin, H. B., Choi, J., Kim, K., & Lee, Y. (2015). Expressive prosody in autism spectrum disorders: Meta-analysis. *Communication Sciences and Disorders*, 20(3), 424-434.
- Sim, H. S. (2007). Physiological processing of prosody in language disordered people: A review of the literature. *Special Education Research*, 6(2), 129-144.
- Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & de

Gelder, B. (2010). I feel your voice: Cultural differences in the multisensory perception of emotion. *Psychological Science*, 21(9), 1259-1262.

Trainor, L. J., Austin, C. M., & Desjardins, R. N. (2000). Is infant-directed speech prosody a result of the vocal expression of emotion? *Psychological Science*, 11(3), 188-195.

Willcox, G. (1982). The feeling wheel: A tool for expanding awareness of emotions and increasing spontaneity and intimacy. *Transactional Analysis Journal*, 12(4), 274-276.

• **김예지 (Yeji Kim)**

이화여자대학교 언어병리학과 석사과정

서울시 서대문구 이화여대길 52

Tel: 02-3277-2120

Email: edge_kim@ewhain.net

관심분야: 아동언어발달, 아동언어장애, 이중언어

• **송혜선 (Hyesun Song)**

이화여자대학교 언어병리학과 석사과정

서울시 서대문구 이화여대길 52

Tel: 02-3277-2120

Email: bl435@naver.com

관심분야: 아동언어발달, 책읽기중재

• **전예솔 (Yesol Jeon)**

이화여자대학교 언어병리학과 석사과정

서울시 서대문구 이화여대길 52

Tel: 02-3277-2120

Email: 212shg15@ewhain.net

관심분야: 청각장애, 말장애, 말명료도, 음성학

• **오유림 (Yoorim Oh)**

이화여자대학교 언어병리학과 석사과정

서울시 서대문구 이화여대길 52

Tel: 02-3277-2120

Email: ddang330@naver.com

관심분야: 아동언어발달, 이중언어

• **이영미 (Youngmee Lee)** 교신저자

이화여자대학교 언어병리학과 교수

서울시 서대문구 이화여대길 52

Tel: 02-3277-4603

Email: youngmee@ewha.ac.kr

관심분야: 난청, 말장애, 음성장애

부록 1. 신뢰도 평가를 위한 감정별 음성 발화 문장 목록

행복	와! 드디어 내일이면 방학이다!
	오예! 오늘은 놀이공원에 간다!
분노	아저씨, 새치기를 하시면 어떡해요!
	너 때문에 나까지 넘어졌잖아!
공포	귀신의 집은 안 들어갈래. 무서워.
	커다란 맹수가 들어오면 어떡하지?
슬픔	할머니가 간밤에 돌아가셨어.
	어떡해, 고양이가 많이 아픈가봐.
놀람	헤엑, 그렇게나 많이 먹을 수 있어?
	너 언제 이렇게 키가 컸어?
중립	인터넷에는 다양한 정보가 있어.
	좌회전하면 바로 빵집이 나와.

한국형 감정 음성 데이터베이스 구축을 위한 타당도 연구

김 예 지 · 송 혜 선 · 전 예 술 · 오 유 림 · 이 영 미

이화여자대학교 언어병리학과

국문초록

본 연구는 운율을 기반으로 감정을 인식하는 능력을 측정할 때 이용할 수 있는 한국형 감정 음성 데이터베이스 (Korean Affective Voice Database, 이하 KAV DB)를 개발하고, 해당 DB가 의사소통장애 관련 임상과 연구에서 활용될 수 있는지를 점검하기 위하여 신뢰도, 민감도, 특이도를 산출하여 그 타당성을 살펴보았다. 본 연구에서는 배우 2명(남 1명, 여 1명)이 의미적으로 중립적인 문장 10개를 행복, 분노, 공포, 슬픔, 놀람, 중립의 6개 정서로 표현하도록 하여 음성을 녹음하였다. 녹음된 음성 중에서 목표 정서가 잘 표현된 문장을 선별하여 타당도 점검을 위한 음성 파일 세트를 구성하였으며, 청자 31명(남 14명, 여 17명)이 시각적 아날로그 평정법을 이용하여 각 음성에서 6개의 정서가 얼마나 반영되었는지 평정하도록 하였다. 그 결과, KAV DB에 대한 청자의 내적 일관성 신뢰도는 .872, 전체 민감도 82.8%, 전체 특이도 83.8%였다. 이를 통하여, 타당도가 확보된 KAV DB는 감정 음성 인식 및 산출과 관련된 연구와 임상 콘텐츠 제작 등에 활용될 수 있을 것으로 기대된다.

핵심어: 운율, 감정 음성, 감정 인식

참고문헌

- 강은주, 황민아, 정미란(2014). 고기능 자폐아동의 운율정보를 통한 정서 인식 능력. *정서 행동장애연구*, 30(3), 79-94.
- 권철홍, 송승규, 김종열, 김근호, 장준수(2012). 감정 인식을 위한 음성 특징 도출. *말소리와 음성과학*, 4(2), 73-78.
- 김찬희, 김영태, 이수정(2013). 맥락 및 운율에 따른 고기능 자폐아동의 정서 이해 특성. *언어청각장애연구*, 18(1), 24-34.
- 마학봉(2012). 사투리 연기를 위한 효과적인 운율 습득방법 연구. *함경도 사투리를 구사하는 인물사투리 지도 사례를 바탕으로 한국예술종합학교 석사학위논문.*
- 박소현, 김지연, 박선영(2016). 의사소통장애인의 운율에 관한 연구 동향. *언어치료연구*, 25(3), 155-171.
- 박찬욱(2010). *말의 운율에서 정서를 읽는 능력의 발달*. 충북대학교 석사학위논문.
- 서상규(2021). *2021 년 국어 기초 어휘 선정 및 어휘 등급화 연구*. 서울: 국립국어원.
- 신희백, 최지은, 김기숙, 이윤경(2015). 자폐스펙트럼장애의 운율 표현에 대한 메타분석. *언어청각장애연구*, 20(3), 424-434.
- 심현섭(2007). 언어장애인 운율손상의 생리학적 기전 연구동향. *특수교육*, 6(2), 129-144.
- 이수정, 김영태(2013). 자폐범주성장애 아동과 청소년의 운율 특성에 관한 연구문헌 분석. *자폐성장애연구*, 13(1), 47-71.
- 최양규, 이지윤(2009). 음성 청취를 통한 인공와우이식 아동의 정서 인식에 관한 연구. *언어치료연구*, 18(3), 137-150.