Linguistic Research 41(3): 611-633
DOI: 10.17250/khisli.41.3.202412.012

Acoustic cues for emotions in Korean vocal expressions[*]

Eunhae Oh
(Konkuk University)

Oh, Eunhae. 2024. Acoustic cues for emotions in Korean vocal expressions. *Linguistic Research* 41(3): 611-633. This study examined the acoustic expression of emotions—specifically happiness, sadness, anxiety, and anger—in speech, with a particular focus on gender-based differences in vocal expressions. Ten phonetic markers, including pitch, intensity, and voice quality, were measured and analyzed through various acoustic parameters. Findings revealed that while both male and female speakers modified acoustic features to convey emotions, distinct gender patterns emerged. Female speakers exhibited stable pitch in sadness and heightened intensity variability when expressing anger, while male speakers showed a tendency to increase pitch variability in sadness and irregularities in anger-related vocalizations. These patterns suggested that social factors may influence how emotions are expressed, with Korean societal norms possibly affecting the vocal differences between male and female speakers. The results are interpreted in light of how gender and societal norms may shape vocal expressions of emotion. **(Konkuk University)**

Keywords vocal emotions, Korean speech production, emotional prosody, acoustic cues, gender differences

1. Introduction

The human voice is a rich source of information and it conveys not only linguistic content but a wide range of emotions. Studies have demonstrated that the voice alone provides reliable cues for accurately discerning a person's emotional state (Scherer 2003; Thompson and Balkwill 2006, 2009; Sauter et al. 2010; Jürgens et al. 2018; Lawson and Mayer 2021). In particular, acoustic features such as pitch, loudness, speech rate and voice quality were reported to be strongly associated with emotions expressed

[*] This work was supported by Konkuk University in 2024.

in specific affective prosody (Scherer 1989; Bryant and Barrett 2008; Pell and Skorup 2008; Sauter et al. 2010; Belyk and Brown 2014). For instance, happiness is often characterized by a fast speech rate, high intensity, and variability in pitch, while sadness is typically associated with a slower speech rate, lower intensity, and less pitch variability (Murray and Arnott 1993; Schirmer and Kotz 2006).

Researchers have identified patterns in these features that are universally recognized across languages. Sauter et al. (2010) compared emotional vocalizations between European native English speakers and the Himba, a semi-nomadic group living in northern Namibia. The study aimed to investigate cross-cultural recognition of nonverbal vocalizations by testing English and Himba listeners' ability to match vocal signals from each other's cultural groups to emotional stories. Participants were presented with nine emotional stories and asked to identify the corresponding emotion from two vocalization sounds. The results showed that the English listeners successfully matched Himba sounds to emotions, and the Himba listeners accurately recognized English sounds. The study demonstrated that nonverbal acoustic cues convey emotions across cultural boundaries and are reliably recognized within cultural groups. Zhang and Pell (2022) also investigated the impact of cultural background on how emotions are expressed and understood in spoken language. The study involved Canadian and Chinese participants, who were tasked with identifying four different emotions (anger, fear, happiness, sadness) expressed in English, Mandarin, and Hindi. Results indicated that although there was an advantage in recognizing ingroup emotions, the Canadian and Chinese participants similarly interpreted emotions conveyed through vocal cues. Furthermore, both groups perceived the speaker's emotions as equal to or more intense than what was actually displayed, particularly for negative emotions. These findings imply a cross-cultural understanding of non-verbal emotional cues, which could allow listeners to swiftly discern emotional states solely through voice, even in languages they do not understand. Past studies have established that while certain acoustic cues of emotion, such as pitch and intensity, are universally recognized, their expression and interpretation can be influenced by cultural and gender-specific factors. For example, Sauter et al. (2010) demonstrated the cross-cultural reliability of nonverbal emotional vocalizations, while Zhang and Pell (2022) found that cultural background modulates emotional recognition and expression. By focusing on the Korean context, this study contributes to a deeper understanding of how gender and culture intersect to shape vocal emotion expression in Korean.

2. Acoustic features in speech emotion

One of the most commonly studied correlates of speech emotion is the fundamental frequency (F0) (for further reviews see Johnstone and Scherer 2000; Juslin and Laukka 2003; Scherer 2003). A frequent discovery within these studies is that emotions impact overarching features of F0. For example, Banse and Scherer (1996) analyzed portrayals of 14 different emotions produced by professional actors, focusing on the intensity and positive or negative valence of the speech emotion. The study demonstrated that listeners can accurately distinguish emotions from vocal cues with significant accuracy and that the changes in F0 greatly contributed in distinguishing different emotional valences. Bänziger and Scherer (2005) investigated the role of intonation in conveying eight different emotional expressions. They conducted a quantitative analysis on a comprehensive corpus of vocal portrayals of emotions and found that basic descriptions of mean F0 or F0 range sufficiently explained the primary differences observed among emotion categories. They further identified distinct differences in F0 contours corresponding to specific emotions, noting that expressions of anger and elation exhibited greater F0 excursions compared to emotions such as sadness or happiness.As reviewed by Scherer (2003), a number of studies on speech emotion have employed a technique to systematically alter acoustic features in real speech utterances (Ladd et al. 1985; Tolkmitt et al. 1988). Among all the variables examined, the F0 range was shown to have the most significant impact on listeners' judgments of speaker attitude and emotional state. For instance, a narrow F0 range was associated with conveying sadness, while a wide F0 range was interpreted as indicating negative emotions such as annoyance or anger.

Studies have also reported meaningful correlations between speech intensity and emotional states. Here, intensity refers to the vocal amplitude or loudness of the acoustic signal, and emotional arousal can lead to changes in intensity, with higher intensity typically associated with negative emotions or aggression. Scherer (2003) showed that intensity, coupled with pitch and voice quality, are crucial in distinguishing speech emotion. Both happiness and anger were linked to higher levels of vocal intensity, while sadness was associated with lower intensity, reflecting more subdued emotional expressions. A study by Patel and Scherer (2011) also examined the relationship between intensity and emotional valence and proposed that not only do strong emotions evoke higher intensity in speech, but these elevated cues also serve as reliable indicators of high-arousal emotional states across different languages and cultural contexts. Furthermore, in the realm of computational linguistics, algorithms have been developed to detect emotional cues in speech based on intensity and other acoustic features. Research by Lee and Narayanan (2005) demonstrated the efficacy of these algorithms in identifying emotional states with considerable accuracy, and further highlighted the critical role of intensity in the acoustic representation of emotions. The prior findings offer an in-depth understanding of how the combination of intensity and F0 contributes to the conveyance of specific emotional expressions.

Other features that have received comparatively less attention are speech rate, voice quality, articulation, and spectral features. Fast speech rates are often associated with high-arousal emotions such as excitement or anger, whereas slower rates tend to accompany low-arousal states like sadness or calmness (Banse and Scherer 1996). Voice quality, including breathiness, roughness, and tension, can also influence the emotional tone of speech. Johnstone and Scherer (2000) discussed how voice quality serves as a subtle yet powerful cue in emotional expression, with breathy voice often signaling sadness or tenderness, and harsh voice indicating anger or fear. Gobl and Chasaide (2003) also showed that changes in voice quality alone can evoke different emotional perceptions in speakers. They argued that voice quality serves a distinct role from other acoustic cues by conveying the valence (positive or negative) of an emotion, whereas pitch, intensity, and duration indicate activation levels. For instance, tense or harsh voices were associated with high activation emotions such as anger, stress, confidence, and interest. In contrast, breathy, whispery, or creaky voices were linked to low activation emotions like relaxation, intimacy, sadness, and boredom, both with a mix of positive as well as negative valences. Additionally, El Ayadi, Kamel, and Karray (2011) showed that combining prosodic features (e.g., pitch and energy) with voice quality measures such as jitter and shimmer enhanced the robustness of emotion recognition systems. Together, these parameters can ensure the robustness and depth of the analysis in examining gender-specific differences in vocal expressions.

Investigating the different acoustic cues used for various emotions is crucial because each emotion uniquely modulates speech characteristics, influencing how it is perceived and recognized. Different emotions emphasize distinct acoustic parameters. In a study on emotional speech in Korean, for example, Kim et al. (2004) analyzed pitch, amplitude (i.e., intensity), and duration at the beginning, middle and end of the short sentences produced in sad, angry and joyful emotions by 4 male Korean speakers. Each feature was controlled individually and then modified simultaneously to examine the contribution of each acoustic parameter on the perception of different emotions in speech. They reported that short duration was the primary feature for conveying anger. Nine out of 10 participants correctly identified the intended angry emotion when pitch and amplitude were controlled. Amplitude was shown to be the key feature for conveying sadness, whereas joyful emotion did not show any dominant features. They concluded that emotional speech generally showed lower pitch and higher amplitude compared to neutral speech but there was no consistent trend in feature variations across different emotions. The study only examined five short sentences repeated five times by four male speakers, further emphasizing the need for larger datasets and inclusion of various acoustic properties to gain a more comprehensive understanding of emotional speech.

While research has identified connections between acoustic cues and speech emotion, it's also important to recognize that the connections must be viewed in the context of social variables, including the gender of speakers and cultural perceptions of emotions. Differences among speakers can influence how emotional speech is perceived, and they add complexity to the decoding of vocal emotion cues. Listeners bring their own biases and experiences to the process of interpreting emotions from speech, further complicating the variability in emotional vocal cues. Studies have shown that female listeners tend to be more accurate than male listeners in recognizing emotions from faces, gestures, and voices (Hall 1978; Schirmer et al. 2005). Lausen and Schacht (2018) examined the impact of gender on recognizing emotions through vocal expressions with 290 participants who were presented with a variety of vocal stimuli. Results showed that performance accuracy varied based on listener and speaker gender. Female speakers were consistently judged more accurately for specific emotions. This pattern was interpreted to indicate that societal roles and biological factors may contribute to females' higher sensitivity to subtle emotional cues, potentially due to their nurturing and affiliative roles. Additionally, studies have demonstrated how females and males tend to use different acoustic cues to express the same emotions. In Schirmer and Kotz (2006a), gender differences were observed in how emotional vocalizations were processed as well. While both female and male participants exhibited similar patterns of activation in many brain regions, differences were noted in the extent of activation and neural connectivity. Specifically, the female participants showed greater activation in brain regions associated with emotional processing, suggesting that they may be more sensitive to emotional cues conveyed through vocalizations. Additionally, the study found that females displayed greater variations in vocal intensity compared to males when expressing emotions. The result implies that female speakers are more likely to utilize variations in intensity to communicate emotional states compared to male speakers. This interplay of biological, psychological, and cultural factors emphasizes the significance of exploring gender-specific vocal strategies within the Korean cultural context.

As for the influence of cross-cultural contexts, Laukka et al. (2014) investigated cultural differences in vocal expressions of emotion by analyzing emotionally-inflected speech segments from hundred professional actors across five English-speaking cultures. Using machine learning, researchers classified expressions based on acoustic features, testing both within and across cultural boundaries. Results showed above-chance accuracy in cross-cultural classification, suggesting shared characteristics in vocal expressions across cultures. However, accuracy was higher within cultures, indicating an in-group advantage. Zhang and Pell (2022) found that Chinese participants with some exposure to English culture exhibited higher accuracy in recognizing emotions in English compared to Hindi. This suggests that their proficiency in English as a second language may have contributed to improved recognition performance. These findings support the idea of cultural differences in vocal expression and align with the dialect theory of emotions, which suggests that familiarity with culturally specific expressive styles leads to better recognition of in-group expressions. Similarly, Liu et al. (2016) demonstrated that Chinese immigrants in Canada have exhibited comparable behavioral reactions to emotion processing as North American participants, suggesting that L2 learners can acquire the perceptual sensitivity to adapt to the speech emotion in the L2 with greater experience to the culture.

Building on previous research, this study examined the acoustic properties of emotions as expressed in speech, with a particular focus on identifying potential gender-based differences in the use of these properties among male and female Korean speakers. Specifically, the study investigated which acoustic properties are characteristic of emotions such as happiness, sadness, anxiety, and anger in speech and whether male and female speakers differ in their use of these properties when expressing the same emotions. To address these questions, the study analyzed a collection of voice samples from professional broadcasters, each expressing multiple Korean sentences in four targeted emotions: happiness, sadness, anxiety, and anger. By analyzing trained broadcasters, who are skilled at modulating their voices for specific emotions, this study aimed to establish a baseline for understanding how acoustic cues differ across emotional states. These results can provide salient acoustic cues for future studies involving more naturalistic samples.

3. Methods

This study used datasets from 'The Open AI Dataset Project (AI-Hub, S. Korea)'. All data information can be accessed through 'AI-Hub (www.aihub.or.kr). The dataset comprised 31,730 observations from 16 speakers (8 females, 8 males), with 4 speakers assigned to each of the 4 different emotions. The four speakers for each emotion delivered the same sentences but the scripts varied across different emotions. There was a total of 14 missing or inaudible files due to poor recording quality. The speakers, all of whom were voice actors, were Seoul dialect users and produced all sentences in a conversational style. Each emotion involved production of different sentences, with accentual phrases (AP) ranging from a minimum of one to a maximum of five. However, no significant differences in the mean length of utterance (MLU), or the average number of words, were observed across the different emotions.

For data analysis, Praat software version 6.0.35 (Boersma and Weenik 2017) was employed to measure 10 phonetic characteristics. These included the median and range of fundamental frequency (Hz) and intensity (dB), as well as the standard deviation (SD) of both F0 and intensity. Additional parameters such as log-transformed speech duration, local shimmer (amplitude variability in dB), local jitter (frequency variability in %), and the Harmonic-to-Noise ratio (HNR, indicative of voice signal noise, in dB), were also quantified. The selection of ten acoustic cues, including F0 median, F0 range, intensity median, intensity variability, jitter, shimmer, and Harmonics-to-Noise ratio, was guided by their established relevance in previous research on emotional speech (Banse and Scherer 1996; Johnstone and Scherer 2000; Gobl and Chasaide 2003; El Ayadi, Kamel, and Karray 2011). In this study, speech duration, defined as the total time taken to articulate a given sentence, was analyzed to compare how different emotions influence the timing of vocal delivery in controlled utterances, rather than the pace of articulation (i.e., speech rate). By incorporating these features, the current study aims to provide a comprehensive framework for examining how gender influences the acoustic manifestation of emotions in Korean.

For statistical analysis, Linear Mixed-Effects Models were utilized, employing packages lme4 version 1.1.1.3 (Bates et al. 2014) and afex 0.16.1 (Singmann et al. 2015). The model incorporated four emotions (Happy, Sad, Anxious, Angry) and Gender (female vs male) as fixed effect predictors, including their interaction effects. Random effects consisted of varying intercepts by subject and scenario, and by subject random slopes for emotion, structured as: Acoustic features ~ Emotion * Gender + (1 + Emotion|Subject) + (1|Scenario). In our acoustic analysis of emotional speech, we employed R Treatment Contrast with "Happy" as the reference level for comparing other emotions. This analysis was chosen because “Happy” is the only emotion with a positive valence among the four emotional speech categories. Additionally, studies have shown that acoustic features extracted from speech signals can effectively discriminate happy emotions from other emotional states (Scherer and Oshinsky 1977; El Ayadi, Kamel, and Karray 2011). By anchoring our comparisons to "Happy", the aim was to evaluate how other emotions differ acoustically from this established base. This method allowed for a systematic assessment of how the acoustic attributes of various emotions diverge from the typical acoustic profile of "Happy."

4. Results

4.1 Speech duration

The variation in speech duration reflected the distinct acoustic characteristics associated with these emotions. The analysis revealed that emotion showed a significant main effect on speech duration (χ^2 = 20.37, p < .001). Compared to the baseline emotion, Happy, estimated coefficients for Anxious (Estimate = -369.070, SE = 107.664, t = -3.428, p = 0.01359) and Sad (Estimate = 379.780, SE = 87.728, t = 4.329, p = 0.00198) emotions indicated notable decreases in duration for Anxious and increases for Sad emotions. Speakers increased their speech duration when expressing anxiety and slowed down when expressing sadness. Gender also showed a significant main effect (χ^2 = 4.99, p = .025), with the coefficient for male speakers (Estimate = -264.068, SE = 108.968, t = -2.423, p = 0.03271), suggesting significantly shorter duration compared to female speakers. As shown in Figure 1, the difference in duration for Sad was more pronounced for male speakers. However, the absence of a statistically significant interaction between emotion and gender indicated that the effect of emotion on speech duration was overall consistent for both female and male speakers.

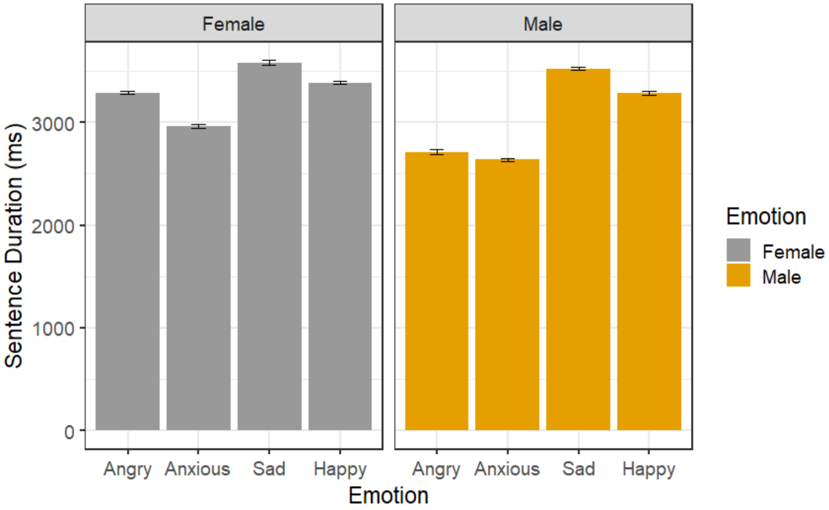

Figure 1. Speech duration (in milliseconds) for four emotions by gender

4.2 F0 median and F0 range

Emotion returned a significant main effect on the F0 median (χ^2 = 15.79, p = .001). Among the emotions with negative valence, only sadness exhibited a lower pitch compared to happiness. (Estimate = -9.515, SE = 2.693, t = -3.533, p = 0.005) Figure 2 shows a significant decrease in overall pitch for Sad across genders. Angry and Anxious showed a decrease in F0 median compared to Happy, but not in a statistically significant manner. The interaction between emotion and gender was not significant.

There were significant effects of emotion on the F0 range (χ^2 = 11.03, p = .012), suggesting that different emotions lead to noticeable changes in pitch variation within an utterance. Although there was a significant main effect of emotion on the F0 range, this effect was moderated by gender, diminishing the apparent main effect of emotion. Neither Anger, Anxious nor Sad emotions showed significant main effects compared to Happy. However, Gender had a significant effect ($\chi^2 = 15.72$, $p < .001$), with male speakers exhibiting a consistently smaller F0 range across all emotions compared to female speakers (Estimate = -91.7953, SE = 14.354, t = -6.395, $p < 0.001$). The interaction between Emotion and Gender was also significant, suggesting gender-specific patterns in how emotions were expressed. Specifically, Anxious and Sad significantly interacted with gender (p = 0.0123 and p = 0.0237, respectively). As illustrated in Figure 2 (right), Anxious emotion was associated with a smaller F0 range in male speakers, while Sad showed the smallest range in female speakers. Taken together with the mean F0, Sad was expressed with generally lower pitch in both genders. Notably, however, sadness among female speakers was reflected in a less varied pitch and thus diminished emotional expressiveness, whereas sadness was expressed with a comparatively broader F0 range.

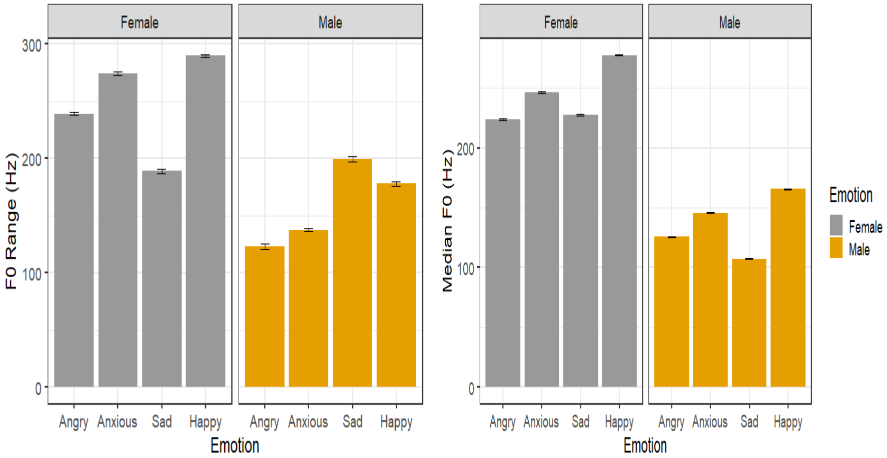

Figure 2. F0 range (left) and F0 median (right, in Hz) for four emotions by gender

4.3 F0 variability

Emotion significantly influenced F0 variability ($\chi^2 = 13.48$, p = .004). Compared to the Happy emotion, Sad significantly reduced F0 variability (Estimate = -9.515, SE = 2.693, t = -3.533, p = 0.004648). In contrast, the effects of Angry and Anxious were not statistically significant. gender also played an important role, showing a significant main effect (χ^2 = 18.61, p < .001). The main effect of gender indicated an overall decrease in F0 variability for male speakers (Estimate = -23.388, SE = 3.989, t = -5.864, p = 0.000113), suggesting distinct gender-based differences in speech variability patterns across emotional states. The interaction between emotion and gender was marginally significant (χ^2 = 7.27, p = .064). Although the influence of emotion on F0 variability was relatively consistent across gender, the effect of Sad differed significantly across gender (Estimate = 19.979, SE = 5.377, t = 3.716, p = 0.003411). As illustrated in Figure 3, female speakers exhibited significantly reduced F0 variability for Sad compared to other emotions, whereas male speakers did not show this distinction as prominently.

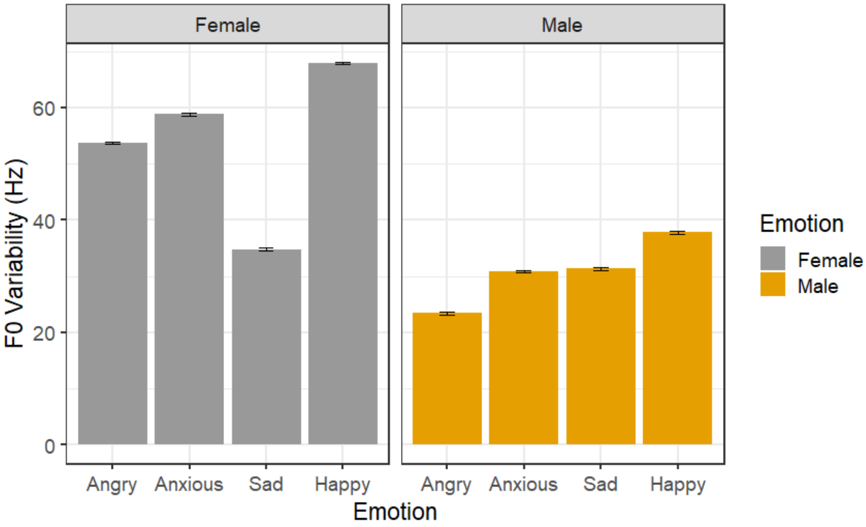

Figure 3. F0 variability (standard deviation in Hz) for four emotions by gender

4.4 Median intensity and intensity range

The analysis showed that emotion had a significant effect on the median intensity (χ^2 = 8.56, p = .036). There was a significant increase in the intensity for Anxious (Estimate = 1.0685, SE = 0.3941, t = 2.711, p = 0.02044), indicating a stronger vocal expression compared to Happy. On the other hand, Sad was associated with a significant decrease in the intensity (Estimate = -1.4837, SE = 0.4059, t = -3.655, p = 0.00419), indicating a softer and a subdued vocal expression. The Gender effect did not reach statistical significance. The effect of sadness on intensity was not uniform across gender (Estimate = -2.9124, SE = 0.8089, t = -3.601, p = 0.00470). As illustrated in Figure 4 (left), Sad emotion in male speech was expressed with notably lower intensity than female speech. The difference may be interpreted as male speakers using different strategies for regulating their emotions compared to females, who showed greater control over F0 features.

The analysis of intensity range with respect to emotion and gender did not reveal significant main effects or interaction effects. Although emotion or gender did not influence intensity range, Sad emotion showed a statistically meaningful decrease in intensity range. As shown in Figure 4 (right), the intensity range for Sad was distinctively smaller especially for the female speakers (Estimate = -3.9925, SE = 1.3721, t = -2.910, p = 0.0175).

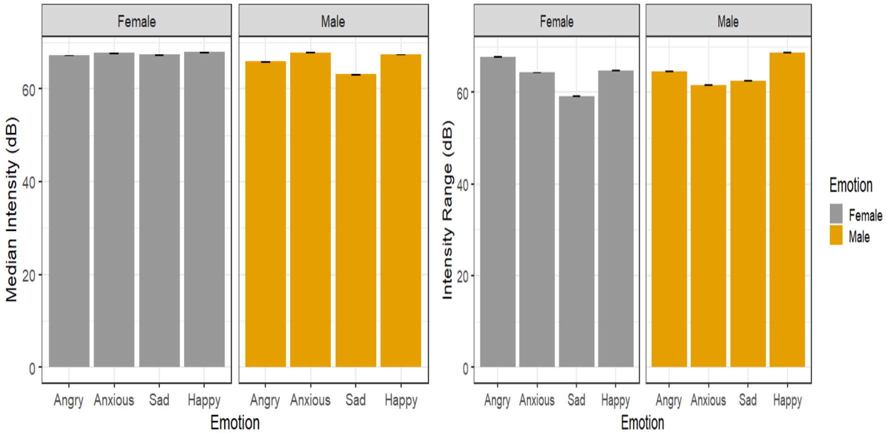

Figure 4. Median intensity (right, in dB) and intensity range (left) for four emotions by gender

4.5 Intensity variability

A marginal main effect of emotion on intensity variability (s.d.) was observed (χ^2 = 6.80, p = .079) and the interaction between emotion and gender was also significant (χ^2 = 10.32, p = 0.0272), indicating that the relationship between emotion and intensity variability was different for the male and female speakers. As shown in Figure 5, the difference was evident for the Angry emotion. The female speakers exhibited a greater increase in intensity variability for Angry compared to Happy, significantly more so than male speakers (Estimate = -2.99127, SE = 0.97103, t = -3.081, p = 0.0276). Sad emotion in female speech, but not male speech, showed a significant decrease in intensity variability compared to Happy emotion (Estimate = 1.97792, SE = 0.79394, t = 2.491, p = 0.0322). That is, when female speakers expressed sadness, their speech intensity was more stable and showed smaller fluctuations than when they expressed happiness. This effect was not observed in male speech.

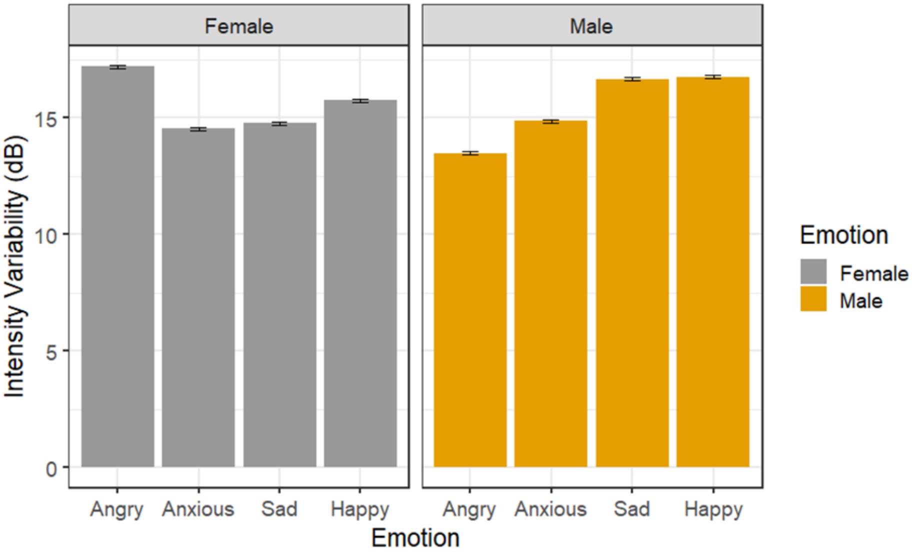

Figure 5. Intensity variability (standard deviation in Hz) for four emotions by gender

4.6 Jitter

Emotion had a discernible but not significant effect on jitter across Emotion (χ^2 = 6.12, p = .106). Gender, however, showed a significant impact on jitter, with male speakers displaying higher levels of jitter (Estimate = 0.0051032, SE = 0.0003292, t = 15.500, p < 0.001). The interaction between emotion and gender was also significant

(χ^2 = 19.68, p < .001). Compared to the Happy emotion, jitter levels were significantly increased in Angry speech only for male speakers (Estimate = 0.0011181, SE = 0.0002299, t = 4.864, p = 0.000311). Also, as shown in Figure 6, jitter was substantially higher in Sad speech for male speakers, while female speakers significantly decreased jitter in Sad speech (Estimate = 0.0034832, SE = 0.0006851, t = 5.085, p = 0.003275). As higher jitter indicates a greater frequency variation or instability in pitch, the Angry and Sad emotions expressed by male speakers are likely to have resulted in a more raspy or unsteady voice quality.

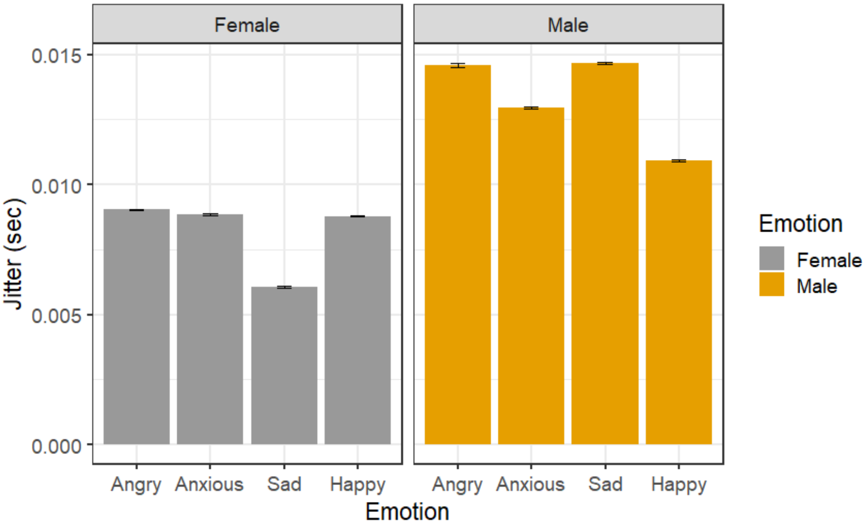

Figure 6. Jitter for four emotions by gender

4.7 Shimmer

Emotion returned a significant influence on voice shimmer (χ^2 = 9.21, p = .028). Shimmer significantly increased in Angry speech (Estimate = 0.011688, SE = 0.001436, t = 8.138, p < 0.001), indicating greater intensity variation compared to Happy speech. On the other hand, Anxiety led to a decrease in shimmer (Estimate = -0.005757, SE = 0.002216, t = -2.598, p = 0.04062). Gender alone also significantly influenced shimmer: the male speakers revealed overall higher shimmer values than the female speakers (Estimate = 0.022735, SE = 0.002340, t = 9.714, p < 0.001). Anger and anxiety altered the vocal shimmer differently for the male speakers. Anger increased the intensity variation in speech, making the voice sound more rough or breathy, whereas Anxiety lowered the variation, resulting in steadier or less rough voice quality. As illustrated in Figure 7, the interaction between emotion and gender also revealed a significant increase in shimmer for Sad speech by the male speakers (Estimate = 0.008981, SE = 0.002851, t = 3.150, p = 0.00953). The female speakers' shimmer values for sadness was lower than the male speakers, which indicates a more controlled and less dynamic vocal expression.

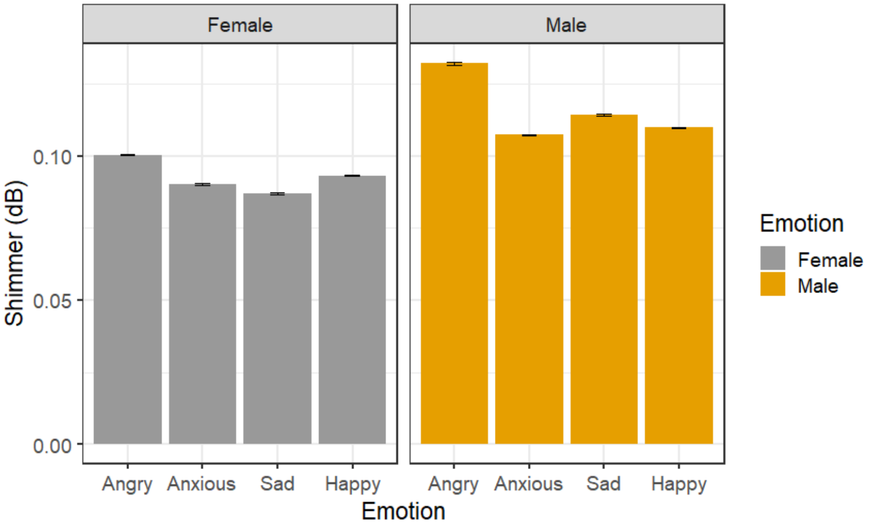

Figure 7. Shimmer for four emotions by gender

4.8 HNR (Harmonic-to-Noise ratio)

Emotion significantly influenced the mean Harmonics-to-Noise ratio (HNR) (χ^2 = 22.34, p < .001). A significantly reduced HNR for Angry emotion (Estimate = -1.9745, SE = 0.2224, t = -8.877, p < 5.02e-06) suggested a compromised vocal clarity for both genders. Anxiety substantially increased HNR (Estimate = 1.4853, SE = 0.1854, t = 8.010, p < 2.17e-06), indicating a clearer and more stable voice quality, whereas Sad emotion did not significantly affect HNR (Estimate = 0.1430, SE = 0.3748, t =

0.382, p = 0.717976). Gender also significantly influenced HNR, with the male speakers showing a lower mean HNR (Estimate = -2.8299, SE = 0.3483, t = -8.124, p < 7.15e-06) than the female speakers. The interaction between emotion and gender revealed that gender modifies the effects of emotion on HNR, particularly for Angry and Anxious emotions (p = 0.001517 and p = 0.000889, respectively). As shown in Figure 8, the male speakers exhibited a significant decrease in HNR for Angry, which signals an increase in noise components in voice quality. Interestingly, anxiety led to increased vocal clarity in male speech, possibly reflecting a modulation to maintain control or composure.

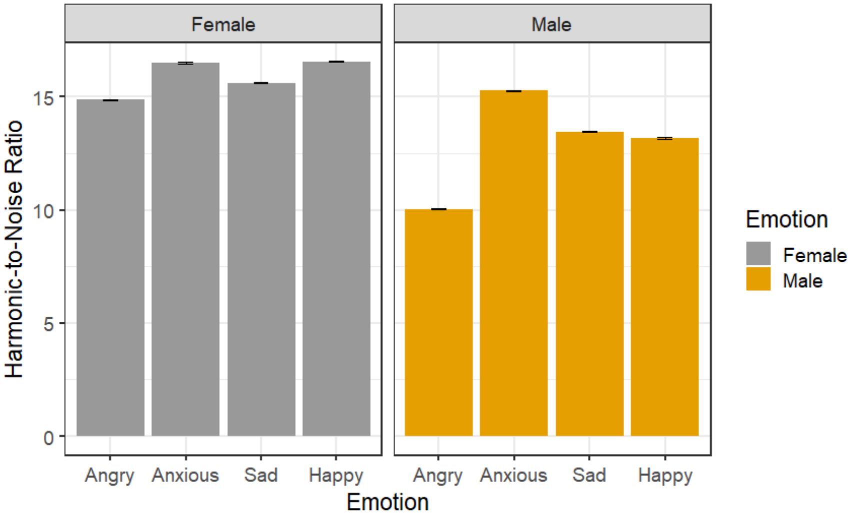

Figure 8. HNR for four emotions by gender

The following table captures the effects of emotion on acoustic characteristics. Across different emotions, certain speech characteristics display universal patterns. Anxiety consistently results in shorter speech duration, reflecting a sense of urgency. In contrast, sadness leads to longer speech, suggesting a more drawn-out and deliberate pace. The F0 cues were also affected by sadness, which generally lowered the median F0 and reduced its variability. Additionally, anxiety appeared to increase loudness, while sadness was delivered in a more quiet voice. Gender-specific differences were also prominent in how emotions were influenced. When angry, the female speakers exhibited increased variability in loudness, reflected in heightened intensity variability. The male speakers showed increased jitter when sad, whereas female speakers display decreased jitter, resulting in a more stable pitch. Anger decreased HNR for both genders, while anxiety improved vocal clarity, particularly in male speech. While no single acoustic feature may be sufficient to accurately detect emotions in speech, the incorporation of speaker gender information can play a crucial role in understanding how various acoustic features collectively contribute to speech emotion recognition.

Table 1. Acoustic characteristics of emotional speech by gender (M, F)

Features	Angry	Anxious	Sad
Speech duration	-	Decrease	Increase
F0 median	-	-	Decrease
F0 variability	-	-	Decrease
F0 range	-	-	Increase (M), Decrease (F)
Median Intensity	-	Increase	Decrease
Intensity range	-	-	Decrease
Intensity variability	Increase (F)		Decrease (F)
Jitter	Increase (M)	-	Increase (M), Decrease (F)
Shimmer	Increase	Decrease	-
HNR	Decrease	Increase	-

5. Discussion

The current study provides a comprehensive analysis of how emotional states influence speech acoustics, particularly focusing on phonetic markers of emotions such as happiness, sadness, anxiety, and anger. The findings highlight significant effects of emotion on speech duration, fundamental frequency (F0), intensity, and voice quality characteristics, offering an insight into the interplay between gender, societal norms, and vocal emotion expression. The analysis showed that both male and female speakers exhibited significant changes in median F0, F0 range, and F0 variability when expressing different emotions. However, the patterns of these changes differed by gender. Female speakers expressed sadness with a low F0, reduced F0 range and variability, suggesting a more stable pitch which could indicate an attempt to maintain emotional neutrality or restraint. These findings align with prior research on gender-specific vocal strategies. For instance, Schirmer and Kotz (2006b) observed that women tend to stabilize pitch to convey sadness, reflecting emotional composure. Conversely, male speakers' increased pitch variability in sadness aligns with observations by Belyk and Brown (2014), suggesting a broader emotional expressiveness. These results support the hypothesis that societal norms and cultural expectations influence gendered patterns in emotional vocalization (Laukka et al. 2014). Male speakers, on the other hand, showed a low F0, minimal F0 range and variability in anger, but a significant increase in F0 range when sad, suggesting a broader range of pitch fluctuation in sadness compared to other emotional states.

These gender-specific patterns were also observed in intensity, suggesting that there may be inherent differences in how males and females acoustically express certain emotions. Female speakers did not differ in median intensity across the emotions. They exhibited high intensity range and variability when expressing anger and the lowest intensity range when expressing sadness. Male speakers, however, demonstrated the lowest median intensity when expressing sadness, indicating an overall subdued voice amplitude. They also displayed the least variability in intensity when conveying anger, suggesting a more uniform level of expressiveness. Interestingly, for male speakers, the variability in intensity for sadness was as high as for happiness, indicating a similar level of expressive variation for the two emotionally contrasting states. In general, male speakers showed a tendency to convey happiness through a broad range and variability in intensity, whereas female speakers demonstrated high intensity variability specifically in expressing anger. Our findings suggest that the emotional vocalizations of Korean speakers reflect deeply embedded cultural norms regarding gender and emotion.

In examining voice-quality characteristics, the study analyzed jitter, shimmer, and the HNR across four emotional states. Jitter and shimmer indicate irregularities in pitch and intensity, respectively, and these variations were observed to differ between male and female speakers in emotional expression. Specifically, jitter was elevated for both genders when expressing anger, suggesting a common trend in the heightened irregularity of pitch in angry speech. However, a different pattern was shown with sad voices: female speakers exhibited notably lower jitter, indicating more stable pitch in sadness, whereas male speakers showed a significant increase in jitter, potentially giving the sad voice a more unpredictable and shaky quality. Regarding shimmer, an increase was observed for expressions of anger and anxiety compared to happiness, with the rise being particularly pronounced for male speakers when expressing anger (see Figure 9). Anger, typically associated with heightened emotional intensity, may have led to greater vocal effort, thereby elevating shimmer values. A high HNR indicates a clearer, more harmonic-rich voice with fewer noise components. Figure 10 shows that female speakers have higher HNR than male speakers across all four emotions. Anger was linked to a lower HNR for both genders, with this effect being particularly pronounced in male speakers. In contrast, anxiety resulted in greater vocal clarity in male speech. Interestingly, anxiety led to an increase in vocal clarity particularly in male speech. Male speakers may have modulated their voice to maintain a semblance of control or composure in an effort to downplay their anxious feelings. This control can manifest as a more stable vocal production, which inadvertently results in a higher HNR. Overall, these vocal characteristics—high jitter, high shimmer, and low HNR—in male speakers' voices when expressing negative emotions suggest a vocal output that is less stable, less smooth, and less clear.

The findings suggest that male and female may process and express emotions differently. According to the results, male may have expressed anger with more physical tension, leading to greater irregularities in pitch (jitter) and intensity (shimmer). This tension can also contribute to a lower HNR, indicating a rougher, less clear voice. The gender differences may be more distinctive in Korean society due to its rigid gender roles in cultural and social norms. For male speakers, societal norms may encourage the expression of anger through more dominant and forceful vocal quality cues, including jitter and shimmer, and a HNR. This vocal roughness and stability may be perceived as strength or authority. Female speakers, on the other hand, might be socialized to express emotions more subtly and variably, resulting in different acoustic profiles especially for angry emotions. High intensity variability in female speech of anger may suggest a strategic negotiation with traditional femininity. Female speakers appear to have used attenuated vocal cues to communicate varying degrees of displeasure within socially acceptable bounds. The implications of these findings extend beyond the acoustic properties of emotional speech. They reveal how gendered vocal strategies may serve as adaptive mechanisms for navigating societal expectations. For example, in a culture like Korea’s, where traditional gender roles are prominent, the observed differences in male and female vocal expressions may reflect broader societal constructs of emotional regulation. Future research should explore whether these patterns are consistent across languages and cultural contexts, contributing to a global understanding of emotional prosody.

In conclusion, the findings emphasize that while both male and female speakers utilize various acoustic cues to express emotions, the realization of these cues differs significantly by gender. The study shows that Korean female speakers exhibit less variation in F0 for sadness and rely more on intensity for anger, whereas Korean male speakers demonstrate less F0 variation and rely more on intensity and voice-quality cues for sadness and anger. These findings also demonstrate the distinct roles of acoustic cues, as highlighted in the context of sadness in female speakers. This distinction suggests that voice quality and range may serve as complementary markers of emotional depth and subtlety. These insights contribute to a deeper understanding of the complex interplay between gender, societal norms, and vocal expression of emotions. Further research is needed to explore how these patterns hold across different languages and cultures, and how they are influenced by societal expectations and norms.

References

Banse, Rainer and Klaus R. Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology* 70(3): 614-636.

Bänziger, Tanja and Klaus R. Scherer. 2005. The role of intonation in emotional expressions. *Speech Communication* 46(3-4): 252-267.

Belyk, Martin and Steven Brown. 2014. Perception of affective and linguistic prosody: An ALE meta-analysis of neuroimaging studies. *Social Cognitive and Affective Neuroscience* 9(9): 1395-1403.

Bryant, Gregory and H. Clark Barrett. 2008. Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture* 8(1-2): 135-148.

Chaplin, Tara M. and Amelia Aldao. 2013. Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin* 139(4): 735-765.

El Ayadi, Moataz, Mohamed S. Kamel, and Fakhri Karray. 2011. Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition* 44(3): 572-587.

Gobl, Christer and Ailbhe Ní Chasaide. 2003. The role of voice quality in communicating emotion, mood, and attitude. *Speech Communication* 40(1-2): 189-212.

Hall, Judith A. 1978. Gender effects in decoding nonverbal cues. *Psychological Bulletin* 85(4): 845-857.

Ishii, Keiko, Jesus A. Reyes, and Shinobu Kitayama. 2003. Spontaneous attention to word content versus emotional tone: Differences among three cultures. *Psychological Science* 14(1): 39-46.

Javanbakht, Arash, Shimon Tompson, Shinobu Kitayama, Abigail King, Clara Yoon, and Israel Liberzon. 2018. Gene by culture effects on emotional processing of social cues among East Asians and European Americans. *Behavioral Sciences* 8(7): 62.

Johnstone, Tom and Klaus R. Scherer. 2000. Vocal communication of emotion. In Michael Lewis and Jeannette M. Haviland-Jones (eds.), *Handbook of emotions*, second edition, 220-235. New York, NY: Guilford Press.

Jürgens, Rainer, Julia Fischer, and Annekathrin Schacht. 2018. Hot speech and exploding bombs: Autonomic arousal during emotion classification of prosodic utterances and affective sounds. *Frontiers in Psychology* 9: 228.

Juslin, Patrik N. and Petri Laukka. 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin* 129(5): 770-814.

Kim, Sang Jin, Kyungkoo Kim, and Minsoo Hahn. 2004. Study on emotional speech features in Korean with its application to voice color conversion. *Proceedings of Interspeech 2004,* 1849-1852.

Ladd, D. Robert, Katharine E. A. Silverman, Friedrich Tolkmitt, Gerald Bergmann, and Klaus R. Scherer. 1985. Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America* 78(2): 435-444.

Laukka, Petri, Daniel Neiberg, and Hillary Anger Elfenbein. 2014. Evidence for cultural dialects in vocal emotion expression: Acoustic classification within and across five nations. *Emotion* 14(3): 445-449.

Lausen, Anika and Annekathrin Schacht. 2018. Gender differences in the recognition of vocal emotions. *Frontiers in Psychology* 9: 882.

Lewis, Michael and Jeannette M. Haviland-Jones. 1993. *Handbook of emotions.* New York, NY: Guilford Press.

Liu, Ping, Simon Rigoulot, and Marc D. Pell. 2016. Cultural immersion alters emotion perception: Neurophysiological evidence from Chinese immigrants to Canada. *Social Neuroscience* 12(6): 685-700.

Matsumoto, David, Traci Consolacion, Hiroshi Yamada, Rieko Suzuki, Bonnie Franklin, and Sumie Paul. 2002. American-Japanese cultural differences in judgements of emotional expressions of different intensities. *Cognition and Emotion* 16(6): 721-747.

Patel, Sejal and Klaus R. Scherer. 2011. Vocal intensity as a reliable cue in the detection of high-arousal emotions. *Emotion Review* 3(3): 327-329.

Paulmann, Silke and Sonja A. Kotz. 2008. Early emotional prosody perception based on different speaker voices. *Neuroreport* 19(2): 209-213.

Pell, Marc D. and Vera Skorup. 2008. Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication* 50(6): 519-530.

Sauter, Disa A., Frank Eisner, Paul Ekman, and Sophie K. Scott. 2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences* 107(6): 2408-2412.

Scherer, Klaus R. 1986. Vocal affect expression: A review and a model for future research. *Psychological Bulletin* 99(2): 143-165.

Scherer, Klaus R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40(1-2): 227-256.

Scherer, Klaus R. and Judith S. Oshinsky. 1977. Cue utilization in emotion attribution from auditory stimuli. *Motivation and Emotion* 1(4): 331-346.

Scherer, Klaus R., Tom Johnstone, and Gunnar Klasmeyer. 2003. Vocal expression of emotion. In Richard J. Davidson, Klaus R. Scherer, and H. Hill Goldsmith (eds.), *Handbook of affective sciences,* 433-456. Oxford: Oxford University Press.

Scherer, Klaus R., Rainer Banse, Harald G. Wallbott, and Thomas Goldbeck. 2001. Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology* 32(1): 76-92.

Schirmer, Annett and Sonja A. Kotz. 2006a. Beyond the right hemisphere: Brain mechanisms mediating vocal emotional processing. *Trends in Cognitive Sciences* 10(1): 24-30.

Schirmer, Annett and Sonja A. Kotz. 2006b. Sex differences in the pre-attentive processing of vocal emotional expressions. *Neuroreport* 16(6): 635-639.

Schuller, Björn, Markus Wöllmer, Florian Eyben, and Gerhard Rigoll. 2009. Prosodic, spectral or voice quality? Feature type relevance for the discrimination of emotion pairs. In Sandra Hancil (ed.), *The role of prosody in affective speech*, 285-307. Bern: Peter Lang.

Singmann, Henrik, Ben Bolker, John Westfall, and Frederik Aust. 2015. *Afex: Analysis of factorial experiments* (R package version 0.16-1). Retrieved from https://CRAN.R-project.org/package=afex.

Thompson, William Forde and Laura L. Balkwill. 2006. Decoding speech prosody in five languages. *Semiotica* 158(1-4): 407-424.

Thompson, William Forde and Laura L. Balkwill. 2009. Cross-cultural similarities and differences. In Patrik N. Juslin and John A. Sloboda (eds.), *Handbook of music and emotion: Theory, research, applications*, 755-791. New York, NY: Oxford University Press.

Tolkmitt, Friedrich J., Elisabeth Krahmer, Gerhard Oberbeck, and Klaus R. Scherer. 1988. The identification of emotional vocal expressions across languages. *Journal of Cross-Cultural Psychology* 19(1): 55-74.

Zhang, S. and Marc D. Pell. 2022. Cultural differences in vocal expression analysis: Effects of task, language, and stimulus-related factors. *PLOS ONE* 17(10): e0275915.

Eunhae Oh
Professor
Department of English Language and Literature
Konkuk University
120 Neungdong-ro, Gwangjin-gu
Seoul 05029, Korea
E-mail: gracekonkuk@gmail.com

Received: 2024. 11. 04
Revised: 2024. 12. 10
Accepted: 2024. 12. 12