BILINGUAL PUBLISHING GROUP

Forum for Linguistic Studies

https://journals.bilpubgroup.com/index.php/fls

ARTICLE

Linguistic Prosody and Melodic Characteristics of Korean Emotion Vocabulary: A Musical-Linguistic Analysis

Soo Yon Yi [1*] , Hyun Ju Chong [2]

[1] *Department of Music Education, The Graduate School, Gachon University, Seongnam-si 13120, Korea*

[2] *Department of Music Therapy, The Graduate School, Ewha Womans University, Seoul 03760, Korea*

ABSTRACT

This study explores the intersection between linguistic prosody and melodic characteristics in Korean emotion vocabulary, aiming to quantify emotional articulation in spoken language and inform melodic construction in songwriting and lyrical composition. Thirty Korean female participants (ages 19–23) were asked to speak emotion-related words representing three target emotions: happiness, anger, and sadness. Acoustic analyses were conducted to examine prosodic features, including fundamental frequency (Hz) and articulation time (ms), which were then translated into musical parameters such as pitch register, pitch range, melodic contour, and tempo. Statistical analyses identified significant differences in prosodic and melodic characteristics across emotional categories. Results showed that the mean pitch corresponded to B3 (253.7 Hz) for happy words, A3 (213.1 Hz) for angry words, and G3 (211.6 Hz) for sad words. Happy words featured high pitch registers, wide ranges, and descending contours; angry words had mid-range registers with rising-falling or descending contours; and sad words exhibited low registers, narrow ranges, and either descending or unisonous contours. In terms of tempo, angry words were articulated most quickly (172 ms), followed by happy (191 ms) and sad (210 ms). Significant differences were found in frequency between happy and angry words (18.5 Hz), and in articulation time between happy and sad (0.02 ms), and angry and sad (0.03 ms). These findings suggest that the prosodic expression of emotion can be meaningfully translated into melodic representation, with potential applications in music composition, song therapy, and

*CORRESPONDING AUTHOR:

Soo Yon Yi, Department of Music Education, The Graduate School, Gachon University, Seongnam-si 13120, Korea; mtsoo@gachon.ac.kr

ARTICLE INFO

Received: 18 July 2025 | Revised: 25 July 2025 | Accepted: 4 August 2025 | Published Online: 24 September 2025

DOI: https://doi.org/10.30564/fls.v7i10.11138

CITATION

Yi, S.Y., Chong, H.J., 2025. Linguistic Prosody and Melodic Characteristics of Korean Emotion Vocabulary: A Musical-Linguistic Analysis. Forum for Linguistic Studies. 7(10): 301–317. DOI: https://doi.org/10.30564/fls.v7i10.11138

affective computing. This framework establishes a foundation for future interdisciplinary exploration linking language, music, and emotion.

Keywords: Linguistic Prosody; Melodic Characteristics; Korean Emotion Vocabulary; Pitch; Pitch Range

1. Introduction

Emotion refers to changes in mental and physical states[1], triggered by the perception of external stimuli and accompanied by psychological and physiological responses[2]. Basic emotions—such as happiness, anger, sadness, and calmness—are commonly categorized along two dimensions: emotional valence (positive/negative) and arousal level (high/low)[3–5]. These emotions can be expressed through both verbal (spoken) and non-verbal (e.g., musical) means.

While a variety of channels exist for emotional expression, identifying efficient and nuanced methods for conveying emotional experiences remains a significant research interest. Emotion-related vocabulary plays a critical role in verbal expression, enabling individuals to label and communicate emotional states. Within spoken language, prosodic features—such as intonation, intensity, and duration—serve as key carriers of emotional nuance and communicative intent[6–8]. Empirical studies have shown that prosodic variations, especially in intonation patterns, differ systematically depending on the emotion being conveyed[9–11]. In the context of Korean, research indicates that frequency range and speech rate vary with emotional tone, and that prosodic features also differ by syllable count and phoneme segmentation[12–15].

Cross-linguistic studies further support the functional significance of prosody. For example, it has been shown that prosodic features of discourse markers in spontaneous Mandarin vary depending on their semantic and pragmatic roles[16], while distinct pitch contours between interrogatives and declaratives have been identified in the Chengdu dialect[17]. These findings reinforce the idea that prosody encodes both emotional and structural functions in language, making it a key modality for nuanced communicative intent across languages.

As interest grows in more intuitive and expressive forms of emotional communication, music has emerged as a compelling medium for emotion expression beyond traditional language. Songs, which combine lyrics with melodic elements, offer a multidimensional platform for conveying emotional content[18,19]. While lyrics communicate semantic meaning, melodies reinforce and amplify emotional tone[20]. Composers often incorporate prosodic cues into melody construction to align musical and emotional intent. Research further supports the notion that musical elements such as melodic contour, pitch range, tempo, rhythm, intensity, and dynamics can convey emotion effectively, especially when paired with emotionally charged lyrics[21–24]. Some scholars have suggested that the prosodic structure of speech should serve as a model when composing emotional melodies[25].

However, previous studies have primarily addressed emotional expression separately within the domains of language or music, with limited systematic investigation into the structural connections and mutual convertibility between these two fields. Research converting prosodic features of linguistic emotional expression into musical elements can provide important theoretical and practical foundations not only for music composition but also for applied fields such as music therapy.

In pursuit of this integration, it is essential to explore the parallels between prosodic features in speech and musical characteristics. Prior studies have demonstrated perceptual and acoustic similarities between spoken prosody and musical melody[11,26]. Both domains involve shared acoustic dimensions: pitch (frequency in Hz), duration (ms), and intensity (dB). For example, pitch can be perceived as high or low in both speech and melody; duration reflects rhythm or speech rate; and intensity corresponds to loudness in speech and dynamic emphasis in music. These shared features suggest that emotional expressions in language and music can be meaningfully compared and converted[27,28].

The aim of this study is to investigate how the linguistic prosodic characteristics of Korean emotion vocabulary can be converted into melodic characteristics through acoustic analysis. Specifically, the study analyzes fundamental frequency and articulation time of emotion vocabulary and maps these prosodic parameters onto melodic components such as pitch register, pitch range, melodic contour, and tempo. Using Praat software, prosodic features are extracted and interpreted within a musical framework based on the equal-tempered semi-tone scale. This interdisciplinary approach provides foundational data applicable to musical composition and therapeutic practices, contributing to the integration of music, linguistics, and emotional expression. The present study addresses the following research questions:

What prosodic features (e.g., fundamental frequency, articulation time) characterize Korean emotion-related vocabulary across distinct emotional categories (happiness, anger, sadness)?
How can prosodic attributes of Korean emotional vocabulary be systematically transformed into melodic parameters (e.g., pitch register, contour, tempo)?
What differences in derived melodic characteristics are observed across emotional categories, and how do these reflect the expressive intent of the original spoken vocabulary?

2. Literature Review

2.1. Theoretical Link between Linguistic Prosody and Song Melodic Features: Roles in Emotional Expression and Interconvertibility

Linguistic prosody refers to suprasegmental variations—such as pitch, duration, intensity, and rhythm—superimposed on speech, which play a crucial role in conveying emotional and pragmatic meaning beyond lexical content [6]. These prosodic parameters—fundamental frequency (F0), speech rate, and intensity—systematically vary according to emotional valence (positive/negative) and arousal (high/low) [3,6,9]. For example, high-arousal emotions such as happiness and anger are typically characterized by elevated pitch and faster speech rate, whereas sadness is marked by lower pitch and slower tempo [12,13].

The expressive function of prosody in spoken language closely parallels that of musical elements—melody, rhythm, and dynamics—in vocal music [21,23]. Both speech and music rely on the same core acoustic dimensions: pitch (frequency), time (duration or tempo), and intensity (amplitude) [26]. Given this overlap, variations in speech prosody—such as pitch contour, articulation rate, loudness—can be systematically mapped onto corresponding musical parameters like melodic pattern, tempo, and dynamics [28]. While prosodic cues may vary in their influence across developmental stages [29], the emotional meaning embedded in prosody can nevertheless be effectively transferred to music, with listeners accurately identifying emotions conveyed in both speech and melody as demonstrated by cross-modal emotion recognition tasks [30]. These shared acoustic foundations also reflect underlying cognitive and perceptual processing commonalities between language and music, suggesting a deep-seated interconnection between the two domains [31]. A compelling illustration of this shared system can be found in songs, where linguistic and musical elements work in tandem to heighten emotional expression. In this context, the fusion of lyrics with melody exemplifies how both systems contribute structurally and affectively to the communication of emotion [32].

Research has shown that individuals with stronger musical abilities tend to exhibit enhanced prosodic perception, suggesting that pitch and timing sensitivity are shared across both linguistic and musical domains [33]. Additionally, prosodic discrimination skills have been found to mediate the relationship between musical aptitude and vocal emotion recognition ability, indicating common processing mechanisms for emotional cues in speech and music [34]. These findings reinforce the feasibility of systematically translating emotional prosody into melodic features and provide empirical support for cross-domain applications.

Studies demonstrating a strong association between musical prosody and emotion have involved having vocalists listen to their own singing and classify the emotions conveyed. These studies examined the relationship between musical features and prosodic characteristics, confirming that musical prosody is integrally linked to both acoustic and linguistic features [33]. A comprehensive review of nearly three decades of research on emotional prosody discussed how linguistic and musical elements share acoustic dimensions such as pitch, duration, and intensity, which are crucial for conveying emotions. The review supports the idea that linguistic prosody and musical melodies share structural features for emotional expression [34]. Another study employed a recently developed measure of speech rhythm to compare the rhythmic patterns of English and French language and classical music. The results showed significant differences between English and French musical themes according to this rhythm metric, which also effectively distinguished the rhythm of spoken English and French. Therefore, the claim that spoken prosody influences the music of a culture is supported by empirical evidence [35].

These converging findings indicate that linguistic prosody and musical structure are underpinned by shared acoustic and affective dimensions—particularly pitch, rhythm, and intensity—central to emotional expression. Moreover, empirical studies show that prosodic patterns in speech can shape musical organization across perceptual and cultural levels. This accumulating evidence strengthens the theoretical link between linguistic prosody and song melody, offering a robust foundation for research on their functional interconvertibility in emotional communication.

2.2. An Empirical Study of Prosody, Melody, and Acoustic Characteristics in Korean Emotional Expression

Recent empirical studies have explored the acoustic underpinnings of Korean emotional prosody and its correspondence to musical elements. This chapter aims to review empirical studies that acoustically analyze the prosodic features of emotional speech and the melodic characteristics of song, as well as investigations exploring the interrelation between these two domains.

A study analyzing Korean prosody examined how acoustic parameters—such as pitch (F0), duration, and intensity—function in conveying emotional and semantic meaning. This research experimentally investigated how these acoustic cues encode the linguistic and emotional functions of Korean prosody. It systematically presented prosodic differences according to sentence types and intonation patterns, offering both theoretical and methodological insights for the acoustic analysis of emotional expression [36].

Building on foundational work in Korean prosody, comparative studies have explored how such acoustic parameters function across languages. A study analyzing emotional prosody in Korean, English, and Chinese used Praat to examine F0 and intensity in happiness, anger, and sadness across 180 sentences from films and TV shows. Results showed English had minimal F0 and intensity variation, while Korean and Chinese expressed sadness with notably lower F0 and intensity than happiness and anger. This highlights how acoustic analysis provides objective measures for cross-cultural comparisons of emotional speech [37].

The empirical study acoustically analyzed emotional expressions produced in two typologically distinct languages, examining how acoustic features vary according to emotion, language, and speaker gender. Using 17 acoustic parameters—including pitch (F0), intensity, spectral features, duration, and perturbation—the study identified emotion-specific acoustic profiles that systematically differed across languages and between male and female speakers. Such findings underscore the cross-linguistic and gender-dependent variability in emotional prosody [38].

An empirical investigation was conducted into the linguistic prosodic features of Korean and their connection to music. The intonation patterns of Korean emotion words were analyzed to acoustically characterize their melodic features. By systematically classifying pitch and stress variations associated with different emotions, the study explored the relationship between phonological and acoustic properties and the melodic characteristics in Korean emotional expression [39].

Beyond identifying prosodic features, some studies have investigated how these acoustic elements can be systematically translated into musical structures. An acoustic analysis of Korean speech was conducted using data collected from 30 male and 30 female participants, analyzed with the Praat software. The results showed that interrogative sentences exhibited higher pitch, while declarative sentences had longer durations. Female speakers demonstrated both higher pitch and longer durations compared to male speakers. These acoustic features were then converted into musical notation, reflecting pitch contour, duration, and accent patterns. This study highlights the potential of transforming the acoustic characteristics of Korean speech into musical elements [40].

In addition to language-specific characteristics, gender differences have also emerged as a significant factor influencing emotional prosody. A study was conducted to analyze acoustic differences in Korean emotional speech—specifically the expressions of happiness, sadness, anxiety, and anger—according to gender. The results showed that female speakers exhibited stable pitch when expressing sadness and greater variability in intensity when expressing anger. In contrast, male speakers demonstrated increased pitch variability in sadness and irregularities in vocal quality during anger. These findings suggest that acoustic analysis can provide valuable insight into how emotional speech expression differs by gender [41].

While not directly focused on prosodic features, related work in sound quality vocabulary offers further evidence of the music-language interface. A study was conducted to analyze the semantic properties of Korean vocabulary used for evaluating sound quality in music listening. Specifically, it examined how the vocabulary employed to assess perceived sound quality varies according to different music genres, providing foundational data on the semantic characteristics of sound quality descriptors. The study identified vocabulary categories that are sensitive to acoustic attributes. Although it did not directly analyze the phonetic properties of the vocabulary itself, the research highlights the close relationship between music and language, suggesting notable parallels between these two domains [42].

Taken together, these empirical investigations offer a cohesive view of how emotion is encoded through prosody and reflected in musical expression. The reviewed studies have empirically analyzed acoustic parameters such as pitch, duration, and intensity to examine the prosodic and melodic structures underlying emotional expression. These findings confirm that prosody and melody share common mechanisms for encoding emotion. This integrative perspective offers a valuable theoretical framework for investigating the interplay between spoken language and music in emotional expression, particularly within the Korean context.

2.3. Rationale and Significance of the Present Study

Recent empirical research has increasingly focused on the acoustic foundations of emotional prosody and its parallels in music, underscoring the critical role of vocal expression in conveying affective states. Language and music, while traditionally regarded as separate domains, share core acoustic parameters—such as pitch, rhythm, and intensity—that are fundamental to emotional communication. Prosody in speech and melody in music both rely on these parameters to encode and express emotion in ways that are perceptually salient and culturally embedded.

Previous studies have independently analyzed the acoustic characteristics of linguistic prosody and musical melody to elucidate their respective roles in emotional expression [36,39,40]. However, integrated research that simultaneously investigates the interaction between these two structures within the context of Korean emotional expression remains relatively limited. While there are studies decoding emotional and semantic information through prosodic features of Korean, such as fundamental frequency (F0), duration, and intensity [36,41], few have examined how linguistic prosody and musical melody might be structurally and cognitively interconnected.

Given the shared reliance on pitch, timing, and dynamic variation, as well as overlapping perceptual and cognitive processing mechanisms between language and music [26,28,31,34], a convergent analytical approach is not only feasible but necessary. Moreover, studies identifying cross-linguistic and gender-based variations in acoustic emotional cues further highlight the need for language-specific, data-driven exploration of how emotion is vocally encoded [37,38].

Accordingly, the present study aims to acoustically analyze the prosodic features of Korean emotion-related vocabulary and convert these features into corresponding melodic structures. By transforming linguistic prosody into musical melody, this study explores the structural continuity between speech and music in affective expression. The findings are expected to enhance theoretical understanding of prosody-melody convergence and to provide practical insights for interdisciplinary applications, including affective computing, music therapy, and emotionally responsive design.

To pursue these objectives, the study adopted a two-phase methodology. First, we conducted a detailed acoustic analysis of Korean emotional vocabulary spoken by native speakers, examining parameters such as fundamental frequency and articulation time. Second, these prosodic features were systematically mapped onto musical dimensions including pitch register, melodic contour, and tempo. The following chapter outlines the procedures used for data collection, analysis, and prosody-to-melody transformation.

3. Materials and Methods

3.1. Participants

This study was approved by the Ewha Womans University Institutional Review Board (IRB No. 115-8) and involved 30 female university students aged 19 to 23 (M = 20.2). This selection was based on prior research indicating that the ability to use prosody stabilizes during adulthood, approximately between the ages of 20 and 40 [29]. Preliminary investigations also confirmed that male participants exhibited a more limited range of prosodic variation. Furthermore, previous studies analyzing prosodic characteristics by gender reported that females tend to employ more dynamic prosodic patterns [38]. Consequently, this study restricted the sample to female participants only, excluding males. Participants were recruited via notices posted on a university bulletin board.

The inclusion criteria were as follows. First, only female participants were selected due to their relatively consistent intonation patterns. Prior studies have shown that males tend to exhibit greater intonational variation, which can reduce the reliability of prosodic data [29,39,40]. Thus, restricting the sample to females enhanced the consistency of acoustic measurements. Second, participants were required to be native speakers of standard Korean without regional dialects. Dialectal variation can significantly influence intonation patterns [43], so participants were interviewed to verify that their speech conformed to the standard dialect. Third, participants had to be able to recognize emotional states and express them verbally in a clear and discernible manner. Written informed consent was obtained from all participants prior to data collection.

3.2. Selection of Emotion Vocabulary

Target vocabulary for each emotion was selected based on three criteria. First, the emotions of interest were happiness, anger, and sadness—three of the most universally recognized basic emotions, classified by valence and arousal [3–5]. Although calmness is sometimes considered a basic emotion, it was excluded due to the inconsistent variability in its prosodic characteristics [39,44]. Second, emotion words with three or four syllables and high frequency of use in daily speech were chosen [45]. Third, since the initial consonants of Korean words can affect intonation (e.g., hard, aspirated, or fricative consonants), vocabulary with diverse starting phonemes was selected to ensure prosodic variation [14]. Based on these criteria, a total of 12 emotion words were selected to represent the three target emotions (see **Table 1**).

3.3. Procedure

Data collection was conducted in a quiet room with ambient noise below 20 dB. Each participant completed two tasks. First, they were asked to sustain the vowel sounds 'a, e, i, o, u' for three seconds each. This task was used to gather baseline data on individual fundamental frequency (f0). Second, participants were instructed to choose one emotion word from each category (happy, sad, angry) and speak about a personal experience related to that emotion. They were encouraged to speak naturally, as if in conversation, to capture spontaneous prosodic patterns.

3.4. Materials

Speech was recorded using an MXL-990 microphone and saved as mono WAV files with a sampling rate of 44,100 Hz and 16-bit resolution. Acoustic analyses were performed using Praat software, which allows for detailed measurement of speech features such as fundamental frequency, pitch, formants, duration, and intensity [13]. Spoken words were segmented into syllables and labeled in Praat. Fundamental frequency (Hz) and articulation time (ms) were analyzed using the Script_toneLabeler tool [35] (e.g., **Figure 1**). Each syllable's frequency was converted to a musical tone based on the equal-tempered semi-tone scale, which divides an octave into 12 equal parts and is used in standard musical tuning [46] (e.g., **Figure 2**).

3.5. Data Collection and Analysis

3.5.1. Participant Homogeneity Check

To confirm group homogeneity, the mean, minimum, and maximum fundamental frequencies (Hz) for each participant's sustained vowel sounds were analyzed. Grubb's test was used to detect outliers. No significant outliers were found (textitp = 0.039), so all data were retained.

3.5.2. Prosodic Comparison by Emotion

To examine prosodic differences between the three emotion categories, one-way ANOVA was conducted on the mean fundamental frequency (Hz) and articulation time (ms). Post-hoc analyses using Scheffé's test were performed to identify specific group differences.

3.5.3. Extraction of Representative Prosodic Patterns

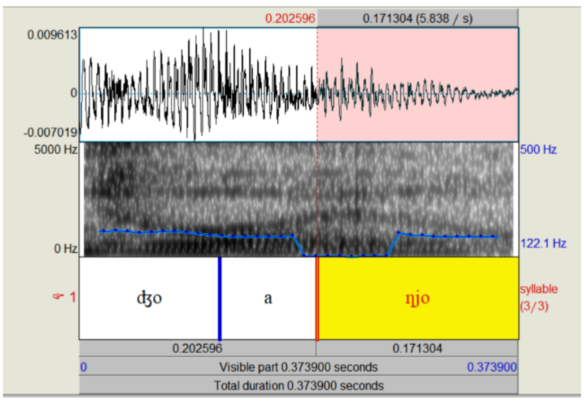
To identify a representative prosodic pattern for each vocabulary item, three steps were followed. First, similar prosodic patterns were clustered using the mclust method. Second, the most representative pattern was selected using maximum likelihood estimation. Third, mean fundamental frequency (Hz) and articulation time (ms) for each syllable in the representative patterns were analyzed.

3.5.4. Conversion to Melodic Characteristics

Prosodic data were transformed into musical features through the following process: Each syllable's frequency (Hz) was converted to a musical pitch using the semi-tone scale. The melodic contour was then derived by connecting the pitch sequence across syllables. Articulation time (ms) for each syllable was converted into a temporal ratio based on a total word duration normalized to 100 ms, which allowed for rhythm and tempo comparisons across words (see **Table 2**).

Table 1. Selected Korean emotion words by initial intonation pattern and syllable count [39].

Emotion	Intonation Pattern	Korean Emotion Words (Meaning)	
		Three Syllables	Four Syllables
Happy	Low starting High starting	ʤo-a-ŋjo (delightful) sɨn-na-ŋjo (fun)	ʤɜl-gʌ-ŋwa-ŋjo (pleasing) hɛŋ-bok-hɛ-ŋjo (joyous)
Angry	Low starting High starting	na-ᴘa-ŋjo (bad) wha-na-ŋjo (furious)	ʌg-ul-hɛ-ŋjo (unfair) ʤɜŋ-o-hɛ-ŋjo (hate)
Sad	Low starting High starting	ʤ-ʧʌ-ŋjo (tired) sl-ᴘʌ-ŋjo (sorrowful)	ʙul-ŋan-hɛ-ŋjo (anxious) ʧm-ul-hɛ-ŋjo (gloomy)



syllable label	f0	semiTone
ʤo	166.5	7.2
a	152.0	5.6
ŋjo	154.5	5.9

Figure 1. Fundamental frequency (Hz) and articulation time (ms) analysis of syllables in vocabularies. Example of syllable segmenting and labeling using Praat (left). Example of fundamental frequency (f0) analysis per syllable using Script_toneLabeler (right) [39].

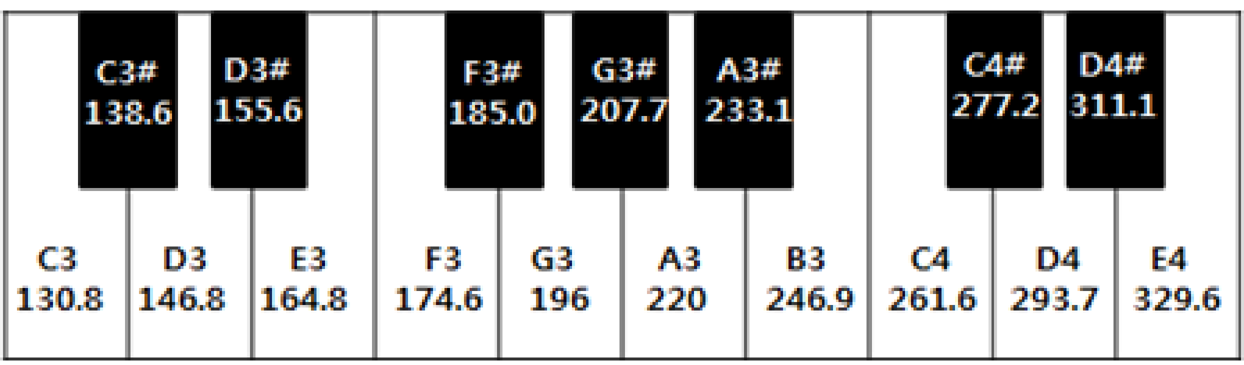


Figure 2. Fundamental frequency (Hz) of mean tone within the semi-tone scale [47].

Table 2. Data collection and analysis.

Purpose	Data Collection and Analysis	Measurement and Method
	⇓	
• To confirm homogeneity among participants	• Mean, maximum and minimum fundamental frequency (Hz) of participants' vowel sounds ('a, e, i, o, u') • Range between minimum and maximum frequency values • Detection of multiple outliers among participants	• Praat program • Grubb's outlier test
	⇓	
• To compare the statistical differences	• Mean frequency (Hz), and articulation time (ms) of 'happy', 'angry' and 'sad' emotion vocabularies • Significant differences among the emotions	• Praat program • Descriptive statistic • One-way ANOVA • Post-hoc: Scheffe's test
	⇓	
• To extract representative prosodic patterns of emotion vocabularies	• Mean frequency (Hz) and mean articulation time (ms) of syllables per representative data	• Mclust method • Maximum likelihood estimation • Praat program
	⇓	
• To obtain melodic characteristics of emotion vocabularies converted from representative prosodic patterns	• Mean tone of syllable converted from mean frequency (Hz) • Melodic contour derived by connecting tones of three- and four- syllables	• Using semi-tone scale
	• Tempo of syllable converted from articulation time (ms) • Rhythm length ratio of three- and four- syllables	• Substituting the total articulation time with 100 (ms) • Converting the time per syllable into a ratio

4. Results

Of the 12 initially selected emotion vocabulary items, 10 were included in the analysis. Two words—na-ᴘa-ɲjo (bad) and ʤɳ-o-hɛ-ɲjo (hate)—were excluded because only 4 (13%) and 2 (7%) participants, respectively, used them during the storytelling task. The low usage frequency indicated that these items were not representative of natural emotional expression in spoken dialogue (see **Table 3**).

The mean fundamental frequency (f_0) for sustained vowel sounds was 206.5 Hz. The minimum and maximum frequencies were 148.1 Hz and 238.9 Hz, corresponding to musical notes D3 and B3, respectively. The mean tone was G3 (see **Table 4**).

Grubb's test for outliers confirmed that all participants fell within the acceptable frequency range ($p = 0.039$), and no outliers were excluded. Grubb's outlier test was conducted to detect any data points falling outside the range of 148.1 Hz (minimum) and 238.9 Hz (maximum). The test yielded a significance probability of 0.039; however, no outliers were identified ($p < 0.05$). Therefore, all participant data were retained for analysis (see **Figure 3**).

Table 3. Analysis of emotion words and number of participants uttered in the target emotions.

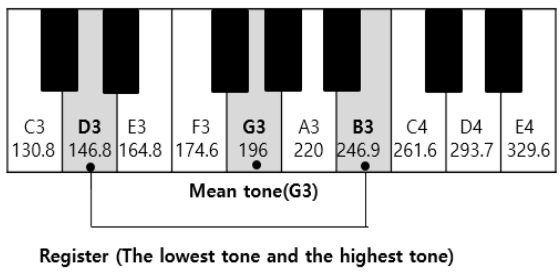
Happy		Angry		Sad	
Emotion Words (Meaning)	**Participants Uttered N (%)**	**Emotion Words (Meaning)**	**Participants Uttered N (%)**	**Emotion Words (Meaning)**	**Participants Uttered N (%)**
ʤo-a-ɲjo (delightful)	16 (53)	*na-ᴘa-ɲjo (bad)	5 (17)	ʤ-ʧʌ-ɲjo (tired)	18 (60)

Table 3. *Cont.*

Happy		Angry		Sad	
Emotion Words (Meaning)	**Participants Uttered N (%)**	**Emotion Words (Meaning)**	**Participants Uttered N (%)**	**Emotion Words (Meaning)**	**Participants Uttered N (%)**
s_{i}n-na-ɲjo (fun)	14 (47)	wha-na-ɲjo (furious)	11 (37)	sl-ᴘʌ-ɲjo (sorrowful)	11 (37)
ʤl-gʌ-ŋwa-ɲjo (pleasing)	16 (53)	ʌg-ul-hɛ-ɲjo (unfair)	16 (53)	ʙul-ŋan-hɛ-ɲjo (anxious)	18 (60)
hɛŋ-bok-hɛ-ɲjo (joyous)	19 (63)	*ʤŋ-o-hɛ-ɲjo (hate)	2 (7)	ʧm-ul-hɛ-ɲjo (gloomy)	14 (47)

*Vocabulary excluded from data analysis.

Table 4. Participants mean fundamental frequencies and pitch ranges.

Fundamental Frequency (Hz)	M±SD	Musical Tone	Register
Mean	206.5 (20.7)	G3	
Minimum	148.1 (31.3)	D3	
Maximum	238.9 (67.2)	B3	

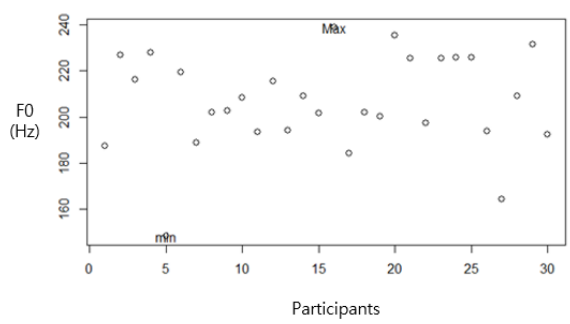


Figure 3. Grubb's outlier test.

4.1. Linguistic Prosodic Characteristics across Emotional Categories in Korean Emotion Vocabulary

The mean frequency (Hz) of emotion was 253.7 Hz for happy, 213.1 Hz for angry, and 211.6 Hz for sad (see **Table 5**). Mean articulation time (ms) was analyzed based on the number of phonological syllables. For three-syllable words, the average articulation times were 170 ms for happy, 144 ms for angry, and 200 ms for sad. For four-syllable words, the values were 170 ms for happy, 159 ms for angry, and 158 ms for sad (see **Table 5**).

One-way ANOVA revealed a statistically significant difference in mean frequency (Hz) among the emotion vocabulary groups, $F(1, 277) = 55.1$, $p < 0.001$ (see **Table 6**). Post-hoc Scheffé analysis indicated a significant difference in mean frequency between happy and angry vocabulary items, with a mean difference of 18.5 Hz. Additionally, significant differences in mean articulation time (ms) were observed between happy and sad words (0.02 ms), and between sad and angry words (0.03 ms), $p < 0.05$ (see **Table 7**).

Table 5. Mean frequency (Hz) and articulation time (ms) of words in the target emotions.

Emotion	Frequency (Hz)		Articulation Time (ms)		
	M	***SD***	**Syllable N.**	***M***	***SD***
Happy	253.7	32.2	3 4	170 170	6 5
Angry	213.1	35.9	3 4	144 159	37 72
Sad	211.6	37.8	3 4	200 158	8 49

Table 6. Statistical differences between emotion words.

Category	SS	df	MS	F	P
Within a group	89383.5	1	89383.5	55.1	0.001*
Between the groups	449278.2	277	1621.9	-	-
Total	538661.6	278	-	-	-

*$P < 0.001$.

Table 7. Statistical differences of mean frequency (Hz) and articulation time (ms) among emotions.

Category	Emotion Vocabularies	*MD*	Std. Error	Lower	Upper	*p*
Mean frequency (Hz)	Happy × Sad	13.5	7.0	−3.7	30.8	0.157
	Happy × Angry	18.5	5.9	4.0	32.9	0.008*
	Angry × Sad	−4.9	7.4	−23.2	13.3	0.800
Mean articulation time (ms)	Happy × Sad	0.02	0.001	−0.001	0.04	0.039*
	Happy × Angry	−0.005	0.009	−0.02	0.01	0.773
	Angry × Sad	0.03	0.008	0.005	0.05	0.012*

*$P < 0.05$.

4.2. Melodic Characteristics Converted from Linguistic Prosody in Korean Emotion Vocabulary

4.2.1. Happy Vocabulary

The melodic characteristics of vocabulary associated with the emotion of happy, derived from prosodic features such as pitch (converted from fundamental frequency), articulation time, and syllable duration ratio. ‘ʤo-a-ŋjo’ (delightful) showed a pitch pattern of A3–G3–F#3 with a total articulation time of 460 ms and a syllable duration ratio of 3:3:4, indicating a slightly lengthened final syllable. ‘s$_{i}$n-na-ŋjo’ (fun) exhibited pitches of B3–A3–G3 over 520 ms, with a 4:2:4 ratio, suggesting rhythmic emphasis on the initial and final syllables. ‘ʤl-gʌ-ŋwa-ŋjo’ (pleasing) had the highest register (G4–F#4), a total duration of 580 ms, and a 4:2:2:2 pattern, reflecting a fast tempo and initial syllable prominence. ‘hεŋ-bok-hε-ŋjo’ (joyous) displayed a pitch sequence of D4–D4–A#3–A3 and the slowest articulation time (720 ms), with a 2:2:3:3 ratio, showing elongation toward the end (see **Table 8**).

4.2.2. Angry Vocabulary

The melodic characteristics of vocabulary associated with the emotion of angry, ‘wha-na-ŋjo’ (furious) showed a pitch contour of C3–A3–F3, with a total articulation time of 410 ms and a syllable duration ratio of 4:2:4, indicating strong emphasis on the first and final syllables. ‘ʌg-ul-hε-ŋjo’ (unfair) exhibited pitches of F#3–G3–E3–D#3 over 660 ms, with a 3:2:2:3 ratio, revealing a slight elongation at the final syllable and balanced pacing in the middle segments (see **Table 9**).

4.2.3. Sad Vocabulary

The melodic characteristics of vocabulary associated with the emotion of sad, ‘ʤ-ʧʌ-ŋjo’ (tired) showed a pitch pattern of F#3–F#3–F3 over a total articulation time of 560 ms, with a syllable duration ratio of 3:2:5, indicating significant elongation on the final syllable. ‘sl-ᴘʌ-ŋjo’ (sorrowful) exhibited a pitch sequence of A3–G#3–G3, with a total duration of

650 ms and a 4:3:3 ratio, reflecting rhythmic prominence at the beginning and stable pacing toward the end. ‘ʙul-ŋan-hɛ-ŋjo’ (anxious) displayed pitches of G#3–G3–F#3–F3 over 630 ms, with a 3:2:2:3 ratio, showing a balanced middle section and gradual lengthening at the final syllable. ‘ʧm-ul-hɛ-ŋjo’ (gloomy) had a pitch contour of A#3–G#3–F3–F3, with a total articulation time of 580 ms and a 4:2:2:2 ratio, highlighting strong initial stress and a generally rapid tempo (see **Table 10**).

Table 8. Melodic characteristics converted from the linguistic prosody of ‘happy’ vocabulary.

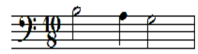
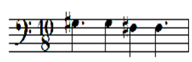
Vocabulary (Meaning)	Syllable	Pitch (Hz)	Musical Note	Time (ms)	Duration Ratio	Prosodic Patterns	Melodic Characteristics
‘ʤo-a-ŋjo’ (delightful)	ʤo a ŋjo	214.0 193.7 180.6	A3 G3 F3#	140 140 180	3 3 4	Fall-fall ↘↘	 ʤo a ŋjo
‘sᵢn-na-ŋjo’ (fun)	sᵢn na ŋjo	253.1 215.8 201.8	B3 A3 G3	220 100 200	4 2 4	Fall-fall ↘↘	 sᵢn na ŋjo
‘ʤl-gʌ-ŋwa-ŋjo’ (pleasing)	ʤl gʌ gʌ gʌ	391.3 386.8 386.8 386.8	G4 G4 G4 G4	220 130 130 130	4 2 2 2	Flat-fall-flat →↘→	 ʤl gʌ ŋwa ŋjo
‘hɛŋ-bok-hɛ-ŋjo’ (joyous)	hɛŋ bok hɛ ŋjo	286.3 287.8 233.2 191.0	D4 D4 A3# A3	170 160 180 210	2 2 3 3	Flat-fall-fall →↘↘	 hɛŋ bok hɛ ŋjo

Table 9. Melodic characteristics converted from the linguistic prosody of ‘angry’ vocabulary.

Vocabulary (Meaning)	Syllable	Pitch (Hz)	Musical Note	Time (ms)	Duration Ratio	Prosodic Patterns	Melodic Characteristics
‘wha-na-ŋjo’ (furious)	wha na ŋjo	254.8 219.0 175.2	C3 A3 F3	160 100 150	4 2 4	Flat-fall →↘	 wha na ŋjo
‘ʌg-ul-hɛ-ŋjo’ (unfair)	ʌg ul hɛ ŋjo	185.5 191.5 165.2 159.1	F3# G3 E3 D3#	200 130 150 180	3 2 2 3	Rise-fall-fall ↗↘↘	 ʌg ul hɛ ŋjo

Table 10. Melodic characteristics converted from the linguistic prosody of ‘sad’ vocabulary.

Vocabulary (Meaning)	Syllable	Pitch (Hz)	Musical Note	Time (ms)	Duration Ratio	Prosodic Patterns	Melodic Characteristics
‘ʤ-ʧʌ-ŋjo’ (tired)	ʤ ʧʌ ŋjo	190.1 187.7 170.5	F3# F3# F3	200 160 200	3 3 5	Flat-fall →↘	 ʤ ʧʌ ŋjo
‘sl-ᴘʌ-ŋjo’ (sorrowful)	sl ᴘʌ ŋjo	224.0 213.1 195.8	A3 G3# G3	224.0 213.1 195.8	4 3 3	Fall-fall ↘↘	 sl ᴘʌ ŋjo
ʙul-ŋan-hɛ-ŋjo’ (anxious)	ʙul ŋan hɛ ŋjo	207.5 192.9 183.5 181.4	G3# G3 F3# F3#	207.5 192.9 183.5 181.4	3 2 2 2	Fall-fall-flat ↘↘→	 ʙul ŋan hɛ ŋjo
‘ʧm-ul-hɛ-ŋjo’ (gloomy)	ʧm ul hɛ ŋjo	230.5 213.96 178.9 170.9	A3# G3# F3 F3	230.5 213.9 178.9 170.9	4 2 2 2	Fall-fall-flat ↘↘→	 ʧm ul hɛ ŋjo

4.3. Melodic Characteristics Converted from the Linguistic Prosody across Emotional Categories

The melodic characteristics converted from the linguistic prosodic features of Korean emotion vocabulary exhibited distinct differences across the three emotional categories. Happy vocabulary showed relatively high pitch registers and wide pitch ranges. Their melodic contours were primarily descending, characterized by either conjunct or disjunct downward movements. In contrast, angry vocabulary was characterized by mid-level pitch registers and wide pitch ranges, with melodic contours displaying dynamic variations such as disjunct descending and rising-falling patterns. The melodic features of sad vocabulary were generally consistent, exhibiting low pitch registers and narrow pitch ranges. Their melodic contours were mainly conjunct descending or unison (monotone). The tempo for sad vocabulary was slower compared to those of happy and angry vocabulary, although some sad words showed relatively faster tempos. Additionally, certain sad vocabulary displayed chromatic progressions—pitch sequences in half-tone intervals—that added emotional nuance to the melodic expression (see **Table 11**).

Table 11. Melodic characteristics converted from the linguistic prosody across emotional categories.

Emotion	Pitch Register	Pitch Range	Melodic Contour	Tempo
Happy	High	Wide	Primarily descending: conjunct or disjunct downward movements	Faster than sad
Sad	Mid	Wide	Dynamic variations; disjunct descending and rising-falling patterns	Faster than sad
Angry	Low	Narrow	Mainly conjunct descending or unison (monotone), chromatic progressions intervals	Generally slower; some faster

5. Discussion

This study examined the melodic characteristics of emotion-related vocabulary by analyzing their prosodic features and converting them into musical parameters. Thirty Korean female participants (aged 19–23) who spoke standard Korean provided spoken samples of vocabulary associated with three target emotions—happiness, anger, and sadness. The prosodic features of these utterances, including fundamental frequency (Hz) and articulation time (ms), were analyzed and then converted into melodic characteristics such as mean tone, pitch register, pitch range, melodic contour, and tempo based on normalized criteria.

Statistical analyses revealed significant differences in both fundamental frequency and articulation time across the three emotional categories. These results reflect the distinct prosodic profiles of Korean emotion vocabulary and suggest perceptible acoustic cues in emotional speech expression. Happy words exhibited significantly higher average frequencies than both angry and sad words, supporting previous findings that associate positive emotional valence with higher pitch levels [9,11]. The absence of a significant frequency difference between angry and sad words suggests that pitch may be more closely related to emotional valence than to arousal, consistent with models distinguishing valence from arousal in acoustic-emotional mapping such as the circumplex model of affect [2,3,25]. Regarding articulation time, sad words showed the longest durations, while angry words were produced with the shortest. This supports the notion that sadness, associated with low arousal states, tends to be expressed through slower temporal patterns, whereas anger, associated with high arousal and urgency, is marked by faster tempos. The similarity in articulation time between happy and angry words may indicate that both emotions share a moderately elevated level of arousal.

The melodic characteristics derived from prosodic features showed clear distinctions across emotional categories in terms of mean tone, pitch range, register, contour, and tempo. The average frequency of happy vocabulary was 253.7 Hz, corresponding to B3; angry vocabulary averaged 213.1 Hz (A3), and sad vocabulary averaged 211.6 Hz (G3). When converted to melodic form, happy was characterized by high pitch registers, wide pitch ranges, and primarily conjunct or disjunct descending contours. Angry displayed mid-range registers, similarly wide pitch ranges, disjunct descending or rising-falling contours, and a generally faster tempo. In contrast, sad exhibited low pitch registers, narrow pitch ranges, unisonous or conjunct descending contours, and the slowest tempos.

The findings of this study revealed consistent patterns between vocal attributes and emotional vocabulary. These vocal features—particularly their temporal and spatial characteristics—can be interpreted through the lens of schema theory, which posits that emotions are cognitively organized and expressed through various sensory modalities, including sound. According to schema theory, individuals structure knowledge into mental frameworks, or schemata, which shape perception, memory, and interpretation of sensory input [48]. In the context of emotional language, these schemata often include spatial metaphors and prosodic expectations. For example, high-pitched, rising intonation patterns are commonly associated with positive emotional states such as excitement or happiness, whereas low-pitched, falling contours are typically linked to sadness or seriousness [11,49]. Such spatial prosodic attributes—pitch height, contour direction, and intonation slope—are cognitively internalized as part of emotion schemas, enabling listeners to rapidly interpret affective meaning through vocal cues. Thus, the spatial configuration of prosody both reflects and reinforces culturally embedded schemata for understanding emotional intent. In alignment with these frameworks, the results suggest that schematic associations are evident in vocal expression of emotion, mapping acoustic features such as pitch register or melodic contour onto spatial-emotional concepts.

Complementing schema theory, dual coding theory explains how verbal and nonverbal information—such as auditory and musical signals—are simultaneously encoded and retrieved, enhancing memory and emotional recognition [50,51]. According to this theory, cognitive processing involves two functionally independent yet interconnected representational systems: a verbal system, which processes linguistic and symbolic content, and a nonverbal system, which encodes imagery and sensory input such as auditory, visual, and musical cues. These systems can operate separately but also interact to reinforce memory and interpretation. In this study, emotional vocabulary functions as auditory signals through prosodic features like pitch, stress, and duration, which are integratively encoded alongside linguistic meanings within both systems. Stronger emotional expressions are more likely to engage dual coding processes, thereby enhancing emotional communicability and recall. By systematically mapping prosodic features onto melodic patterns, this study offers an empirical model of multimodal emotion transformation, demonstrating how emotions are conveyed across sensory modalities like language and music. For example, the transformation of happiness, anger, and sadness into distinctive pitch registers, melodic contours, and tempos concretely illustrates the integration of multimodal affective cues.

Building on this theoretical and practical foundation, the present study contributes significantly by objectively identifying emotion-specific prosodic patterns based on acoustic indices such as fundamental frequency and articulation time, and translating these into musical parameters like pitch register, melodic contour, and tempo. This quantitative and reproducible approach provides a multidimensional foundation for exploring affective expression similarities between language and music, enabling novel interpretations of emotional complexity that are difficult to capture through linguistic means alone [9,11,26,39,40].

These findings have practical applications in various fields such as affective computing, emotion-aware speech synthesis, and affective user interface design. In particular, they can serve as foundational data for musical expression that maximizes emotional impact in music therapy. However, it is important to note that the efficiency of dual coding may vary across cultural contexts and individual differences, as emotional interpretation of prosody and music can be influenced by personal auditory preferences, cultural schemata, or neurodivergent perceptual patterns. Future research should therefore examine how these factors mediate dual coding efficacy to ensure broader applicability of multimodal emotion models.

However, this study has several limitations. First, the conversion of speech prosody into musical parameters employed the 12-tone equal temperament system. While providing a structured method for mapping speech pitch onto musical notes, this framework inadequately captures microtonal pitch variations inherent in natural speech (e.g., quarter tones or finer intervals), potentially leading to loss of expressive subtlety. Moreover, melodies generated via discrete pitch mapping may sound unnatural or musically constrained. Future research should consider additional musical dimensions such as harmony, dynamics, and ornamentation to enrich emotional expression, alongside exploring alternative tuning systems or continuous pitch-mapping techniques that better reflect natural prosody's fluidity.

Second, reflecting Korean prosodic and stress characteristics posed challenges. Korean exhibits relatively minimal stress impact on lexical meaning and lacks pronounced syllabic intensity variation [52]. Consequently, attempts to translate speech intensity into musical dynamic markings failed to produce salient perceptual contrasts. More effective strategies may involve analyzing syllabic pitch patterns (intonation) relative to perceived stress rather than relying solely on intensity measures. Third, the complexity and interaction of emotions were not fully addressed. The study focused on discrete emotions—happiness, anger, and sadness—while real-life emotional experiences are often complex and dynamic. Future research should incorporate multidimensional emotional states and temporal dynamics. Finally, although the theoretical and analytical contributions are substantial, further empirical validation and technical refinement are essential for practical applications. To maximize emotional efficacy in fields such as music therapy, affective interactive systems, and creative arts, integrating diverse musical features and conducting rigorous evaluations of user experience will be critical. Future studies should also extend this work by incorporating cross-cultural and developmental comparisons and employing neuroscientific methods to deepen understanding of the integrative processing of language and music modalities in emotion perception.

6. Conclusions

The present study quantified to a deeper understanding of emotional articulation in spoken language by demonstrating that specific prosodic cues—such as pitch register, melodic contour, and articulation tempo—systematically vary according to emotional content. These prosodic features, when translated into melodic elements, offer a novel empirical basis for designing emotionally congruent melodies. In the context of songwriting and lyrical composition, this suggests that emotional authenticity can be enhanced by aligning melodic structures with the inherent prosodic profiles of emotion-related words. For instance, lyrics conveying sadness may be musically reinforced through descending contours, narrow pitch ranges, and slower tempos, while joy-themed lyrics may benefit from wider pitch intervals and brisker tempi. This integration of speech-derived prosodic features into melody design opens new avenues for data-informed songwriting, where emotion is not only lyrically described but also melodically embodied. Furthermore, this framework invites future research into cross-linguistic variation, affective congruency in music perception, and applications in music-based interventions that aim to resonate with specific emotional states.

Author Contributions

Conceptualization, S.Y.Y.; methodology, S.Y.Y. and H.J.C.; software, S.Y.Y.; validation, H.J.C.; formal analysis, S.Y.Y.; investigation, S.Y.Y. and H.J.C.; resources, S.Y.Y. and H.J.C.; data curation, S.Y.Y.; writing—original draft preparation, S.Y.Y.; writing—review and editing, S.Y.Y. and H.J.C.; visualization, S.Y.Y.; supervision, H.J.C.; project administration, S.Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data supporting the findings of this study are available from the first author upon reasonable request (mtsoo@gachon.ac.kr).

Acknowledgments

This study was based on the research design, methodology, and selection of Korean emotion-related vocabulary from Yi, Oh, & Chong (2016). The present paper is a revised and reorganized version of Yi's doctoral dissertation.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Plutchik, R., 2000. A psychoevolutionary theory of emotion. In: Plutchik, R. (ed.). Emotion in the practice of psychotherapy: Clinical implications of affect theories. American Psychological Association: Washington, DC, USA. pp. 59–79.

[2] Juslin, P.N., Västfjäll, D., 2008. Emotional responses to music: The need to consider underlying mechanisms. Behavioral and Brain Sciences. 31(5), 559–575. DOI: https://doi.org/10.1017/s0140525x08006079

[3] Eerola, T., Vuoskoski, J.K., 2011. A comparison of the discrete and dimensional models of emotion in music. Psychology of Music. 39(1), 18–49. DOI: https://doi.org/10.1177/0305735610362821

[4] Ekman, P., 1992. An argument for basic emotions. Cognition and Emotion. 6, 169–200. DOI: https://doi.org/10.1080/02699939208411068

[5] Zentner, M., Grandjean, D., Scherer, K.R., 2008. Emotions evoked by the sound of music: Characterization, classification, and measurement. Emotion. 8(4), 494–521. DOI: https://doi.org/10.1037/1528-3542.8.4.494

[6] Bänziger, T., Scherer, K.R., 2005. The role of intonation in emotional expressions. Speech Communication. 46, 252–267. DOI: https://doi.org/10.1016/j.specom.2005.02.016

[7] Kent, R., 2004. Normal aspect of articulation. In: Bernthal, J., Bankson, N. (eds.). Articulation and phonological disorders. Prentice Hall: Englewood Cliffs, NJ, USA.

[8] Scherer, K.R., Bänziger, T., 2004. Emotional expression in prosody: A review and an agenda for future research. In Proceedings of the 1st International Conference on Speech Prosody; pp. 1–14. Available from: https://sprosig.org/sp2004/PDF/Scherer-Baenziger.pdf

[9] Banse, R., Scherer, K.R., 1996. Acoustic profiles in vocal emotion expression. Journal of Personality and Social Psychology. 70, 614–636. DOI: https://doi.org/10.1037/0022-3514.70.3.614

[10] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., et al., 2001. Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine. 18, 32–80. https://ieeexplore.ieee.org/document/911197

[11] Juslin, P.N., Laukka, P., 2003. Communication of emotions in vocal expression and music performance: Different channels, same code? Psychological Bulletin. 129, 770–814. DOI: https://doi.org/10.1037/0033-2909.129.5.770

[12] Lee, S.H., 2002. K-ToBI Labelling system: A study of description system of Korean prosodic structure. The Linguistics Association of Korea Journal. 10(2), 1–18. Available from: https://scispace.com/pdf/korean-intonation-patterns-from-the-viewpoint-of-f-0-45v4hgx295.pdf

[13] Park, M.Y., Park, M.K., 2007. An analysis of emotional speech using distribution chart of f(0). Korean Linguistics. 27, 233–254. DOI: https://doi.org/10.17790/kors.2007..27.233

[14] Cho, S., Ban, T., Choi, J., et al., 2021. Study on the phonological characteristics of Korean anger speech-acts of responses. Journal of East Asian Cultures. (85), 151–174. DOI: https://doi.org/10.16959/jeachy..85.202105.151

[15] Yi, S.P., 2014. An acoustical analysis of emotional speech using close-copy stylization of intonation curve. Phonetics and Speech Science. 6(3), 131–138. DOI: https://doi.org/10.13064/ksss.2014.6.3.131

[16] Guo, X., Lai, Q., 2021. Semantic relations and prosodic features of ranhou in spontaneous Mandarin conversation. Forum for Linguistic Studies. 3(1). DOI: https://doi.org/10.18063/fls.v3i1.1250

[17] Jiang, H., 2019. Intonational pitch features of interrogatives and declaratives in Chengdu dialect. Forum for Linguistic Studies. 1(1). DOI: https://doi.org/10.18063/fls.v1i1.1082

[18] Boxill, E.H., 1985. Music therapy for the developmentally disabled. Aspen Systems Corporation: Rockville, MD, USA.

[19] Robarts, J.Z., 2003. The healing function of improvised songs in music therapy with a child survivor of early trauma and sexual abuse. In: Hadley, S. (ed.). Psychodynamic music therapy: Case studies. Barcelona Publishers: Gilsum, NH, USA.

[20] Bruscia, K.E., 1998. The dynamics of music psychotherapy. Barcelona Publishers: Gilsum, NH, USA.

[21] Baker, F.A., Wigram, T., 2005. Songwriting: Methods, techniques and clinical applications for music therapy clinicians, educators and students. Jessica Kingsley Publishers: London, UK.

[22] Gerardi, G.M., Gerken, L., 1995. The development of affective responses to modality and melodic contour. Music Perception. 12, 279–290. DOI: https://doi.org/10.2307/40286184

[23] Juslin, P.N., Sloboda, J.A., 2010. Handbook of music and emotion: Theory, research, applications. Oxford University Press: London, UK.

[24] Thompson, W.F., Robitaille, B., 1992. Can composers express emotions through music? Empirical Studies of the Arts. 10, 79–89. DOI: https://doi.org/10.2190/nbny-akdk-gw58-mtel

[25] Coutinho, E., Dibben, N., 2013. Psychoacoustic cues to emotion in speech prosody and music. Cognition and Emotion. 27(4), 658–684. DOI: https://doi.org/10

.1080/02699931.2012.732559
[26] Ilie, G., Thompson, W.F., 2006. A comparison of acoustic cues in music and speech for three dimensions of affect. Music Perception. 23(4), 319–330. DOI: https://doi.org/10.1525/mp.2006.23.4.319
[27] Rodero, E., 2011. Intonation and emotion: Influence of pitch levels and contour type on creating emotions. Journal of Voice. 25(1), 25–34. DOI: https://doi.org/10.1016/j.jvoice.2010.02.002
[28] Ross, D., Choi, J., Purves, D., 2007. Musical intervals in speech. Proceedings of the National Academy of Sciences. 104(23), 9852–9857. https://doi.org/10.1073/pnas.0703140104
[29] Friend, M., Bryant, J.B., 2000. A developmental lexical bias in the interpretation of discrepant messages. Merrill Palmer Quarterly: A Journal of Developmental Psychology. 46(2), 140–167. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5502114/
[30] Doh, S., Kim, J., Lee, H., 2023. Textless speech-to-music retrieval using emotion similarity. Preprint. arXiv:2303.10539. DOI: https://doi.org/10.48550/arXiv.2303.10539
[31] Jansen, N., Harding, E., Loerts, H., et al., 2023. The relation between musical abilities and speech prosody perception: A meta-analysis. Journal of Phonetics. 101, 101278. DOI: https://doi.org/10.1016/j.wocn.2023.101278
[32] Brown, S., 2017. A joint prosodic origin of language and music. Frontiers in Psychology. 8, 1894. DOI: https://doi.org/10.3389/fpsyg.2017.01894
[33] Kraxenberger, M., Menninghaus, W., Roth, A., et al., 2018. Prosody-based sound-emotion associations in poetry. Frontiers in Psychology. 9, 1284. DOI: https://doi.org/10.3389/fpsyg.2018.01284
[34] Larrouy-Maestri, P., Poeppel, D., Pell, M., 2024. The sound of emotional prosody: Nearly three decades of research and future directions. Perspectives on Psychological Science. 20(4), 623–638. DOI: https://doi.org/10.1177/17456916231217722
[35] Patel, D., Daniele, J.R., 2003. An empirical comparison of rhythm in language and music. Cognition. 87(1), 35–45. DOI: https://doi.org/10.1016/S0010-0277(02)00187-7
[36] Cho, T., 2021. Linguistic functions of prosody and its phonetic encoding with special reference to Korean. Japanese/Korean Linguistics. 29, 1–22. Available from: https://site.hanyang.ac.kr/documents/11105253/13205893/Cho_2022_JK-29-Chapter+1.pdf
[37] Zhao, D., Chang, W.H., 2025. Acoustic features of emotional prosody: A cross-linguistic analysis of English, Chinese, and Korean. Language and Linguistics. 107, 23–56. DOI http://dx.doi.org/10.20865/202510702
[38] Takahashi, O., Tanaka, K., Kobayashi, K., et al., 2021. Melodic emotional expression increases ease of talking to spoken dialog agents. In Proceedings of the 9th International Conference on Human-Agent Interaction (HAI '21), Kyoto, Japan, 9–11 November 2021; pp. 84–92. DOI: https://doi.org/10.1145/3472307.3484180
[39] Yi, S.Y., Oh, J.H., Chong, H.J., 2016. Tonal characteristics based on intonation pattern of the Korean emotion words. Journal of Music and Human Behavior. 13(2), 67–83. DOI: https://doi.org/10.21187/jmhb.2016.13.2.67 (in Korean)
[40] Choi, J.H., Park, J.M., 2024. Acoustic analysis and melodization of Korean intonation for language rehabilitation. Journal of Music and Human Behavior. 21(1), 49–68. DOI: https://doi.org/10.21187/JMHB.2024.21.1.049 (in Korean)
[41] Oh, E.H., 2024. Acoustic cues for emotions in Korean vocal expressions. Linguistic Research. 41(4), 611–633. DOI: https://doi.org/10.17250/khisli.41.3.202412.012 (in Korean)
[42] Lee, E.Y., Yoo, G.E., Lee, Y., 2024. Analysis of semantic attributes of Korean words for sound quality evaluation in music listening. Journal of Music and Human Behavior. 21(2), 107–134. DOI: https://doi.org/10.21187/JMHB.2024.21.2.107 (in Korean)
[43] Clopper, C.G., Smiljanic, R., 2011. Effects of gender and regional dialect on prosodic patterns in American English. Journal of Phonetics. 39(2), 237–245. DOI: https://doi.org/10.1016/j.wocn.2011.02.006 (in English)
[44] Roy, S., 2016. Future directions of modelling the uncertainty in the cognitive domain. In: Vaidya, A. (ed.). Decision Making and Modelling in Cognitive Science. Springer: New Delhi, India. (in English)
[45] Kim, J.Y., 2014. A study on direction of emotional vocabulary education for building emotional literacy. Journal of Cheong Ram Korean Language Education. 49, 321–345. DOI: https://doi.org/10.26589/jockle..49.201403.321 (in Korean)
[46] Seong, C.J., 2013. Script_toneLabler_cj.praat [computer program]. Available from: http://blog.naver.com/cj_seong (cited 1 December 2024).
[47] Lindsay, P.H., Norman, D.A., 2013. Human Information Processing: An Introduction to Psychology, 2nd ed. Academic Press: London, UK.
[48] Rumelhart, D.E., 2017. Schemata: The building blocks of cognition. In: Spiro, R.J., Bruce, B.C., Brewer, W.F. (eds.). Theoretical Issues in Reading Comprehension. Routledge: New York, NY, USA. pp. 33–58.
[49] Scherer, K.R., 2003. Vocal communication of emotion: A review of research paradigms. Speech Communication. 40(1–2), 227–256. DOI: https://doi.org/10.1016/S0167-6393(02)00084-5
[50] Scherer, K.R., Ellgring, H., 2007. Multimodal expression of emotion: Affect programs or componential appraisal patterns? Emotion. 7(1), 158–171. DOI:

https://doi.org/10.1037/1528-3542.7.1.158

[51] Klasen, M., Chen, Y.H., Mathiak, K., et al., 2014. Multimodal emotion processing in autism spectrum disorders: An integrative review. Frontiers in Human Neuroscience. 8, 897. DOI: https://doi.org/10.1016/j.dcn.2012.08.005

[52] Ko, H.J., 2012. Reconsideration of the methodology of Korean accent study. Journal of North-east Asian Cultures. 33, 257–268. DOI: https://doi.org/10.17949/jneac.1.33.201212.014 (in Korean)