The use of suprasegmentals in the production of emotion-affected speech between native English and native Korean speakers

by

Catherine Coates

Submitted to the

University Honors College in partial fulfillment

of the requirements for the degree of

Bachelor of Philosophy

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

UNIVERSITY HONORS COLLEGE

This thesis was presented

by

Catherine Coates

It was defended on

July 10, 2020

and approved by

Doctor Shelome Gooden, Associate Professor at the University of Pittsburgh, Department of Linguistics

Doctor Veronica Sardegna, Lecturer at Duquesne University, Department of Instruction and Leadership in Education

Doctor Absesalam Soudi, Lecturer at the University of Pittsburgh, Department of Linguistics

Thesis Advisor: Doctor Marta Ortega-Llebaria, Associate Professor at the University of Pitsburgh, Department of Linguistics

The use of suprasegmentals in the production of emotion-affected speech between native English and native Korean speakers

Catherine Coates

University of Pittsburgh, 2020

This study contributes to the literature investigating emotion in speech by examining the use of suprasegmental features in emotion production in native English and native Korean speakers by studying four specific research questions: (1) do native English speakers and native Korean speakers differ in their use of suprasegmentals to express emotion in their native language?; (2) do the participants differ in their use of suprasegmentals to produce emotion between their native (L1) and second language (L2)?; (3) which suprasegmentals are used to express emotion?; and (4) how are the suprasegmentals used to produce the different emotions? In order to answer these questions, 4 L1 English L2 Korean speakers and 4 L1 Korean L2 English speakers were asked to produce 50 words (25 English, 25 Korean) in five different emotional affects: happiness, sadness, neutral, fear, and disgust, creating a database of 2000 total emotional productions (50 words *5 emotions *8 participants). These emotional productions were then examined for four different suprasegmental features: duration, intensity, pitch, and voice quality.

Results indicate that there is high similarity in the use of suprasegmental features between the two native language groups, except for intensity maximum and pitch average. As for the participants' L2s, evidence of L1 transfer was found for intensity maximum and pitch average. Meanwhile, all four suprasegmental categories (i.e. duration, intensity, pitch, and voice quality) were found to have a significant effect on emotion production, though patterns among the suprasegmental categories themselves were uncovered in the process.

Table of Contents

List of Tables

List of Figures

Preface

I would like to thank first and foremost Dr. Marta Ortega-Llebaria for her invaluable support and guidance throughout this research, and who without this thesis would not have been possible. I would also like to thank the other committee members, Dr. Abdesalam Soudi, Dr. Shelome Gooden, and Dr. Veronica Sardegna for devoting their time to the review of this thesis. Finally, I would like to thank the University Honors College for generously providing the funding necessary to carry out this research.

1.0 INTRODUCTION

This study was originally inspired by a question as to what type of miscommunications may occur for a person when portraying emotional speech in a second language setting: would a native speaker mistake a non-native speaker's polite tone for one of arrogance? From where do these differences in emotion productions originate? Does a person differ in their emotion production in their L1 compared to their L2? It is these questions that lead to an examination of the current existing research on emotion, and the subsequent discovery of a gap in the literature; while previous studies have examined either emotion production itself or emotion differences between language groups, none have appeared to combine the two perspectives to determine how emotion production may differ across languages, both for speakers' L1s and their L2s. Thus, the current study is a contribution to the combining of these two perspectives, but first, an explanation of the relevant emotion concepts is in order (section 1.1), followed by a more in-depth analysis of the exact nature of the gap in literature, which leads to the research questions that this study poses in order to bridge said gap (section 1.2).

1.1 Relevant Emotion Concepts

With regards to terminology, the concepts of suprasegmentals, valence, arousal, and affect play an important role in the emotion literature. First, suprasegmentals are of particular importance to the production of emotion, as emotion in speech is mainly carried by the use of suprasegmental features (Briefer, 2012). Segmental features are those which can be analyzed in discrete units, such

as consonants and vowels, while suprasegmental features are those which continue over the discrete units and affect how those consonants and vowels sound, such as pitch, duration, intensity, and voice quality (Crystal, 1981). Moreover, the same suprasegmental features used in the expression of emotion are used as well to convey rhythm and intonation to sentences. Interestingly, English and Korean have strikingly different rhythmic and intonation patterns. In terms of rhythm, English is a stressed-timed language, while Korean is a syllable-timed language (Lee & Song, 2019). In other words, for English speakers, unstressed syllables are reduced in duration, while stressed syllables may become twice as long as their unstressed counterparts; in contrast, Korean does not have lexical stress, so each syllable is roughly the same in duration (Jeon, 2015). In terms of intonation, English is a stress-accent language and Korean is a phrase language. This means that in English, pitch-accents (or pitch turning points) are anchored in syllables with primary stress. As a result, the pitch contour of a sentence looks like a series of pitch turning points linked to the long-duration stressed syllables. In contrast, pitch-contours in Korean are defined by accentual phrases. Each accentual phrase, which contains one or more words, has the same phonological form (i.e. LH LH). Consequently, the pitch contour of a sentence looks like a regular series of peaks and valleys with tuning points aligning with the beginning and end of the words within the accentual phrase.

Second, emotional valence is defined as the extent to which an emotion causes positive or negative feelings, while arousal is defined as how intensely an emotion makes a person feel (Kuppens et al., 2013). For example, fear is a negative valence and high arousal emotion, while happiness is a positive valence and high arousal emotion. It is possible for an emotion to have several valences, such as surprise, which can make a person feel either negatively or positively (Dupuis & Pichora, 2014). Third, there is the emotional affect, which is the specific tone of voice

that a person uses to express a particular emotion. Previous studies have posited that some emotions are more universal than others, particularly those of happiness, sadness, disgust, fear, surprise, contempt, and anger (Eckman, 1999). Presumably, these emotions are the most easily recognizable for any language group, and may be easily recognizable due to their similar use of suprasegmentals.

Now, while these concepts are all important on their own, what is of particular importance for this study is how their relationships to one another affect emotion production. For instance, previous studies typically point towards the importance of four distinct suprasegmental categories in the production of emotion: duration, intensity, pitch, and voice quality. The use of these suprasegmental features can change depending on the arousal or valence of an emotional affect. For example, duration, which refers to the length of the word, relates to speech rate. Faster speech rates (i.e. shorter words) tend to reflect excitement or higher arousal (Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005). In terms of valence, negative emotions are thought to have longer durations than positive ones (Briefer, 2012). Pitch refers to how fast the vocal folds vibrate and is measured in cycles per second or Hertz. Slower or faster vocal fold vibrations are perceived respectively as low or high pitch. In general, high-pitched words may convey high arousal emotions, while low-pitched words may convey low arousal emotions (Bachorowski & Owren, 1995; Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005; Murray & Arnott, 1993). Intensity refers to how loud or quiet the word is, with a higher intensity correlating to a louder word, and a lower intensity correlating to a quieter word. Typically, high intensities are found to correspond with high arousal emotions (Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005; Sobin & Alpert, 1999). Voice quality can be defined as the manipulation of laryngeal and supralaryngeal features to color the voice of the individual; the current study specifically measured

voice quality in terms of creaky and breathy voice. Creaky voice—sometimes referred to as vocal fry—is when the larynx tightly compresses the vocal folds together, which leads to further compactness. Breathy voice is when the vocal folds allow more air to pass through while they vibrate, resulting in a whisper or sigh-like vocalization. Again, previous studies have found a correlation between high arousal emotions and an increase in both creaky and breathy voice production (Bachorowski & Owren, 1995; Briefer, 2012; Juslin & Scherer, 2005; Murray & Arnott, 1993).

In sum, the relationship between suprasegmental features, valence, arousal, and affect influence emotion production in that high arousal emotions, such as fear, become shorter, higher, louder, and breathy or creaky voiced. Less attention has been paid to the exact relationship between suprasegmentals and emotional valence, although negative emotions are thought to be longer. Nevertheless, these are the emotion concepts that one should pay attention to when examining emotion production.

1.2 Bridging the Gap in Emotion Literature

While previous studies have examined either the use of suprasegmentals in emotion production or emotion differences between language groups, no study has done a comprehensive comparison of the use of suprasegmentals in emotion production across languages, both in terms of L1 and L2. Instead, such emotion studies are either examinations of the effect of suprasegmentals in one particular language, or are cross-linguistic emotion perception-based tasks that examine how affects are interpreted differently by separate language groups, with no in-depth study on the effect of suprasegmentals.

First, the monolingual emotion production studies that examine the effect of suprasegmental features take a multitude of somewhat different, though relatively similar approaches. Banse & Scherer (1996) collected professional actors to portray 14 different emotional affects (happiness, contempt, disgust, worry, etc.) and measured their productions in terms of pitch, intensity, and duration; they helped discover that suprasegmentals index the degree of arousal and valence of an emotional affect. Sobin & Alpert (1999) again measured pitch, intensity, and duration in four different emotional affects (fear, anger, sadness, and joy) and found further evidence for the differentiation of suprasegmental use between affects, namely in terms of pause duration. Interestingly, studies that examined the effect of voice quality on emotion production typically *only* looked at voice quality. Patel et al. (2011) examined emotional speech in 10 native French speakers and found that voice quality had a significant effect on the expression of joy, relief, anger, fear, and sadness. Gobl & Chasaide (2003) examined one utterance with seven different voice qualities and found that certain voice qualities were associated with certain affective attributes (i.e. anger with tense voice, boredom with lax-creaky voice, etc.). While these studies have been fundamental in establishing the relationship between suprasegmentals and emotion production, all of them still come from a monolingual perspective. There is much to be gained from studying the use of suprasegmentals from the perspective of different language groups, especially since it has already been established that speakers of different language backgrounds do not always interpret emotional affects in the same way.

Now, the studies that find these cross-language differences in emotional interpretation typically do so from a perception task perspective, rather than that of a production task. For example, Shochi et al. (2016) examined a selection of native Japanese emotional affects in terms of how accurately native French speakers were able to interpret them; results indicated that some

emotional affects were better interpreted than others (i.e. surprise was relatively well-interpreted, but arrogance was not). Abelin & Allwood (2000) studied the emotional interpretation of L1 Swedish productions across a multitude of other L1 speakers (Swedish, English, Finnish, and Spanish) and found that certain affects had higher agreement across the different language groups than others (i.e. anger, fear, sadness, and surprise were well-interpreted, but not shyness). Pfitzinger et al. (2011) examined emotional interpretation differences between native Hebrew and native German speakers and found that they had the greatest divergence when it came to rating the valence of an emotional production (i.e. the native Hebrew speakers tended to rate an utterance as being more negative in comparison to the native German speakers). While these studies often acknowledge the important role that suprasegmental features play in perhaps being the cause of these differences in emotional interpretation, they have not conducted an in-depth study that has specifically examined the use of these suprasegmental features in different emotional affects cross-linguistically. Additionally, few, if any, studies have examined how speakers produce emotion in their L2 as opposed to their L1; therefore, it is possible that there are elements of L1 transfer (i.e. the use of L1 features in a person's L2) in the emotion production of a person's L2 that have yet to be examined.

Thus, this study attempts to bridge the gap in existing literature between monolingual emotional production tasks and cross-language emotional perception tasks by conducting research that examines the use of suprasegmental features (i.e. duration, intensity, pitch, and voice quality) in the production of five different emotional affects (i.e. happiness, sadness, disgust, fear, and neutral) for 4 L1 English / L2 Korean speakers and 4 L1 Korean / L2 English speakers. By studying four different suprasegmental features across five different emotional affects, a more nuanced description of the relationship between suprasegmentals, valence, arousal, and emotional affects

can be provided. Additionally, the cross-linguistic nature of the study allows for a more holistic view of the patterns of emotion production in terms of suprasegmental use, as well as L1 vs. L2 emotion production.

In order to fully grasp how the use of suprasegmentals differ across both the emotional affects (i.e. happiness, sadness, disgust, fear, and neutral) and the language groups (i.e. L1 English L2 Korean and L1 Korean L2 English), the following four research questions are the focus for this study:

1. Do the two native language groups (i.e. English and Korean) differ in their use of suprasegmentals to produce emotion?
2. Do the participants differ in their use of suprasegmentals to produce emotion between their L1 and L2?
3. Which suprasegmentals (i.e. duration, intensity, pitch, and voice quality) are used to express emotion?
4. How are the suprasegmentals used to produce the different emotions?

The answers to these questions will not only help to bridge the gap in the literature, but may also work as a starting point for future fields of study; examining how different language groups differ in their emotional production may lead to further inquiries in terms of what this means for second language instruction and cross-cultural communication. Thus, this database is multi-faceted in that it lends itself to numerous fields of thought.

2.0 METHODS

2.1 Emotional Affects and Word Sets

The first step in the methodological process was deciding which emotional affects the study should examine; after consideration of the universality, arousal, and valence of each affect, this study decided to focus on happiness, sadness, disgust, fear, and a neutral baseline.

First, the universality of the emotions had to be considered due to the cross-linguistic nature of the research. It was important to check and control for an equal understanding of each emotion in each language, so that the speakers were not asked to produce an emotional affect that had a slightly different interpretation in English than in Korean. As mentioned in the introduction, there is Ekman (1999) and his theory of seven universal emotions—disgust, fear, sadness, anger, contempt, happiness, and surprise. In order to ensure that each emotion's core concept would be well understood for both the Korean and English speakers, these were the emotional affects first considered for selection.

However, this study also had to consider the valence and arousal of each emotional affect. As mentioned earlier, surprise could have either a positive or negative valence, so it was automatically eliminated for ease of the production task. Additionally, some studies distinguish between a "cold anger" and a "hot anger" (Banse & Scherer, 1996; Patel et al., 2011), which potentially differ on the valence and arousal scale, so anger was also eliminated. Finally, of the remaining affects (happiness, sadness, disgust, fear, and contempt), there was an overabundance of affects with a negative valence and high arousal rating (i.e. disgust, fear, and contempt); thus, contempt was ultimately eliminated, leaving happiness, sadness, disgust, fear, and the addition of

a neutral baseline. A complete description of the valence and arousal distinctions for the five emotional affects included in this study can be found in Table 1 below.

Table 1 Valence and arousal rating of each emotional affect

Emotion	**Valence**	**Arousal**
Disgust	Negative	High
Fear	Negative	High
Happiness	Positive	High
Sadness	Negative	Low
Neutral	0	Low

After deciding which emotional affects were going to be included, the study had to decide which specific words would be included in the database. Now, it is important to note here that all words chosen to be in the database were emotion-laden for one of the five emotional affects (i.e. happy-laden, sad-laden, disgust-laden, fear-laden, or neutral-laden). Emotion-laden words are words that elicit or express a certain emotional feeling in a person, such as the word *roach* eliciting a feeling of disgust (Pavlenko, 2008). This study specifically chose emotion-laden words relating to each emotional affect, because a future research study may use this database to conduct a perception task that examines the relationship between overt emotion (i.e. emotional affect) and covert emotion (i.e. emotion-laden category). It is important to distinguish emotion-laden words from emotion words, which refer to the emotional state of a person (i.e. happy, sad, etc.), and emotion-related words, which refer to behaviors associated with each emotion (i.e. scream, punch, tears). However, Pavlenko (2008) does note that there may be some overlap between emotion-laden and emotion-related words; for instance, the word *tears* may be emotion-related, but still elicit a feeling of sadness within a person.

Now, the first step in selecting emotion-laden words for each emotional affect was to ensure that each potential word did in fact elicit that particular emotion within a person. In order to accomplish this, there were two surveys handed out, one in English and one in Korean. Each survey consisted of 140 potentially emotion-laden words, with approximately 28 words per emotional affect. Participants completed the survey in their L1.

The participants were instructed to select with which emotions they associated each word out of the following six options: happiness, sadness, disgust, fear, neutral, and other. If they selected other, they were asked to specify, although this option was never chosen. Any word that did not have a majority emotion category selected (i.e. above 50% selection rate) was immediately eliminated from consideration. The rest of the results of the emotion surveys were considered in conjunction with the frequency and length of each word.

In order to ensure that each chosen word was equally recognizable for each language group, this study decided to take into account the frequency of each potential word and create relatively similar numbers in both frequency average and standard deviation, while still keeping in mind word length and the results of the surveys. Unfortunately, one database that had information regarding word frequency for both English and Korean could not be found, so the database clearPOND was used for English word frequency, while the database Kokoma was used for Korean word frequency. Frequency itself was measured in occurrences per every million words. In terms of word length, any word that exceeded four syllables was excluded, keeping in mind ease of production for the participants. Words of five syllables or more were also excluded in order to control for the duration suprasegmental factor; while not a focus of this study, future studies using this database may want to examine the use of suprasegmental features across emotion-laden

categories, rather than emotional affects (i.e. do sadness-laden words have the longest duration in comparison to the other categories, no matter the emotional affect produced?).

Final decisions were made through careful trial and error with consideration regarding the strength of the association each word had to the emotion, its effect on the emotion category's average frequency, and its length of four syllables or less. For example, the English word *friend* had a 100% association rate with happiness, but its frequency rate of 419.29 was considerably too high in comparison to other words, such that it skewed the average frequency in comparison to the other emotion categories.

The words with the highest majority rule that did not skew the average were chosen as the final word lists. A table detailing the exact average frequency and standard deviation for each languages' emotion categories can be found in Table 1. The average frequency for all the emotion categories, excluding disgust, varied from 21.94-23.71 with a range of 1.77. The standard deviation for these groups varied from 6.46-8.00 with a range of 1.54. Disgust was analyzed separately from the other emotion categories, because the researcher found that the frequency for disgust-laden words was considerably lower than those of other categories. This is presumed to be due to the aversion many people have towards disgust-laden words, and their avoidance of use in public (Pavlenko, 2008). As a result, the average frequency for the English disgust words was 5.42 and the average frequency for the Korean disgust words was 5.25, leaving a range of 0.17. The standard deviation for the English disgust words was 6.63, and the standard deviation for the Korean disgust words was 7.97, leaving a range of 1.34.

In the end, five words for each emotion category for each language were selected, leading to a total of twenty-five words in both English and Korean and fifty words overall. A complete list of the selected words can be found in Appendix A.

Table 2 Frequency and standard deviation of each emotion-laden word set for both language groups

	Frequency (per million of words)	Standard Deviation
Happiness		
English	22.37	6.46
Korean	22.60	8.00
Sadness		
English	21.94	7.37
Korean	23.71	6.90
Disgust		
English	5.42	1.97
Korean	5.25	1.89
Fear		
English	22.18	6.63
Korean	22.87	7.97
Neutral		
English	22.93	7.57
Korean	23.52	7.54

2.2 Participants

There were eight total participants, four native speakers of English who were learners of Korean and four native Korean speakers who were learners of English. All four L1 English speakers were female students with an age range of 18-21 and an average age of 18.75. The L1

Korean participants consisted of three females and one male, with an age range of 28-61 and an average age of 46.75. Two of the L1 English participants were heritage speakers of other languages, one being Russian and the other being Tagalog. None of the L1 Korean participants were heritage speakers of another language.

In terms of proficiency in their L2, all L1 English participants at the time of the recordings had completed two semesters worth of Korean classroom language instruction at the University of Pittsburgh and considered themselves beginner speakers of Korean. One of the L1 Korean participants was a foreign language professor at the university, and considered a high intermediate speaker of English, while another was a graduate student and considered an advanced speaker of English. The remaining two L1 Korean participants did not permanently live in the U.S. at the time of recordings, but had extensive English classes during their schooling, and were considered beginner speakers of English.

2.3 Recording

Recordings were conducted in the month of July, 2019, with the exception of one held in September of the same year. All were recorded in the University of Pittsburgh's linguistics lab in a sound-proof booth with a head-device microphone through the use of the program Pratt.

Participants were first instructed that they would have to read through the list of fifty words five separate times, each time changing the affect of their voice to fit one of the emotion-laden categories. Participants could begin with the emotion category of their choosing, but each emotion category recording began with the emotion-laden words belonging to that category. Before each recording, participants were also made to read aloud a short emotional scenario (see Appendix B)

corresponding with the particular emotion category in order to prepare them for the emotional affect that they were being asked to portray. These prompts were always read aloud in the participant's native language, and the participant could refer back to them at any point if they wished to do so.

Each participant had a total of 250 analyzed sound files, which led to 2,000 overall sound files split evenly between English and Korean. During the recordings, the researcher would sometimes ask a participant to go back and repeat particular words if the participant had fallen out of the particular emotional affect that was requested. These segments, while recorded, were not included in the final analysis.

2.4 Coding

All sound files were analyzed using Pratt. For each sound file, one tier was created in order to mark each syllable. A boundary was placed at the end of each syllable offset, with the same boundary also marking the onset of the next syllable. For syllable coding, simple numerical order was used. An example can be found in Figure 1 of Participant #7 (a native English speaker) producing the English word *apology* with a sad affect.

In terms of the coding of the sound files themselves, a six point coding system was used to mark the following factors: (1) which participant it was, (2) to which language the word they were speaking belonged, (3) what their native language was, (4) what emotional affect their voice portrayed, (5) to which emotion-laden category the word belonged, (6) to which set number the word belonged in that particular emotion-laden category. For example, the sound file displayed in Figure 1 above is coded as 7EESS2. This is because it was produced by Participant #7, the

participant's native language was English, the word itself was produced in English, the emotional affect was sadness, the word itself belonged to the sadness-laden category, and the word *apology* is the second set out of the five sadness-laden English words.

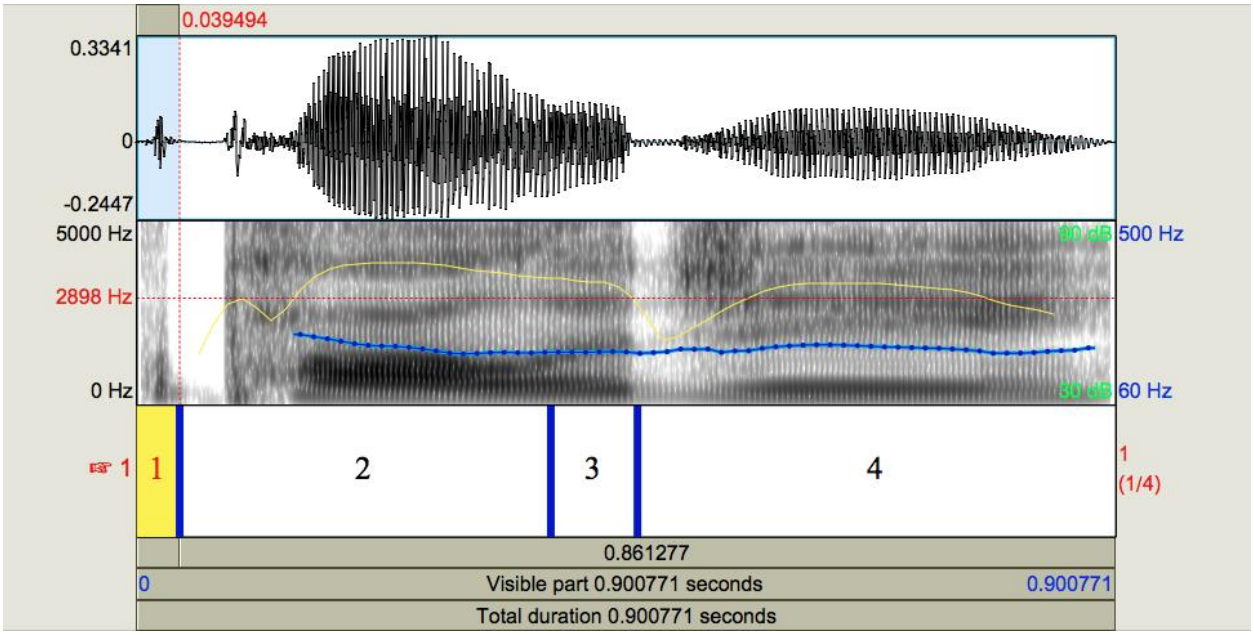

Figure 1 Participant #7 producing the sadness-laden word *apology* with a sad affect

2.5 Scripts & Measurements

A script was used to analyze the four different suprasegmental categories for each of the 2000 words. This study examined fourteen different acoustic measurements in total, a comprehensive list of which can be found in Table 2. All fourteen features related to either the duration, pitch (minimum, maximum, range), intensity (minimum, maximum, range), or voice quality (creaky and breathy voice) of the words in order to capture their suprasegmental characteristics.

In order to control for the male voice among the native Korean speakers, the script was adjusted to add 30 Hertz to each of the male participant's pitch points. Previous studies have estimated the average pitch range of a female voice to be 165 to 255 Hertz, while the average male voice typically ranges from 85 to 155 Hertz (Cheng et al., 2016; Hancock & Rubin, 2014; Jacobi & Schweers, 2017; Mendoza et al., 1996; Pemberton et al., 1998). A cursory analysis of the male participant's pitch to that of his female counterparts show the pitch differences to be relatively minor; the addition of any value over 30 Hertz could have potentially made the male participant's pitch too high in comparison. Thus, 30 Hertz was chosen as the adjustment value.

Table 3 Suprasegmental features and their definitions

Suprasegmental Feature	**Definition**
Duration	The length of the segment in milliseconds.
F_0Q1	The pitch of the first quartile of the segment.
F_0Q2	The pitch of the second quartile of the segment.
F_0Q3	The pitch of the third quartile of the segment.
Intensity Minimum	The lowest amplitude value of the segment.
Intensity Maximum	The highest amplitude value of the segment.
Intensity Average	The mean amplitude value of the segment.
F_0 Minimum	The lowest pitch value of the segment.
F_0 Maximum	The highest pitch value of the segment.
F_0 Average	The mean pitch value of the segment.

F0 Minimum / F0 Maximum:	The ratio of the pitch range of the segment.
Intensity Minimum / Intensity Maximum:	The ratio of the intensity range of the segment.
Jitter	The measurement of creaky voice in the segment, with a value between 0-2 or 0%-200%.
Shimmer	The measurement of breathy voice in the segment.

3.0 RESULTS

3.1 General Results

A univariate and a multivariate repeated measures ANOVAs with participants' L1 (English or Korean) as the within-subjects factor and spoken emotion (disgust, happiness, fear, neutral, or sadness) and participants' L2 (English or Korean) as the between-subjects factors were performed on the duration, intensity (minimum, maximum, average, range), pitch (minimum, maximum, average, range), and voice quality (creaky voice and breathy voice) measurements extracted from the 2000 recorded words (50 words * 5 emotions * 8 participants (4 English L1-Korean L2; 4 Korean L1-English L2). The univariate ANOVA showed statistically significant results for emotion ($p=0.000$), indicating that the participants do indeed use suprasegmentals to express emotion. The univariate ANOVA also showed statistically significant results of emotion and its interaction with L1 (emotion * L1) at $p=0.009$, indicating that the participants' use of some of these acoustic cues was modulated by their L1. In contrast, L2 and its interactions never reached statistical significance.

The multivariate ANOVA detailed the above significant factors for each one of the acoustic measurements. For the emotion factor, significance at $p < 0.01$ was obtained for the duration, intensity (minimum, maximum, average), and pitch (minimum, maximum, average) measurements. Marginally significant differences were obtained for the voice quality measurements and the pitch range and intensity range measurements. The interaction emotion * L1 was significant for only two measurements, namely pitch average and intensity maximum. Consistent with the univariate ANOVA, results from the multivariate ANOVA showed no

statistically significant differences for L2 and their interactions with any of the measurements, revealing that in whatever manner participants used suprasegmentals to express emotions in their L1, they used them in the same way in their L2. For example, native English speakers produced sad words with similarly longer durations in both their L1 English and L2 Korean.

It is important to know the details of these significant differences for the emotion factor in each measurement, and for the emotion * L1 interaction for the pitch average and intensity maximum measurements. To that end, the relevant post-hoc analysis and graphs are portrayed in the sections below.

3.2 Details of the Emotion differences in each acoustic measurement

3.2.1 Duration

There was a significant effect of word duration on emotion ($F(4,24)=5.237$, $p=0.004$) showing that participants used differences in duration to express different emotions. As seen in Figure 1 below, the sad affect words tended to be consistently the longest in duration (M=715.5 ms, SD=101), neutral affect words were consistently the shortest in duration (M=572.6 ms, SD= 82), with disgust, fear, and happiness following in between (M=657.1, SD=79; M=690.6, SD=136, M=626.3, SD=69). Fear affect words had more variability in terms of their range. Post-hoc analyses with the Bonferroni adjustment showed a significant difference between the production of neutral and fear ($p=0.010$) and neutral and sadness ($p=0.001$), but not between any of the other emotional affect relations.

In general terms, there seems to be a relation between speech rate and emotional valence. Sadness and fear, both of which are negative valence emotions, are spoken with significantly slower speech rates than words with a neutral emotion. This slow rate seems especially relevant in the emotion of sadness, since its variation is much smaller than that of fear. Emotions with a non-negative valence (i.e. neutral and happiness) were spoken with comparatively faster speech rates. It is also worthy to note that, while not reaching significance, disgust—the third negative valence emotion—still had consistently longer durations than happiness. Thus, participants used

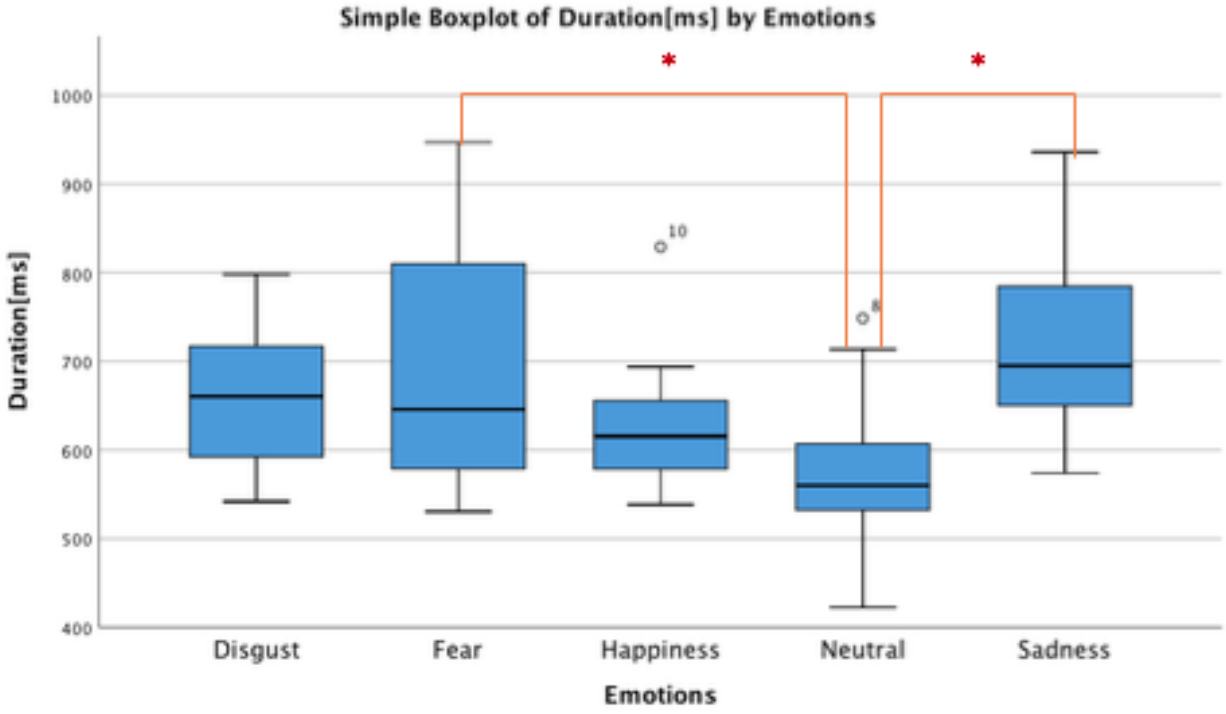

Figure 2 Duration of words per emotional affect. Duration is measured in milliseconds.

a slower speech rate to express emotions with negative valence with a particular slow rate for sadness.

3.2.2 Intensity

The multivariate ANOVAs showed a significant effect for each intensity on emotion at $p < 0.05$ showing that the 8 participants used intensity to express emotion (minimum: $F(4,24)=6.564$, $p=0.001$; maximum: $F(4,24)=16.553$, $p=0.000$; average: $F(4,24)=15.735$, $p=0.000$; range: $F(4,24)=2.990$, $p=0.029$). In contrast, the interaction Emotion*L1 was significant only for the intensity maximum measurement ($F(4,24)=2.980$, $p=0.039$), showing that the 4 native Korean and the 4 native English participants had differences in the maximum intensity they used. Consequently, the graphs in Figure 3 that depict intensity minimum (3a), intensity average (3c), and range (3d) pull together the 8 participants, while L1 English participants and L1 Korean participants are portrayed separately from one another in 2b, which depicts intensity maximum.

Despite that post-hoc analyses with the Bonferroni adjustment found no significant differences between the emotional affects for either intensity minimum, intensity average, or intensity range, general tendencies show a trend of fear affect and happiness affect words consistently producing both the highest minimums (Fear: M=36.5, SD=6.3; Happiness: M=36.2, SD=6.1; Neutral: 34.4, S.D=5.1; Disgust: M=33.7, SD=6.7; Sadness: M=31.2, SD=5.3) and highest averages (Fear: M=63.2, SD=8.7; Happiness: M=63.9, SD=8.2; Neutral: M=59.4, SD=6.9; Disgust: M=60.4, SD=7.7; Sadness: M=56.5, SD=6.5). In the same vein, the sad affect words consistently produced the lowest minimums and lowest averages. In terms of intensity range, patterns are more difficult to discern, although sadness is the only emotional affect with a mean that falls below the 0.5mark. This indicates that the range between their intensity minimum and intensity maximum tended to be larger in comparison to the other affects. In other words, the

sad affect words were slightly more likely to vary in their volume patterns for each production. A visual inspection of the spectrograms showed that words with a sad affect started relatively loud, but then became quieter, while the other affects stayed comparatively consistent in their volume points throughout a word.

Now, in terms of intensity maximum, post-hoc analyses with the Bonferroni adjustment showed statistically significant differences between the L1 English (M= 72.04, SD=6.5)and L1 Korean (M=61.4, SD=5.5) group (p=.039). For all five affects, the L1 English participants produced statistically significantly higher values for their intensity maximums in comparison to the L1 Korean participants, which means that their maximum volume points were louder than those of the L1 Korean speakers.

For the native English speakers, general trends showed that fear and happiness were typically the affects with the highest intensity maximum points in comparison to the other three affects. The sad affect words were the lowest (i.e. quieter, M=67.1, SD=5.9) with their intensity maximums in comparison to both the fear and happy affect words (M=76.5, SD=4.6, M=76.3, SD=5.9). Post-hoc analyses with the Bonferroni adjustment found these differences to reach statistical significance (i.e. sadness vs. fear at p=0.044 and sadness vs. happiness at p=0.039). While not reaching significance, disgust (M=71.3, SD=6.2) and neutral (M= 69, SD=5.6) also had higher intensity maximums than sadness.

For the native Korean speakers, post-hoc analyses with the Bonferroni adjustment found no statistically significant differences between the affects. General trends show that fear and happiness had the highest median values (i.e. their productions were more consistently loud, M=62.4, SD=5.5; M=63.02, SD=4.2), although disgust (M=62.2, SD=7) and neutral (M=61.1, SD=6.2) were variable enough with their productions that a trend is not as easily discernable as it

is with the L1 English productions. Still, sadness was consistent in producing the lowest intensity maximum values in comparison to the other four affects.

In sum, general trends suggest a connection between intensity and emotional arousal. Fear and happiness, both high arousal emotions, were consistently the loudest in their intensity minimum, maximum, and average productions. In contrast, sadness, a low arousal emotion, was consistently the quietest in its minimum, maximum, and average productions. There is a distinction between the two language groups in that the native Korean speakers did not get as loud as the native English speakers with their maximums, suggesting that they were more restricted with their intensity threshold and were quieter in their productions than the L1 English speakers.

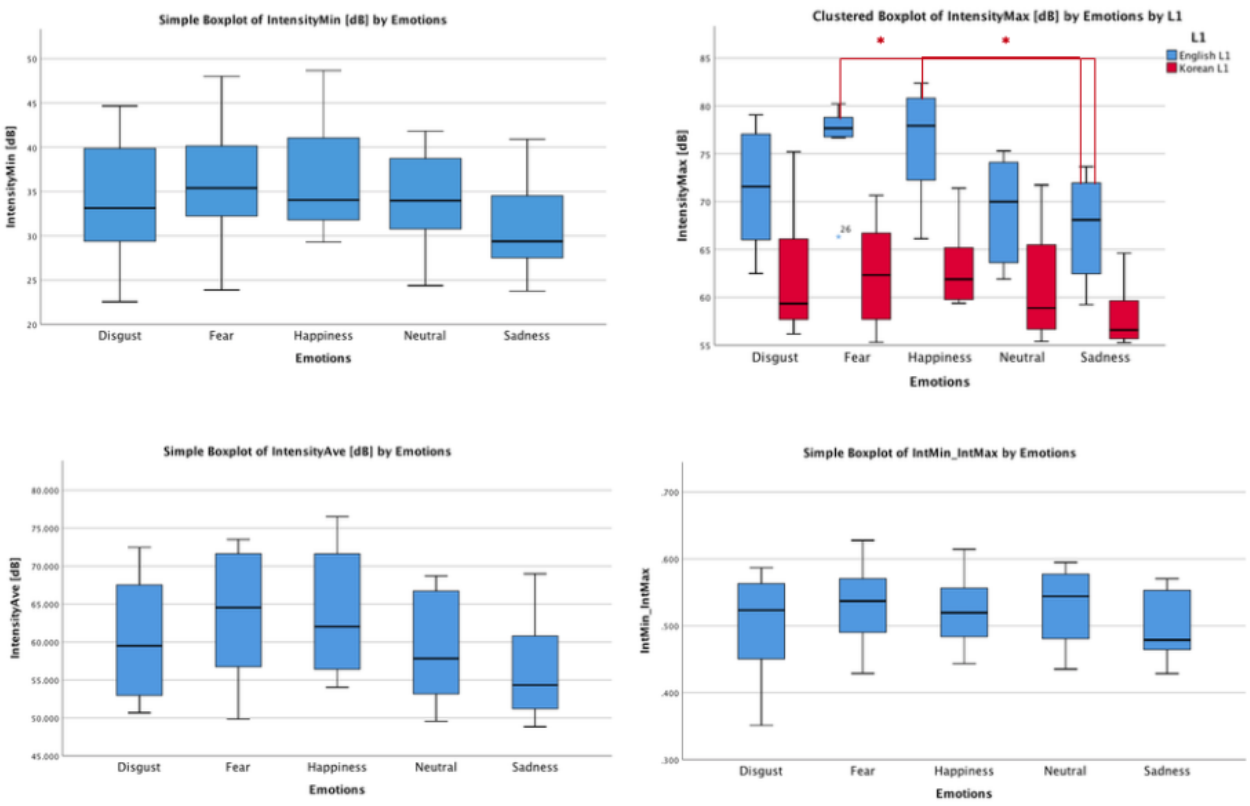

Figure 3 Intensity of words per emotional affect. Intensity is measured in decibels. 3a (top left) depicts intensity minimum, 3b (top right) intensity maximum, 3c (bottom left) intensity average, 3d (bottom right) intensity range.

3.2.3 Pitch

F0Q1, F0Q2, and F0Q3 were not included in the analysis, since Pratt provided too many statistical errors to come to any accurate results. Still, the Univariate Repeated Measures ANOVAs showed significant differences on emotion for each remaining pitch measurement (minimum: $F(4,24) = 5.117$, $p=0.004$; maximum: $F(4,24) = 6.840$, $p=0.001$; average: $F(4,24) = 12.784$, $p=0.000$; range: $F(4,24) = 2.990$, $p=0.039$) showing that differences in pitch conveyed different emotions. With regards to the interaction Emotion*L1, all pitch measurements except for pitch average obtained non-significant differences showing that, in general, native English and native Korean speakers made a similar use of pitch to convey emotion. Consequently, in Figure 4 below, graphs portraying pitch minimum (4a), pitch maximum (4b), and pitch range (4d) collapsed the 8 participants, while in (4c), pitch average was depicted for the 4 native English separately from the 4 native Korean speakers.

As depicted in Figure 4a, fear words obtained the highest pitch minimum values (fear: M=178.94, SD=44.0; happiness: M=163.30, SD=36.97; neutral: M=145.28, SD=31.84; disgust: 133.69, SD=20.98; sadness: M=142.26, SD=23.50). Post-hoc analyses with the Bonferroni adjustment showed statistically significant differences between fear and disgust ($p=0.002$) and fear and sadness ($p=0.028$). While not reaching significance, happiness and neutral follow the same general trend in that they were consistently lower in their minimum pitch productions than fear.

Figure 4b depicts pitch maximum values for each emotion. In general, happiness (M=350.75, SD=99.13) and fear (M=324.42, SD=96.15) had higher means than neutral (M=260.60, SD=33.27), sadness (M=255.59, SD=35.15), and disgust (M=274.47, SD=48.08). Post-hoc analyses corroborated statistically significant differences for happiness with disgust ($p=0.024$), with neutral ($p=0.004$), and with sadness ($p=0.003$).

Pitch range is depicted in Figure 4d. Post-hoc analyses found no statistically significant differences. Still, it is worth noting that the general trend indicates that the happy affect words had consistently lower range values than the other four affects. This means that the differences between the intensity minimum values and intensity maximum values were larger, indicating that each happy affect word saw more variability in pitch in comparison to the other four affects, which were more likely to stay consistent.

Post-hoc analyses found statistically significant differences between the two language groups for pitch average ($p=0.005$). The native English speakers produced significantly higher pitch averages ($M=245.08$ Hz, $SD= 54.8$) than the native Korean speakers ($M= 187.1$, $SD=17.1$) for the five emotional affects. For native English speakers, statistically significant differences were found between the affects for fear and disgust ($p=0.000$), fear and neutral ($p=0.000$), fear and sadness ($p=0.000$), happiness and disgust ($p=0.000$), happiness and neutral ($p=0.000$), and happiness and sadness ($p=0.000$). Essentially, fear and happiness ($M=303$ Hz, $SD=26$, $M=304$ Hz, $SD= 49$) produced significantly higher pitch averages than the other three affects (disgust $M=210$, $SD=18$, neutral $M=209$, $SD=25$, sadness $M=207$, $SD=11$. General trends show that fear was more consistent in these high value productions, while happiness had more variability.

With regards to pitch average (Figure 4c), the native English speakers generally produced higher pitch averages than the native Korean speakers for the five emotional affects. However, the general pattern was similar between the 4 native English speakers and the 4 native Korean participants in that, for all participants, fear (English: $M=303$ $SD=26$; Korean: $M=196$ $SD=12$) and happiness (English: $M=304$, $SD=49$; Korean: $M=203$, $SD=23$) had higher averages than neutral (English: $M=209$, $SD=25$; Korean: $M=182$, $SD=7$), disgust (English: $M=210$, $SD=18$; Korean: $M=180$, $SD=7$), and sadness (English: $M=207$, $SD=11$; Korean: $M=171$, $SD=8$).

Moreover, happiness had a larger standard deviation than fear. For native English speakers, these differences reached statistical significance for fear and disgust (p=0.000), fear and neutral (p=0.000), fear and sadness (p=0.000), happiness and disgust (p=0.000), happiness and neutral (p=0.000), and happiness and sadness (p=0.000). For the native Korean speakers, post-hoc analyses found statistically significant differences between fear and sadness (p=0.007), happiness and disgust (p=0.022), happiness and neutral (p=0.033), and happiness and sadness (p=0.000).

In general, there seems to be a relationship between pitch and emotional arousal, at least in terms of fear and happiness; fear typically had the highest pitch in terms of the minimum, while happiness typically had the highest pitch in terms of the maximum. This suggests a further difference in pitch pattern between the two; each fear affect word was more likely to stay consistently high in pitch throughout the segment, while the happy affect words were more likely

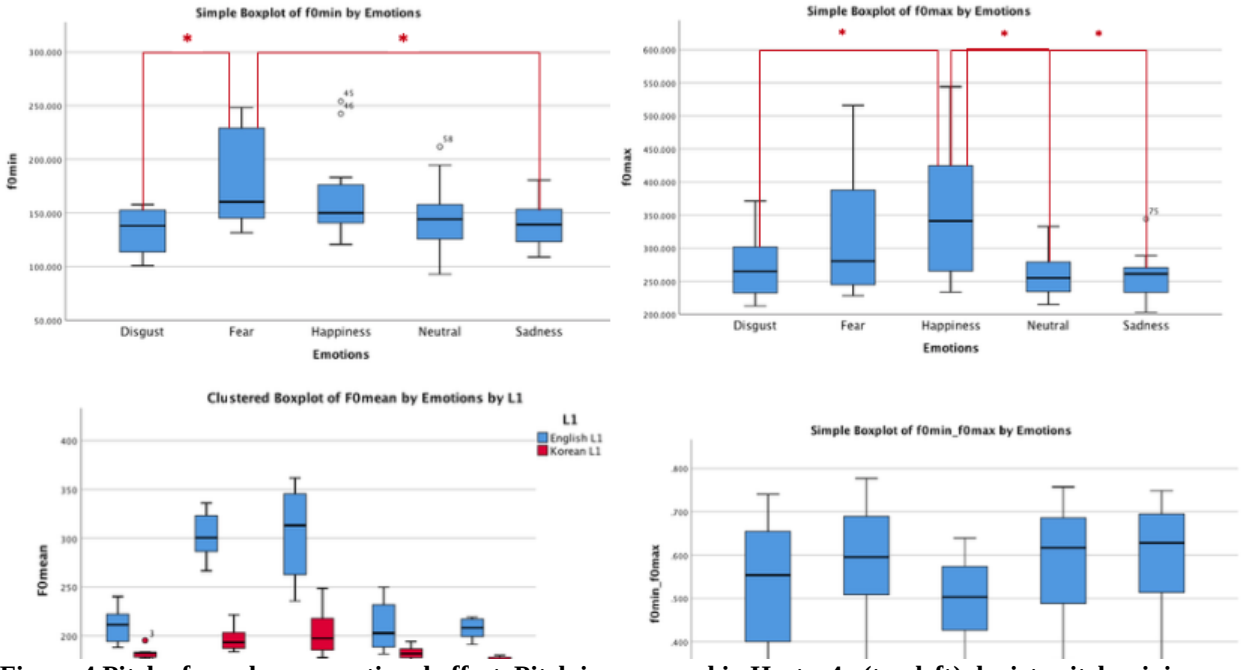

Figure 4 Pitch of words per emotional affect. Pitch is measured in Hertz. 4a (top left) depicts pitch minimum, 4b (top right) pitch maximum, 4c (bottom left) pitch average, 4d (bottom right) pitch range.

to produce a variable pattern throughout the word, such as down-up or up-down. This is further supported by both the fact that happiness had more variability in its pitch averages and that it had consistently lower ratio values than the other affects. In terms of the two language groups, the native English speakers were once again slightly more pronounced in their differences between the affects, while the native Korean speakers were overall deeper in their productions.

3.2.4 Voice Quality

The univariate ANOVAs showed significant differences on emotion for each voice quality measurement (jitter: $F(4,24)=3.153$, $p=0.032$; shimmer: $F(4,24)=3.085$, $p=0.035$), showing that differences in voice quality conveyed different emotions. No significant differences between language groups were found in terms of voice quality. The nonsignificant scores for the Emotion*L1 interaction showed that there were no differences between native English speakers and native Korean speakers in terms of voice quality.

As shown in the jitter values depicted in Figure 5a, disgust is the most consistent in producing high amounts of creaky voice than happiness, neutral, sadness, and fear. Similarly, fear is the most consistent in producing the lowest amounts of creaky voice than happiness, neutral, sadness, and disgust. Post-hoc analyses with the Bonferroni adjustment confirmed the significance between disgust and fear ($p=0.022$).

Figure 4b depicts shimmer values for each emotion. Although post-hoc tests found no significant differences, general trends show that disgust and sadness typically produce the highest amounts of breathy voice, though disgust is more consistent with these productions. Fear produces

the lowest amount of breathy voice the most consistently out of all the affects. Happiness and neutral generally fall between these two points on the scale.

In sum, there is not as clear of a relationship between emotional valence or arousal with voice quality as there is for the other suprasegmental categories. Disgust, a negative valence and high arousal emotion, was the affect most likely to use voice quality adjusters such as creaky and breathy voice. In contrast, fear, another negative valence and high arousal emotion, was the least likely affect to use either of these voice qualities. Meanwhile, it is worth noting that sadness, a negative valence and low arousal emotion, also saw a comparatively higher use of breathy voice.

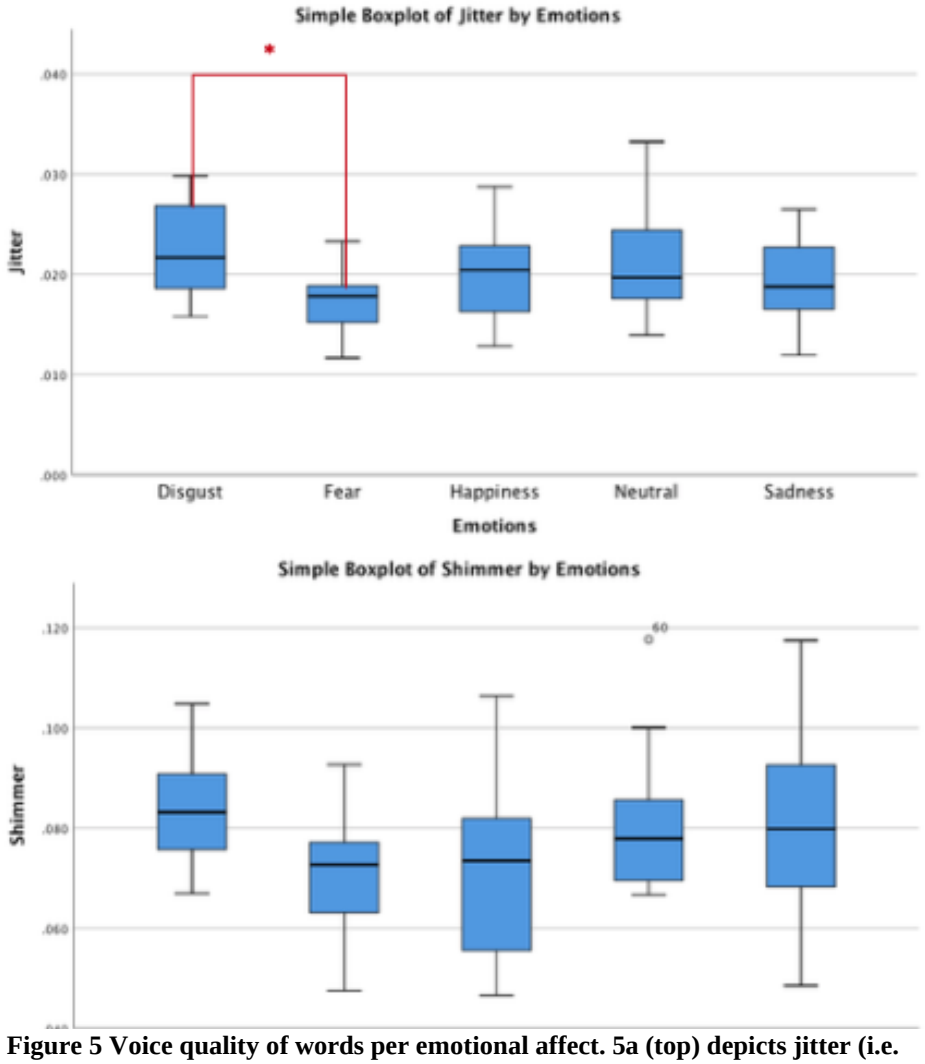

Figure 5 Voice quality of words per emotional affect. 5a (top) depicts jitter (i.e. creaky voice) and 5b (bottom) depicts shimmer (i.e. breathy voice)

4.0 DISCUSSION

4.1 RQ1: Do the two native language groups differ in their use of suprasegmentals to produce emotion?

The first research question this study posed was if native English speakers and native Korean speakers used suprasegmentals in different ways to produce emotion. After running univariate and multivariate ANOVAs, only two of the eleven suprasegmental features examined were statistically different between the two language groups, namely intensity maximum and pitch average. The native Korean speakers were significantly quieter than the native English speakers in terms of the intensity maximum, and significantly lower in pitch (i.e. deeper) in terms of the pitch average. Nevertheless, while these differences were significant, the overall pattern between the emotional affects was still the same for both native language groups, albeit slightly more pronounced for the native English speakers. For example, the emotional affect with the lowest intensity maximums was sadness for both native English and native Korean speakers.

In general, the two native language groups were remarkably similar in their use of suprasegmentals to express emotion, with the remaining nine of the eleven features measured not having any significant differences between the two groups. This similarity may stem back to Eckman (1999) and his idea of the universal or "basic" emotions that can be found in all societies. As mentioned earlier, all the emotions examined in this study come from this universal emotion list (with perhaps the exclusion of neutral). Since they are universal emotions, it may be that the manner in which to use suprasegmentals to express them may also be relatively universal. This is further supported by cross-linguistic emotion perception studies, such as Shochi (2016), which

examined identification rates of Japanese emotional affects for native French speakers that were learning Japanese; results showed that the L2 speakers had the most difficulty with identifying more nuanced emotions, such as arrogance. This trend of nuanced emotions often being the ones to be the most difficult to identify for non-native speakers of a language may indicate a greater difference in the use of suprasegmentals to express them across linguistic boundaries. Therefore, future studies may find greater significant differences in the production of emotional affects such as arrogance, obviousness, doubt, etc. Meanwhile, this study added further support to the universal emotions theory by supplying results that indicate highly similar use of suprasegmentals in the expression of happiness, sadness, neutral, fear, and disgust between two typologically unrelated languages.

4.2 RQ2: Do the participants differ in their use of suprasegmentals to produce emotion in their L1 & L2?

The second research question posed by this study was whether or not the participants would differ in their use of the suprasegmentals from their L1 to their L2. As the multivariate ANOVAs showed, the interaction between emotion and the participants' spoken language (i.e. either their L1 or L2) was statistically non-significant, meaning that the participants used the suprasegmentals to express the emotions in the same way, no matter which language they were speaking. This is not surprising when taking into account the high similarity between the two language groups in the use of the suprasegmentals, as discussed in the subsection above. Still, this means that, even for the suprasegmental factors that did reach significance between the two native groups—intensity maximum and pitch average—the interaction remained statistically non-significant.

This lack of interaction between the speakers' L1 and L2 suggests a level of L1 transfer, which has been highly reported in a number of cross-linguistic suprasegmental studies (Diaz, 2017; Li, 2011; Meng, 2009; Seddighi, 2010; etc.), though not necessarily in terms of emotion. As previous studies have shown, the amount of L1 transfer typically decreases as level of proficiency increases (Bu, 2012; Slabakova, 2000; Yamashita, 2002; etc.) The native English speakers of this study were all entering their second year as Korean students and rated themselves as capable of maintaining beginning-intermediate conversations; two of the native Korean speakers were either a doctoral student or a professor at an American university, and thus rated themselves as highly proficient in English, while the other two native Korean speakers spent their entire schooling up to the age of 18 learning English in a classroom setting and rated themselves as intermediate. Though level of proficiency was not a factor examined in this study, the overall L2 proficiency levels were presumably not high enough to negate or lessen the effect of L1 transfer. Alternatively, the use of suprasegmentals such as intensity maximum and pitch average may be so subtle as to still go unnoticed by even those of a relatively high proficiency. Nevertheless, the factor of spoken language was not a significant one for this study.

4.3 RQ3: Which suprasegmentals are used to express emotion?

The third research question posed by this study was which of the eleven suprasegmentals in this study—all either related to duration, intensity, pitch, or voice quality—would be significantly used in the expression of emotion. As the results showed, all eleven suprasegmentals ended up being used to express the five emotional affects. These results are in line with previous studies who have provided significant effects of these suprasegmentals in emotion expression (for

an overview, see: Briefer, 2012). However, there was a notable difference in the p-values for these suprasegmentals; all duration, intensity, and pitch measurements (with the exception of intensity range and pitch range) produced $p < 0.01$, while the voice quality p-values were $p=0.032$ for jitter and $p=0.035$ for shimmer. While the different p-values do not indicate a possible hierarchy by themselves, this result invites a further analysis in order to examine whether there is in fact a possible hierarchy within the suprasegmentals in terms of their use for the expression of emotion, with duration, intensity, and pitch having a slightly greater importance than that of voice quality. This is a revealed benefit of examining all these suprasegmentals under one study.

4.4 RQ4: How are the suprasegmentals used to produce the different emotions?

The fourth and final research question examined by this study was how the suprasegmentals differed in their use across the five emotional affects. A complete description of the use of each suprasegmental feature for each emotional affect can be found in Table 4, juxtaposed against their emotional value ratings.

For duration, the negative valence emotions (i.e. fear, sadness, and disgust) were typically longer than the positive valence emotion (i.e. happiness) and the neutral affect, sadness being particularly long. This is in line with previous findings that show negative valence reactions tend to last longer in duration than those of a positive valence (Briefer, 2012). The particularly long duration of sadness can possibly be explained by the fact that, in addition to being negative valence, it is also low arousal; low arousal emotions have been shown to be longer in duration than high arousal emotions (Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005). Meanwhile, the

variability in fear's durations suggests that it is being affected by both its status as a negative valence emotion (i.e. longer durations) and as a high arousal emotion (i.e. shorter durations).

In terms of intensity, the three high arousal emotions (i.e. fear, disgust, and happiness), were consistently louder than the low arousal emotion (i.e. sadness) and the neutral affect, especially fear and happiness. These findings are in agreement with previous studies that have shown high arousal emotions to be louder than their low arousal counterparts (Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005; Sobin, 1999).

For pitch, two of the high arousal emotions, namely fear and happiness, were consistently among the highest-pitched affect, although there is a slight difference in that the results revealed a pattern that happiness saw more variability in pitch within a word, while fear stayed consistently high. This is in line with previous studies that state high arousal emotions are higher pitched than low arousal emotions (Bachorowski & Owren, 1995; Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005; Murray & Arnott, 1993). Interestingly, the third high arousal emotion, disgust, was among the most consistently low-pitched affects, along with sadness. While these low pitches were expected for sadness, as a low arousal emotion, the disgust pitch results go against those of previous studies; this may be in part due to the differences in the exact levels of arousal and valence for disgust. While both disgust and fear are negative valence / high arousal emotions, disgust has a higher rating of negative valence than that of fear (Kollias et al., 2019); although valence effects on pitch are not as abundant as those found for arousal, it is possible that disgust's particularly high rating of negative valence is the reason for its low pitches, as shown in other studies that have examined human vocalizations (Briefer, 2012). Additionally, disgust has a comparatively low rating of arousal than fear (Kollias et al., 2019); thus, disgust may not be as constrained by the arousal factor as fear, leading disgust to have lower pitches.

In terms of voice quality, disgust, a negative valence / high arousal emotion, was the most likely to make use of both creaky and breathy voice, which is in line with previous studies that show an increase in non-modal phonation for higher arousal emotions (Bachorowski & Owren, 1995; Banse & Scherer, 1996; Briefer, 2012; Juslin & Scherer, 2005; Murray & Arnott, 1993). However, this stands in direct contrast to the findings for the other negative valence / high arousal emotion (i.e. fear), which was the least likely to use breathy or creaky voice. Indeed, previous studies that have examined voice quality did not explicitly distinguish between disgust and fear and were often perception-based rather than production-based, such as Pfitzinger (2011). Therefore, it is possible that the results of this production-based emotional affect study point towards further nuance within the negative valence / high arousal subfield. More research would need to be conducted in the future to corroborate these findings. Meanwhile, sadness, a negative valence / low arousal emotion, also had comparatively high use of breathy voice. While this also goes against the association of high arousal with higher rates of voice quality production found in previous studies, other research that has specifically examined the acoustic qualities of sadness have found a use of breathy voice (Erickson et al., 2006; Gobl et al., 2002; Yanushevskaya et al., 2005). This further suggests that the association between voice quality and arousal may be more nuanced than previously thought, and that further research should be conducted.

In sum, the results of this study were in agreement with previous research on the use of suprasegmentals in that negative valence emotions were shorter and high arousal emotions were louder and generally higher pitched. However, nuances were discovered in that not every high arousal emotion was high pitched (i.e. disgust) and that voice quality has a more complicated relationship with valence and arousal than previously found in other studies. Thus, benefits arise

from conducting a production-based task that examines emotional affects in conjunction with valence and arousal, rather than one or the other.

Table 4 Summary of suprasegmental's results against the emotional affects and the emotional affects' values.

Emotion	**Emotional Value**		**Acoustic Measurements**			
	Valence	**Arousal**	**Duration**	**Pitch**	**Intensity**	**Voice**
Disgust	Negative	High	Longer	Lowest	Louder	Most irregular
Fear	Negative	High	Longer	Highest	Louder	Regular
Happiness	Positive	High	Shorter	Highest	Louder	Regular
Sadness	Negative	Low	Longest	Lowest	Softer	Regular
Neutral	0	Low	shortest	Low	Softer	Regular

5.0 CONCLUSION

The main conclusion that should be gleaned from this study is that conducting a cross-linguistic production task allows for a more nuanced interpretation of the relationship between emotional affects and the use of suprasegmentals. Having the participants produce five distinct emotional affects allowed this study to uncover patterns within the emotions themselves, such as the distinction between disgust and fear in terms of voice quality, which a simple arousal vs. valence perception task may have glossed over; the relatively lower significance of voice quality may also suggest an overall hierarchy among the suprasegmentals in the role of emotion production. Additionally, the cross-linguistic aspect revealed that even two typologically unrelated languages can display high similarity in the use of suprasegmentals to express emotions, which may lend support to the Eckman (1999) theory that the five emotions in this study (happiness, sadness, disgust, fear, and neutral) are indeed universal. Still, the differences found between the two native language groups in terms of intensity maximum and pitch average, as well as the trace evidence of L1 transfer in these features, could suggest that there is a slight differentiation between the two languages for emotional production.

5.1 Limitations and Future Studies

The main limitation of this study was the relatively small sample size, with only four representatives for both native English and native Korean speakers. Therefore, the conclusions from this study are not meant to be applied to the general population of each language group, but

rather as a starting point for further cross-linguistic comparison. Future studies may do well to focus on the use of one particular suprasegmental in emotion production with an overall larger sample size.

Future research based off the findings of this study could be taken in many directions. First, future production-oriented tasks should consider the differences in the use of suprasegmentals between emotional affects of the same valence and arousal subfield, as the distinction between fear and disgust in terms of voice quality suggests there is more to be uncovered. Additionally, other emotional affects could be examined for their use of suprasegmentals, specifically ones that have been shown to cause misinterpretation across native language groups. Plus, it is the hope of this study that the 2000 word database created in the process could be used for a cross-linguistic perception task in the future. While results show high similarity between native English and native Korean production, a perception task may reveal differences in the pattern of interpretation between the language groups, as well as further examination of the relationship between overt emotion vs. covert emotion (i.e. the emotional affect with which the word is produced vs. the emotion associated with the word). As mentioned in the introduction, the results of this study have implications for other linguistic subfields, particularly how emotion should be taught in the second language classroom, as well as what this may mean for cross-cultural communication. In sum, this study may work as a starting point for further production studies, perception studies, and/or cross-linguistic studies in a multitude of subfields.

Appendix A LIST OF DATABASE WORDS

	English	*Korean Romanization*	*English Translation*
Happy-laden	1 laughter	1 *satang*	*candy*
	2 hug	2 *mujigae*	*rainbow*
	3 flower	3 *kiseu*	*kiss*
	4 healthy	4 *jangmi*	*rose*
	5 winner	5 *seungjin*	*promotion*
Sad-laden			
	1 guilt	1 *jakbyeol*	*farewell*
	2 apology	2 *hakdae*	*abuse*
	3 cancer	3 *myoji*	*cemetery*
	4 pity	4 *oeroum*	*loneliness*
	5 funeral	5 *bingon*	*poverty*
Disgust-laden			
	1 roach	1 *bakwibeolle*	*roach*
	2 mold	2 *bakteria*	*bacteria*
	3 stain	3 *goreum*	*pus*
	4 puke	4 *konmul*	*snot*
	5 rot	5 *banggwi*	*fart*
Fear-laden			
	1 tortured	1 *yuryeong*	*ghost*
	2 kidnapped	2 *napchi*	*kidnap*
	3 threat	3 *angmong*	*nightmare*
	4 murderer	4 *gyeongnyeol*	*violence*
	5 demon	5 *angma*	*devil*
Neutral-laden			
	1 soap	1 *pen*	*pen*
	2 grass	2 *jandi*	*grass*
	3 newspaper	3 *yakguk*	*pharmacy*
	4 pen	4 *teibeul*	*table*
	5 weather	5 *uisang*	*wardrobe*

Appendix B EMOTIONAL SCENARIOS

Below are the emotional scenarios that the participants read aloud before eliciting the appropriate emotional productions. Each participant read the scenarios in their native language. The Korean scenarios are a direct translation from the English.

Happiness:

It is your birthday. The weather is perfect and you are surrounded by your family and friends. Everybody is laughing. They surprise you with a gift and you spend the rest of the night smiling.

오늘은 당신의 생일입니다. 날씨도 좋으며 가족과 친구들과 함께 있습니다. 모두가 웃고 있습니다. 깜짝 선물도 받고 하루 종일 즐겁습니다.

Sadness:

You return home alone. There is an email on your phone saying that you did not get the job. You feel guilt and shame when you realize that you still have not paid the rent. The house is cold and empty.

집에 홀로 들어옵니다. 핸드폰으로 이메일을 확인하는데 취직 시험에 불합격했다는 메시지입니다. 그리고 아직 월세를 내지 못 했다는 사실에 슬퍼집니다. 텅 빈 집은 춥기만 합니다.

Disgust:

You walk into the kitchen and notice an overwhelming odor. The sink is filled with dirty dishes and the trash is starting to spill onto the floor. You see that the bread is now covered in mold.

부엌에 들어 갔는데 역겨운 냄새가 진동을 합니다. 싱크대에는 더러운 식기가 쌓여 있고 쓰레기는 바닥에 흩어져 있습니다. 곰팡이가 핀 빵도 보입니다.

Fear:

It is midnight. You are walking down an abandoned street alone when a van screeches to a stop beside you and two masked figures jump out and throw you in the back. You scream right before one of them knocks you unconscious.

밤 열두 시입니다. 아무도 없는 길을 홀로 걸어 가고 있는데 미니밴이 갑자기 옆에 서더니 복면을 한 두 사람이 차에서 내려 당신을 차 안으로 밀어 넣습니다. 비명을 질러 보지만 주먹으로 맞은 후 의식을 잃습니다.

Neutral:

You walk into the classroom and sit down. You pull out your laptop and get ready to take notes. You notice that the professor has not arrived yet. You say hi to some classmates and wait for class to begin.

당신은 교실로 들어가 자리에 앉습니다. 노트북 컴퓨터를 꺼낸 후 노트할 준비를 합니다. 교수님은 아직 강의실에 안 오신 것 같습니다. 몇 명의 반 친구들에게 인사도 하면서 수업이 시작되길 기다립니다.

BIBLIOGRAPHY

Abelin, A., and Allwood, J. (200). Cross linguistic interpretation of emotional prosody. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*. 2000.

Altarriba, J., & Bauer, L. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology*, *117*(3), 389-410. doi: 10.2307/4149007

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351-373.

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

Banse, R. & Scherer, K.R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614-636. https://doi.org/10.1037/00223514.70.3.614

Briefer, E. F. (2012). Vocal expression of emotions in mammals: mechanisms of production and evidence. *Journal of Zoology*, *288*(1), 1-20.

Bu, Jiemin. (2012). A study of relationships between L1 pragmatic transfer and L2 proficiency. *English Language Teaching*, *5*(1), 32-43.

Campbell, L. (2010). Language isolates and their history, or, what's weird anyway? *Linguistic Society of America*, *36*(1), 16-31. https://doi.org/10.3765/bls.v36i1.3900

Cason, N., Marmursztejn, M., D'Imperio, M., & Schön, D. (2019). Rhythmic abilities correlate with L2 prosody imitation abilities in typologically different languages. *Language and Speech*, *63*(1), 1-17. doi: 10.1177/0023830919826334

Cheng, J. T., Tracy, J. L., Ho, S., & Henrich, J. (2016). Listen, follow me: Dynamic vocal signals of dominance predict emergent social rank in humans. *Journal of Experimental Psychology: General*, *145*(5), 536–547. https://doi.org/10.1037/xge0000166

Crystal, D. (1981). Segmental Phonology. In: Clinical Linguistics, *Disorders of Human Communication*, (pp. 23-57). Springer-Verlag Wien. https://doi.org/10.1007/978-3-7091-400107_2

Dauer, R.M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics*, *11*(1), 51-62.

Diaz, A. (2017). Towards improving Japanese EFL learners' pronunciation: The impact of teaching suprasegmentals on intelligibility. *APU Journal of Language Research*, 2, 69-84.

Do, H.J., & Choi, H.J. (2015). Korean twitter emotion classification using automatically built emotion lexicons and fine-grained features. In *Proceedings of the 29th Pacific Asia Conference on Language, Information, and Computation: Posters* (pp. 142-150).

Dupuis, K., & Pichora-Fuller, M. K. (2014). Intelligibility of emotional speech in younger and older adults. *Ear & Hearing*, *35*(6), 695-707.

Ekman, P. (1999). Basic emotions. In T. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (pp. 45-60). John Wiley & Sons.

Erickson, D., Yoshida, K., Menezes, C., Fujino, A., Mochida, T., and Shibuya, Y. (2006). Exploratory study of some acoustic and articulatory characteristics of sad speech. *Phonetica*, *63*(1), 1-25. doi: 10.1159/000091404

Fontaine, J., Poortinga, Y., Setiadi, B., & Markam, S. (2010). Cognitive structure of emotion terms in Indonesia and the Netherlands. *Cognition and Emotion*, *16*(1), 61-86. doi: 10.1080/02699933014000130

Gobl, C., and Chasaide, A.N. (2003). The role of voice quality in communication emotion, mood, and attitude. *Speech Communication*, *40*(1-2), 189-212. https://doi.org/10.1016/S0167-6393(02)00082-1

Gobl, C., Bennett, E., and Chasaide, A.N. (2002). Expressive synthesis: How crucial is voice quality? In *Proceedings of 2002 IEEE Workshop on Speech Synthesis, 2002*. (pp. 91-94). IEEE. doi: 10.1109/WSS.2002.1224380

Gussenhoven, C. (2002). Intonation and interpretation: Phonetics and phonology. *Speech Prosody 2002, International Conference*.

Hancock, A. B., & Rubin, B. A. (2015). Influence of Communication Partner's Gender on Language. *Journal of Language and Social Psychology*, *34*(1), 46-64. https://doi.org/10.1177/0261927X14533197

Jacobi, T., & Schweers, D. (2017). Justice interrupted: The effect of gender, ideology, and seniority at supreme court oral arguments. *Virginia Law Review*,*103*(7), 1379-1485.

Jeon, H.S.(2015). Prosody. In L. Brown & J. Yeon (Eds.), *Handbook of Korean Linguistics* (pp.41-58). New York: Wiley-Blackwell.

Juslin, P.N., & Scherer, K.R. (2005). Vocal expression of affect. In J.A. Harrigan, R. Rosenthal, & K.R. Scherer (Eds.), *Series in Affective Science. The new handbook of methods in nonverbal behavior research* (pp. 65-135). Oxford University Press.

Kollias, D., Panagiotis, T., Mihalis, N., Papaioannou, A., Zhao, G., Schuller, B., Kotsia, I., and Zafeiriou, S. (2019). Deep affect prediction in-the-Wild: Aff-Wild database and challenge, deep architectures, and beyond. *International Journal of Computer Vision*, *127*(6-7), 907-929.

Kuppens, P., Tuerlinckx, F., Russell, J. A., & Barrett, L. F. (2013). The relation between valence and arousal in subjective experience. *Psychological Bulletin*, *139*(4), 917-940. https://doi.org/10.1037/a0030811

Lee, H.Y., & Song, J. (2019). Evaluating Korean learners' English rhythm proficiency with measures of sentence stress. *Applied Psycholinguistics*, 40(6), 1-14. https://doi.org/10.1017/S0142716419000298

Lee, Y., Vakoch, D.A., & Wurm, L.H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527-542.

Li, M., Zhang, S., Li, K., Harrison, H., Lo, W.K., and Meng, H. (2011). Design and collection of an L2 English corpus with a suprasegmental focus for Chinese learners of English. In *ICPhS* (pp. 1210-1213).

Mendoza, E., Valencia, N., Muñoz, J., and Trujillo, H. (1996). Differences in voice quality between men and women: Use of the long-term average spectrum (LTAS). *Journal of Voice*, *10*(1), 59-66. https://doi.org/10.1016/S0892-1997(96)80019-1

Meng, H., Tseng, C.Y., Kondo, M., Harrison, A., and Viscelgia, T. (2009). Studying L2 suprasegmental features in Asian Englishes: A position paper. In *Tenth Annual Conference of the International Speech Communication Association*.

Murray, I., & Arnott, J. (1993). Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2), 1097-1108. https://doi.org/10.1121/1.405558

Patel, S., Scherer, K.R., Bjorkner, E., and Sundberg, J. (2011). Mapping emotions into acoustic space: The role of voice production. *Biological Psychology*, *87*(1), 93-98.

Pavlenko, A. (2008). Emotion and emotion-laden words in the bilingual lexicon. *Bilingualism: Language and Cognition*, *11*(2), 147-164. doi:10.1017/S1366728908003283

Pemberton, C., McCormack, P., and ussell, A. (1998). Have women's voices lowered across time? A cross ectional study of Australian women's voices. *Journal of Voice*, *12*(2), 208-213. https://doi.org/10.1016/S0892-1997(98)80040-4

Pfitzinger, H.R., Amir, N., Mixdorff, H., and Basel, J. (2011). Cross-language perception of Hebrew and German authentic emotional speech. In *ICPhS* (pp. 1586-1589).

Roach, P. (1982). On the distinction between "stress-timed" and "syllable-timed" languages. In D. Crystal (Ed.), *Linguistic Controversies* (pp.73-79). London: Arnold.

Scherer, K.R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin*, *99*(2), 143-165. https://doi.org/10.1037/00332909.00.2.143

Seddighi, S. (2010). An account of Iranian EFL pronunciation errors through L1 transfer. *Iranian Journal of Applied Language Studies*, *2*(2). 197-214.

Shochi, T., Brousse, A., Guerry, M., Erickson, D., & Rilliard, A. (2016). Learning effect of social affective prosody in Japanese by French learners. *Speech Prosody*, *1*(35) 970-973.

Slabakova, R. (2000). L1 transfer revisited: the L2 acquisition of telicity marking in English by Spanish and Bulgarian native speakers. *Linguistics*, *38*(4), 739-770.

Sobin, C, & Alpert, M. (1999). Emotion in speech: The acoustic attributes of fear, anger, sadness, and joy. *Journal of Psycholinguistic Research*, 28(4), 347-365.

Sumner, M. (2015). The social weight of spoken words. *Trends in Cognitive Sciences*, *19*(5), 238-239.

Wenk, B., and Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics*, *10*(2), 193-216.

Xu, Y. (2019). Prosody, tone, and intonation. *The Routledge Handbook of Phonetics*, 314-256.

Yamashita, J. (2002). Mutual comensation between L1 reading ability and L2 language proficiency in L2 reading comprehension. *Journal of Research in Reading*, *25*(1), 81-95. https://doi.org/10.1111/1467-9817.00160

Yanushevskaya, I., Gobl, C., and Chasaide A.N. (2005). Voice quality and f0 cues for affect expression: Implications for synthesis. In Interspeech-2005, 1849-1852.

Zhang, J., Teo, T., & Wu, C. (2018). Emotion words modulate early conflict processing in a flanker task: Differentiating emotion-label words and emotion-laden words in second language. *Language and Speech*, 62(4), 641-651. doi: 10.1177/0023830918807509