

# EMOTIONAL PROSODY OF LOVE AND SORROW: L1 ENGLISH, TTS, AND EFL LEARNERS

**Ikuyo Kaneko**, Juntendo University  
**Noriko Yamane**, Hiroshima University

This study examined English emotional prosody and identified differences in vocal expression of love and sorrow among first language (L1) speakers, learners of English as a foreign language (EFL), and text-to-speech systems (TTS). American professional narrators (AE) and Japanese college students (JP) were instructed to read aloud and record a love letter and a condolence letter. The same letters were also read by synthetic speakers on a web-based TTS. The measurements targeted “booster expressions”—lexical grammatical items of various types which increase the effectiveness of emotional utterances (Quirk & Greenbaum, 1973). The results showed that onset consonant duration, pitch range, intensity, and speech rate had significant differences among groups. These findings suggest that all these features contribute significantly to how AE express love and sorrow, which is still not available through the current default system of TTS. The findings have an impact on the broader topic of speaking English as a foreign language inside and outside classroom environments.

**Cite as:** Kaneko, I., & Yamane, N. (2022). Emotional prosody of love and sorrow: L1 English, TTS, and EFL learners. In J. Levis & A. Guskaroska (eds.), *Proceedings of the 12th Pronunciation in Second Language Learning and Teaching Conference*, held June 2021 virtually at Brock University, St. Catharines, ON. <https://doi.org/10.31274/psllt.13259>

## INTRODUCTION

Previous studies show that prosody plays an important role in successful communication (Chun & Levis, 2020; Levis 1999; Levis & Levis, 2018; Thompson & Balkwill, 2006), and it is important to facilitate L2 learners to be aware of the differences of intonation structure between L1 and L2 (Mennen, 2015). In L2 English classrooms of junior and senior high schools in Japan, it is difficult for learners to be aware of the intonational differences on their own, partly because emotional vocal expressions are not systematically taught, and partly because the materials used are emotion-neutral so that the speakers’ emotion behind the message does not stand out (Uechi & Tanizawa, 2004).

It is well known that pitch realization varies across emotions (Bänziger & Scherer, 2005; Pell, 2001; among others), but it is based on global measures of sentences as a whole. Given the principle of focus prosody (Gussenhoven, 2008; Ishihara, 2011; Katz & Selkirk, 2011; Lee & Xu, 2012; Pierrehumbert & Beckman, 1988; Selkirk, 2002), it would be reasonable to assume that there is a specific timing where acoustic prominence is added. There may be generalized prosodic patterns in expressing types of emotions. For example, positive assessment may be a “high then loud then quiet” pattern—raised pitch, increased loudness, lengthened vowels, sharp drop in loudness and silence or other low intensity (Ward, 2019, p. 98). An interesting observation here is that native speakers of English may utilize pitch, loudness and intensity in different specific points within a sentence, in order to convey their attitudes and emotions. However, it would not be realistic to instruct English L2 students to mimic a variety of acoustic features in different specific timepoints. There are also certain words that tend to add acoustic enhancement: For example, some positive/negative emphasis

words accompany lengthening of onset consonants and/or vowels of the emphasized words by the intensification of its meaning (Nieburhr, 2010). In Japanese, the meaning of degree adjectives and adverbs is often intensified by lengthening a consonant and a vowel (Hamano, 1998) and this kind of lengthening is widespread across Japanese dialects, which can be formulated in terms of mora augmentation (Davis & Ueda, 2002). It is also known that slow speech rate contributes to the perception of sadness, and a fast rate may contribute to the perception of happiness, perhaps cross-linguistically (Carl et al., 2022).

In spoken English, a variety of adverbs may be used to enhance speakers' expressions of emotion. "Booster" expressions are various types of lexical grammatical items which increase the effectiveness of emotional utterances (Quirk & Greenbaum, 1973), so they can be good candidates to show their acoustic prominence in emotional utterances. Holmes (1990) points out that boosters appear to carry unique intonation patterns (e.g., rise-fall), but few studies describe the details of the phonetic characteristics of boosters (Kaneko, 2010a, 2010b). Thus, it would be a good first step to explore the realization of acoustic features of boosters in different types of emotions.

The purpose of this study is to identify acoustic features that effectively convey different types of emotions in speaking English, so that Japanese learners of English can actively employ them in speaking English to succeed in communication.

## **Research Questions**

Based on previous studies, booster expressions can be enhanced by a diversity of suprasegmental properties such as lengthening of segments, raising of pitch and intensity, and speech rate. Our research questions are given below:

- 1) Which booster words are prominent in professional L1 English narrators in read speech?
- 2) To what extent are acoustic features (i.e., consonant duration, pitch, intensity, and speech rate) different between professional narrators of L1 English and Japanese learners of L2 English in readings of a love letter and a condolence letter?

## **METHODS**

### **Participants**

The participants in the experiment were six American professional narrators (AE: three females and three males), and six Japanese university students (JP: three females and three males). All of the American professional narrators were recruited at the English Language Education Council, Inc. (ELEC) in Tokyo. The Japanese learners were all university students in Japan, whose TOEIC test average score was 574 (Highest score: 795, Lowest score: 280). Moreover, six synthetic speakers (three females and three males) on a web-based text-to-speech system (TTS) were used as controls. NaturalSoft (2020) was adopted for TTS voices, and we extracted US English voices in order to compare these voices to American professional narrators.

## Reading Materials

The linguistic materials used in this experiment were adopted from Kaneko (2010a, 2010b). They were a 185-word love letter and a 186-word condolence letter written by a native speaker of North American English. The love letter was written on the assumption that the writer has a girlfriend/boyfriend with whom s/he has been going out for three years, while the condolence letter was written on the assumption that the writer was conveying feelings and sympathy to the family of a nurturing high school teacher who had suddenly passed away. The letters were adjusted to be gender non-specific by the authors so that the setting of the letters was appropriate for all participants. The linguistic materials are presented in Tables 1 and 2, where sentences for analysis are underlined.

**Table 1**

*Reading Material for the Production Experiment (Love Letter)*

---

Dear My Love,

Today is our third anniversary, and I want to take this time to tell you how much I love you. The truth is the last three years with you have been the happiest of my life. Just the sight of your smile in the morning, and your tender kiss carry me through the day. Let me take you back to the time we first met; I remember the moment when I saw you approach along the path and sit beside me. During all the hours I had spent sitting in the same place, I had never seen such a beautiful smile. I remember the look in your eyes told me how special this moment had become when I plucked up the courage to ask you your name. Since then my feelings for you have grown with every second that passes by.

Today, on the anniversary of that first special moment, I want you to know how deeply I have fallen in love with you. I hope that I can convey how precious you are to me and how I relish every moment we are together.

Yours always,

○○ (a participant's name)

---

## Table 2

### Reading Material for the Production Experiment (Condolence Letter)

---

Dear Mrs. Smith,

You may not know me, but I wanted to express my sincere condolences to you and your family for your recent loss. Your husband, Mr. Smith, was truly my mentor. He was my high school teacher until I left for university last year. During the time when I knew Mr. Smith, he always pushed me to do my best in my studies, and he inspired me with many words of wisdom through the years. I can sincerely tell you that I would not be the person that I am without Mr. Smith.

I realize that this is a very difficult time for you and your family, but I want you to know that Mr. Smith made a big difference in the lives of his students, and the mark he left on this world will not soon be forgotten.

I truly wish that I could have attended the funeral service and honored Mr. Smith's life in this way, but unfortunately my busy class schedule at university has not allowed me to attend. Please know that my thoughts are with you and your family during this difficult time.

Respectfully yours,

○○ (a participant's name)

---

## Tasks

AE and JP participants were instructed to read the letters orally to make audio letters. They were asked to produce their utterance at a comfortable speaking rate and to put the emotion they considered appropriate into the words. Before the recording, the participants were provided enough time for practice and were verbally confirmed that they had no unknown words or expressions. They repeated the text five times. When they made a mistake, they were asked to repeat the sentence where they made a mistake from the beginning. Their utterances were recorded individually. JP participants received no specific teaching or instruction about how they should read the letters, neither from native speakers of English nor from teachers of English.

The same letters were also read by TTS, on the assumption that the speech created should be neutral in emotion. The speech speed was set as 0 or the default speed.

## Targets for Analysis

In the love letter, “much,” “never,” and “such” are booster expressions, while “truly” and “sincerely” are regarded as booster expressions in the condolence letter. Authors observed pitch contours of all sentences including boosters in Praat (Boersma & Weenink, 2021) and selected the sentences that show major differences between AE and JP. From the two types of letters, sentences below were targeted for analysis.

- a. *I want to take this time to tell you how **much** I love you.* (Love letter)
- b. *I had **never** seen **such** a beautiful smile.* (Love letter)

- c. *Your husband, Mr. Smith, was **truly** my mentor.* (Letter of condolence)  
d. *I can **sincerely** tell you that I would not be the person that I am without Mr. Smith.* (Letter of condolence)  
(Bold faced words: boosters; underlined words: neighboring content words)

## Measurement

For each participant, one repetition of the text was selected from the total of five repetitions uttered. Second or third trial of the recordings was selected randomly. The measurement of the target words was done for the F0, intensity, duration, and speech rate in the speech samples using Praat. The details will be shown in the following section.

The measurement of pitch was done in semitones to normalize for gender. Pitch range here means the semitone difference between boosters and neighboring non-boosters (i.e., neighboring function words were not considered). The difference should show the degree of prominence of the boosters, so that the greater the difference, the greater the prominence. For example, the pitch range for “much” is the subtracted value of “tell” from “much,” and the one of “love” from “much.” Here we took the average of these two values. If there are no content neighboring words on both sides (e.g., “never,” “sincerely”), only one side of the content word was included for the calculation. The rationale of looking at the pitch range comes from focus prosody. Speakers orally highlight parts of speech to listeners, which is new in discourse and contrastive in semantics (Gussenhoven, 2008; Katz & Selkirk, 2011). Despite some differences in realization across types and languages (Mizuguchi & Tateishi, 2020), the major consensus is that pitch is raised on focus words and is compressed on pre-focal/post-focal words (Ishihara, 2011; Lee & Xu, 2012; Pierrehumbert and Beckman, 1988; Selkirk, 2002). However, it is also found that L2 learners’ pitch range of focus words is smaller compared to native speakers of English (Fujimori et al., in press), thus it is expected that professional speakers may have a greater pitch range for the boosters compared to learners.

The measurement of intensity (dB) was made for neighboring content words as well. If there were neighboring content words on both sides, both were measured. Since the average intensity varies across speakers, the calculation was done for the ratio per speaker; booster intensity (dB) was divided by the neighboring non-boosters (dB). (Non-boosters mean content words that are not boosters.) Compared to F0, the effect of intensity on prosody is not understood well. Visceglia et al. (2012) reports that L1 English speakers raised the intensity on focus words more robustly compared to L2 English of Mandarin speakers, suggesting that their L1 use of acoustic correlates of focus may be transferred to L2 English. As for Japanese focus prosody, Lee and Xu (2012) found that the distribution of the intensity is similar to F0, in that the focus words should carry higher intensity and the pre-focus and post-focus words lower it. Thus, it is worth seeing how intensity in L2 English of Japanese behave in focus locations.

Based on our auditory impression, the length of onset consonant was determined as a target of measurement as well. This is inspired by Keating et al. (2004), who clarified a phenomenon of domain-initial strengthening by showing EPG data where initial consonants of prosodic domain often get lengthened in several languages. One of their interesting findings is the degree of lengthening correlates to the height of the prosodic domain where the word is situated: As the domain is higher, the initial consonant is stretched longer. This reasoning can be extended to initial consonants of

booster words, since a booster may create a new focus domain by inserting a prosodic boundary at the left edge (Pierrehumbert & Beckman, 1988), which would match with an upper phrase such as intonational phrase. Thus it may not be surprising if the lengthening of the onset consonant of a booster takes place, at least for native speakers of English. As for L1 Japanese speech, no robust result has been reported about lengthening of an initial consonant. Therefore, some difference may emerge between groups.

Speech rate was also measured, because it is generally known that L2 speech is slower than L1 speech (Derwing et al., 2009), and sad speech is slower and happy speech is faster (see Carl et al., 2022, for within-language differences), but it is not known how this pattern emerges in oral reading of letters and how professional narrators and learners perform. We measured the duration of the whole paragraph of the love letter (185 words) and the condolence letter (186 words). Then we divided the number of syllables by seconds (sec).

## RESULTS

### Words of Acoustically Enhanced Boosters

RQ1 was about which booster words would be made prominent by professional L1 English narrators. This is done by the comparison between AE, TTS, and TTS voices that convey “neutral” emotion (Robinson, et al., 2019). Statistical analysis was done by one-way ANOVAs (JASP, 2022). The results indicate that not all boosters were acoustically enhanced, and not all acoustic features were employed by professional narrators. This is summarized in Table 3.

**Table 3**

*Enhanced Prosodic Features of Boosters in the Love Letter and the Condolence Letter: AE vs. TTS*

Type of letter	Boosters	Consonant length	Pitch range	Intensity
Love Letter	<i>much</i>	<b>A &gt; T</b>	n.s.	n.s.
	<i>never</i>	n.s.	n.s.	<b>A &gt; T</b>
	<i>such</i>	n.s.	n.s.	n.s.
Letter of Condolence	<i>truly</i>	<b>A &gt; T</b>	n.s.	n.s.
	<i>sincerely</i>	n.s.	n.s.	n.s.

*Note.* A = American English speakers; T = Text-to-Speech Voice. A > T denotes A is statistically larger than T ( $p < .05$ ).

In love letter readings, “much” was enhanced in consonant duration, and “never” was enhanced only in intensity, and “such” showed no difference. In the letter of condolence, “truly” was enhanced in consonant length only. “Sincerely” has no significant difference between AE and TTS.

### American Professional Narrators vs. Japanese Learners of English

RQ1 asked about the extent to which acoustic features were different between professional narrators of L1 English and Japanese learners of L2 English in the letter reading. The results showed that there

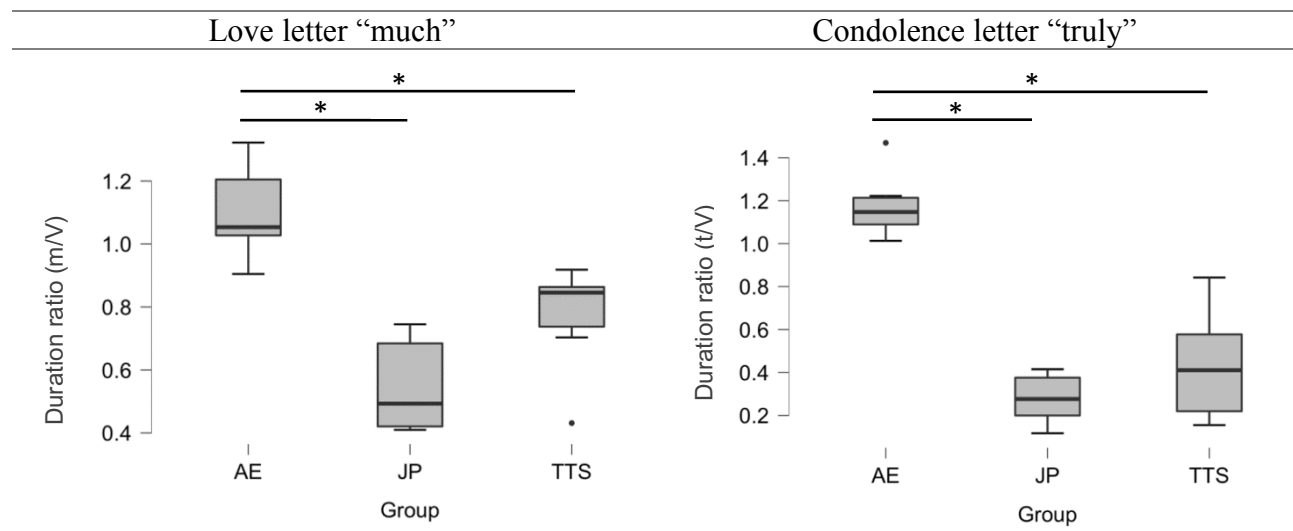
were significant differences in the length of onset consonant of boosters, pitch range, intensity, and speech rate.

### Onset Consonant Length

The onset consonant length was statistically different in the words “much” in the love letter and “truly” in the condolence letter between groups (Figure 1).

**Figure 1**

*Onset consonant duration. Duration of the onset consonant divided by duration of the following vowel. Group comparison: American (AE), Japanese (JP) and TTS (Text-to-Speech).*



The results show the consonant [m] in the word “much” was significantly longer in Americans, compared to TTS and Japanese ( $F(2, 15) = [9.559]$ ,  $p < .01$ ). Tukey’s HSD test for multiple comparisons found that the mean value of m/V was significantly different between AE and JP ( $p = 0.006$ , 95% CI = [0.019, 0.109]), and different between AE and TTS ( $p = 0.004$ , 95% CI = [0.022, 0.112]), but there was no difference between TTS and JP ( $p = 0.081$ ).

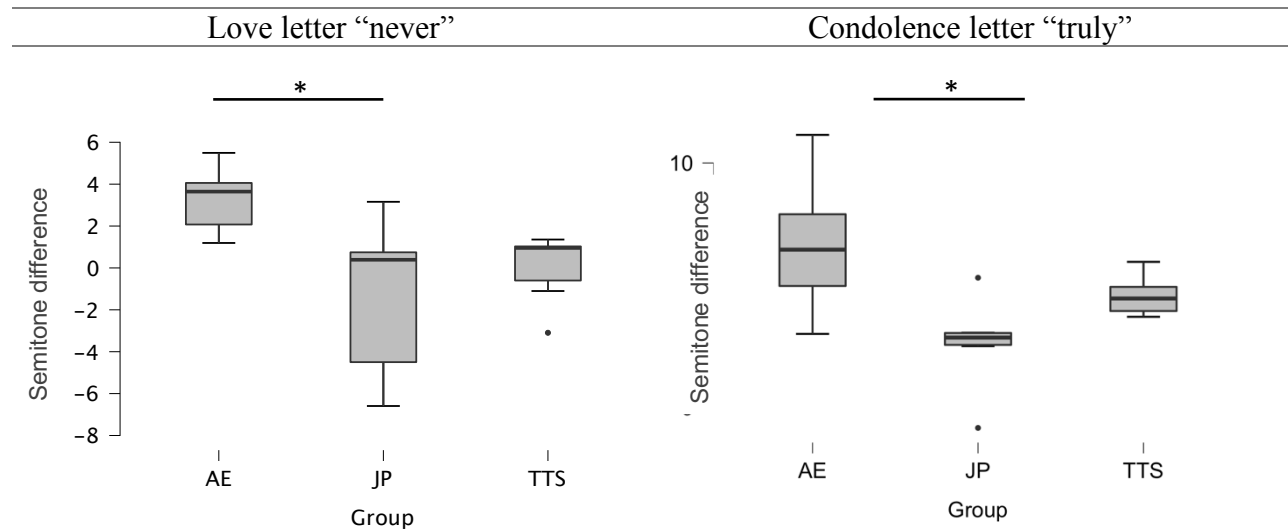
For consonant duration, Americans’ [t] sound in “truly” was significantly longer than that produced by both Japanese and TTS voices ( $F(2, 15) = [37.629]$ ,  $p < .001$ ). Tukey’s HSD test for multiple comparisons found that the mean value of t/V was significantly different between AE and JP ( $p < 0.01$ , 95% CI = [0.613, 1.191]), and different between AE and TTS ( $p < 0.001$ , 95% CI = [0.457, 1.034]), but there was no difference between TTS and JP ( $p = 0.362$ ).

### Pitch range

The pitch range was statistically different in the words “never” in the love letter and “truly” in the condolence letter between groups (Figure 2).

**Figure 2**

*Pitch range*



Note. Semitone (st) difference of “never” and “seen.” The average of (st of “truly” – st of “Smith”) and (st of “truly” – st of “mentor”). Group comparison: American (AE), Japanese (JP) and TTS (Text-to-Speech)

In the love letter readings, the pitch range of “never” was significantly different between groups ( $F(2, 15) = [4.620], p < .05$ ). Tukey’s HSD test for multiple comparisons found that the mean value of “never” was significantly different between AE and JP ( $p < 0.05$ , 95% CI = [0.556, 8.656]), but there was no statistically significant difference between AE and TTS ( $p = 0.124$ ), or between JP and TTS ( $p = 0.675$ ).

In the condolence letter readings, the pitch range of “truly” was significantly different between groups ( $F(2, 15) = [5.689], p < .05$ ). Tukey’s HSD test for multiple comparisons found that the mean value of “truly” was significantly different between AE and JP ( $p < 0.05$ , 95% CI = [0.927, 7.140]), but there was no statistically significant difference between AE and TTS ( $p = 0.223$ ), or between JP and TTS ( $p = 0.263$ ).

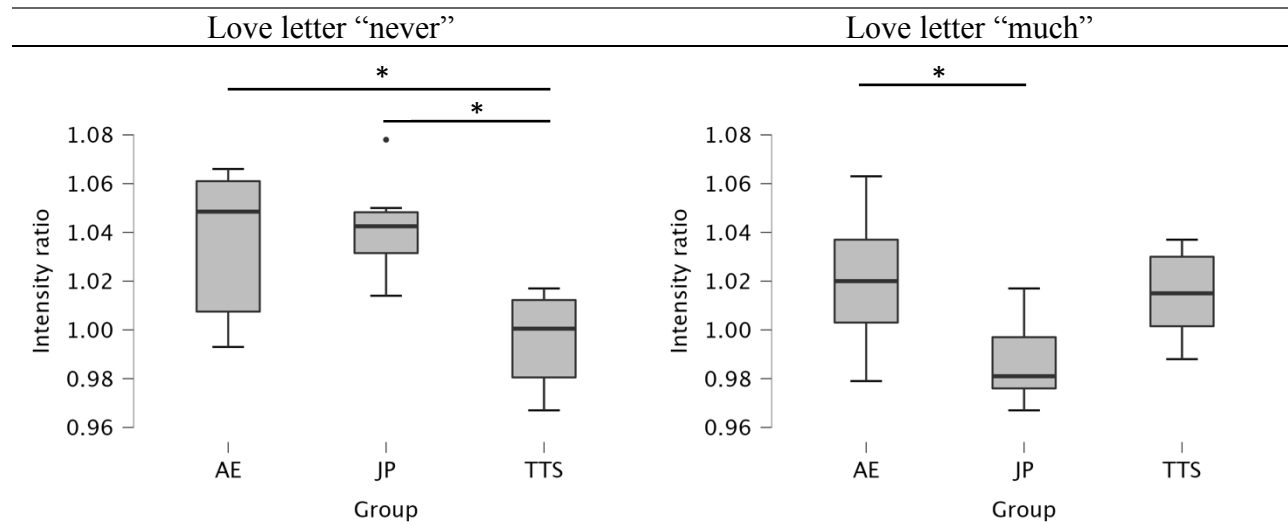
**Intensity**

The intensity was statistically different in the words “never” and “much” in the love letter (Figure 3) while no word intensity difference was found in the condolence letter between groups.



**Figure 3**

*Intensity. The intensity ratio of booster against non-boosters*



*Note.* The average of (dB of “never” / dB of “seen”); There were no non-boosters on the left of “never.” The average of (dB of “much” / dB of “tell”) and (dB of “much” / dB of “love”). Group comparison: American (AE), Japanese (JP) and TTS (Text-to-Speech).

In the love letter readings, intensity of the word “never” was significantly different at least between 2 groups ( $F(2, 15) = [5.689], p < .05$ ). Tukey’s HSD test for multiple comparisons found that the mean value of “never” was significantly different between AE and TTS ( $p < 0.05$ , 95% CI = [0.001, 0.079], and different between JP and TTS ( $p < 0.05$ , 95% CI = [0.008, 0.015], but there was no statistically significant difference between AE and JP ( $p = 0.897$ ).

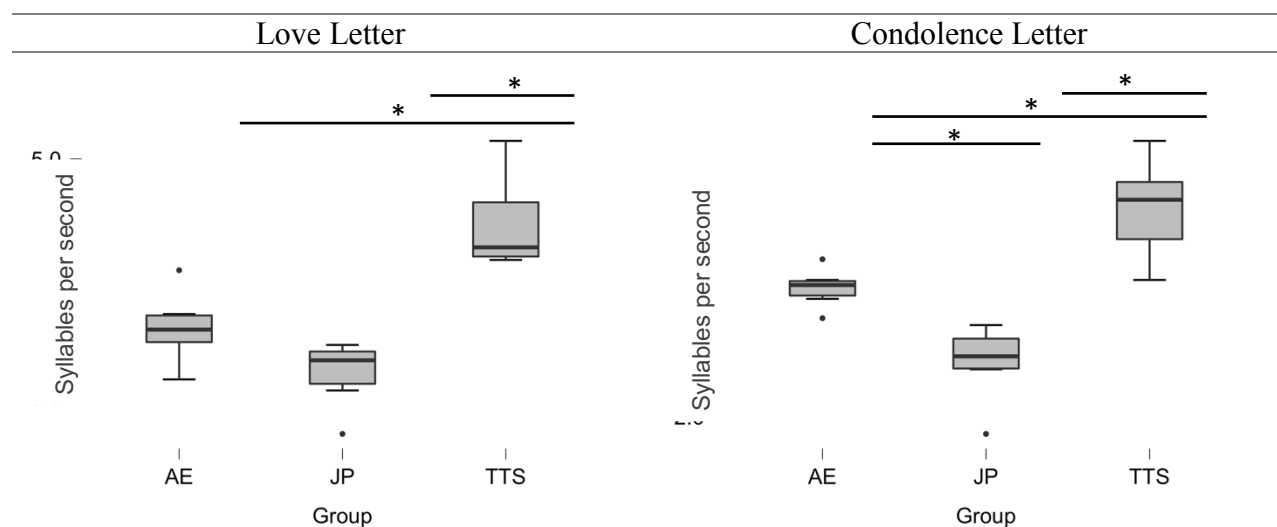
Intensity of the word “much” was significantly different at least between 2 groups ( $F(2, 15) = [4.285], p < .05$ ). Tukey’s HSD test for multiple comparisons found that the mean value of “much” was significantly different between AE and JP ( $p < 0.05$ , 95% CI = [0.002, 0.074], and there was no statistically significant difference between AE and TTS ( $p = 0.909$ ), or between JP and TTS ( $p = 0.087$ ).

### Speech Rate

The results show TTS was significantly faster than the American readers, who were also significantly faster than the Japanese readers (Figure 4).

**Figure 4**

Speech rate. Group comparison: American (AE), Japanese (JP) and TTS (Text-to-Speech).



The speech rate of the love letter was significantly different in at least 2 groups ( $F(2, 15) = [22.035]$ ,  $p < .001$ ) Tukey's HSD test for multiple comparisons found that the mean value of articulation rate was significantly different between AE and TTS ( $p < 0.01$ , 95% CI = [-1.380, -0.370]), and different between JP and TTS ( $p < .001$ , 95% CI = [-1.764, -0.754]), but there was no statistically significant difference between AE and JP ( $p = 0.152$ ).

The speech rate of the condolence letter was significantly different between groups ( $F(2, 15) = [24.461]$ ,  $p < .001$ ). Tukey's HSD test for multiple comparisons found that the mean value of articulation rate was significantly different between AE and JP ( $p < 0.05$ , 95% CI = [0.115, 0.864]), and different between AE and TTS ( $p < 0.01$ , 95% CI = [-0.893, -0.144]), and different between JP and TTS ( $p < 0.001$ , 95% CI = [-1.382, -0.634]).

## DISCUSSION AND CONCLUSION

In regard to RQ1, only a subset of boosters (i.e., “much,” “never,” “truly”) indicated significant differences between AE and TTS (Table 3). In professional speakers' readings, duration of the onset consonant of boosters of “much” and “truly” is lengthened, and intensity of “never” was enhanced compared to TTS. The result of consonant lengthening of [m] and [t] may be compatible with the articulatory strengthening at domain-initial positions (Cho, 2002; Keating et al., 2004). In Cho's kinematic studies, lengthening of consonants and vowels are observed at the edges of prosodic domain, which can be partly derived from syntactic constituents. Although it was not discussed how this phenomenon can be affected by a speaker's attitudes, the prosodic domain may be restructured by this lengthening, where a booster could behave like a focus-initial position (see Truckenbrodt, 1995, for formalized patterns of prosodic restructuring under focus). It should be safe to say that lengthening of a consonant closure could be employed as one of the enhancement strategies.

It was also found that the American narrators read slower to express their emotions of love or sadness compared to TTS. It is generally said that sad speech is slower in rate than happy speech (see Carl et al., 2022, for within-language differences), and this is true for narrators' readings. In principle, it should be possible for professional narrators to read the love letter with a happy emotion in a faster pace than neutral TTS, but actually it turned out to be slower, suggesting that they adopted a slow style. In spontaneous speech, the average speech rate of native speakers is found to be 4.7 syllables per second (Derwing & Munro, 2001), while the read speech varies depending on the styles: For example, one study (Arnfield et al., 1995), based on the reading of *North Wind and the Sun*, shows a range of 3.3. and 5.9 syllables per second. The result of the love letter reading of professionals here is still within this range, but the speech rate of the condolence letter was slower than the range.

In regard to RQ2, statistically significant differences between Japanese learners of English and professional narrators include i) consonant lengthening in “much” and “truly,” ii) pitch range of “never” and “truly,” iii) intensity ratio of “much,” and iv) speech rate of the condolence letter (underlined in Table 4). Among them, the poorest performance JP did was consonant lengthening. The consonant duration ratio of JP was shorter than both TTS and professional narrators. This can be ascribed to the phonological differences of English and Japanese: Stop consonants such as [t] in Japanese are weakly aspirated therefore VOT is shorter than English (Riney et al., 2007). Similarly, consonant [m] in Japanese also involves only a weak constriction of labial closure, which could make the consonant duration shorter than its English counterpart. Interestingly, however, professional narrators lengthened them significantly longer than TTS. We conjecture consonant lengthening is a strategy of acoustic enhancement that the professionals employ in order to show emotions effectively in vocal performance. Therefore, it is not just about the segmental difference between the two languages but may be seen as a way to make the speech expressive. The next thing Japanese speakers need to work on was pitch range modulation. This was expected, as previous research often reported L2 English intonation is less dynamic. This is partly because of the strong resistance to change pitch accent location that is lexically determined in Japanese, which makes global pitch contour less variable compared to English. What JP handled well is enhancement of intensity of “never.” In contrast, JP did not enhance “much.” Both are quantity boosters rather than quality boosters, but “never” indicates extreme negativity, while “much” shows a positive meaning. Because the material used in this experiment was limited, the reason why JP enhanced intensity only in “never,” not in “much,” is unclear.

**Table 4**

*Prosodic Features in the Love Letter and the Condolence Letter: Comparing AE, TTS, and JP*

Type of letter	Boosters	Consonant length	Pitch range	Intensity	Speech rate
Love Letter	<i>much</i>	<u>A &gt; J, T</u>	n.s.	<u>A &gt; J</u>	T > A, J
	<i>never</i>	n.s.	<u>A &gt; J</u>	<u>A, J &gt; T</u>	
	<i>such</i>	n.s.	n.s.	n.s.	
Letter of Condolence	<i>truly</i>	<u>A &gt; T, J</u>	<u>A &gt; J</u>	n.s.	T > <u>A &gt; J</u>
	<i>sincerely</i>	n.s.	n.s.	n.s.	

*Note.* A = American English speakers; T = Text-to-Speech Voice; J = Japanese speakers. “A > J, T” denotes A is statistically larger than J and T ( $p < .05$ ). “J, T” denotes no statistically significant difference between J and T, though J is larger in number than T.

As a pedagogical implication of emotional speech, consonant lengthening teaching may be worth considering in class, by showing kinematics of lips and tongue of speakers of English, and by showing acoustic signals in Praat, and by giving them a chance to compare them with their own speech and discuss the difference between them in class. Another thing that is not taught in general English classes is about the modulation of speed of reading emotional letters. Speech rate or articulation rate may become faster by repetitive oral reading practice, but it should also be taught that speed may convey different impression to listeners. Professional speakers not only change the speed globally depending on the kind of emotion, but also vary the speed locally. One of the strategies to make the speech expressive was to make the constriction of the closure of initial consonant of boosters stronger and longer. The remaining issues are what kinds of consonants are likely to undergo lengthening, and whether it is different depending on grammatical position. These can be investigated through an independent study which requires a proper experimental design with carefully controlled tokens.

In this study, we have explored how professional narrators read the emotional messages and what acoustic differences could be detected between professionals and Japanese speakers. These materials would be of good use for students' prosody education, which should give them a chance to perceive and produce emotional speech, and promote their awareness of differences between Japanese and English intonation, which may in some way to contribute to successful cross-cultural communication.

## **ABOUT THE AUTHORS**

**Ikuyo Kaneko** (ikaneko@juntendo.ac.jp) is an associate professor in the Faculty of Health Care and Nursing at Juntendo University. Her research interests include phonetics, phonology, and second language acquisition. She is currently interested in emotional expressions observed in both speaking and writing produced by Japanese learners of English. For the analysis of writing, she adopts Appraisal theory of Systemic Functional Linguistics (SFL) as a framework. These days, she is working on a text mining approach to examine the usage of emotional expressions. She teaches English grammar for TOEFL, nursing English, and English phonetics, aiming to foster Japanese people who have a practical command of English.

**Noriko Yamane** (yamanen@hiroshima-u.ac.jp) is an associate professor at Hiroshima University. Her interests include applied phonetics and phonology. She is passionate about synthesizing theory and practice, visualizing tongue movements by ultrasound, acoustic movements by Praat, and integrating co-speech gestures or haptic techniques in pronunciation teaching, so that students and prospective teachers can be empowered to enjoy learning language as science. Currently, she is working on articulatory training of English L/R pronunciation for Japanese using ultrasound visual feedback, timing of co-speech gestures in L1 and L2, and stability of sensorimotor synchronization between intonations and body movements. She teaches English communication, English pronunciation methods, phonetics and phonology, introduction to linguistics, and introduction to phonetic analysis. She is also a content expert of A Visual Language Learning Tool of The University of British Columbia (<https://enunciate.arts.ubc.ca/>).

## REFERENCES

- Arnfield, S., Roach, P., Setter, J., Greasley, P., & Horton, D. (1995). Emotional stress and speech tempo variability. *Proceedings of the ESCA/NATO Workshop on Speech Under Stress*, 13–15.
- Bänziger, T., & Scherer, K. R. (2005). The role of intonation in emotional expressions. *Speech Communication*, 46(3–4), 252–267. <https://doi.org/10.1016/j.specom.2005.02.016>
- Boersma, P., & Weenink, D. (2021). *Praat: doing phonetics by computer* (Version 6.1.40) [Computer program]. <https://www.praat.org>
- Carl, M., Icht, M., & Ben-David, B. M. (2022). A cross-linguistic validation of the test for rating emotions in speech: Acoustic analyses of emotional sentences in English, German, and Hebrew. *Journal of Speech, Language, and Hearing Research*, 65(3), 1-10. [https://doi.org/10.1044/2021\\_JSLHR-21-00205](https://doi.org/10.1044/2021_JSLHR-21-00205)
- Cho, T. (2002). *The effects of prosody on articulation in English*. New York: Routledge.
- Chun, D. M., & Levis, J. M. (2020). Prosody in second language teaching: Methodologies and effectiveness. In C. Gussenhoven & A. Chen (Eds.), *The Oxford handbook of language prosody* (pp. 619–632). New York: Cambridge University Press.
- Davis, S., & Ueda, I. (2002). Mora augmentation processes in Japanese. *Journal of Japanese Linguistics*, 18(1), 1–24. <https://doi.org/10.1515/jjl-2002-0104>
- Derwing, T., & Munro, M. J. (2001). What speaking rates do non-native listeners prefer? *Applied Linguistics*, 22(3), 324–337. <https://doi.org/10.1093/applin/22.3.324>
- Derwing, T. M., Munro, M. J., Thomson, R. I., & Rossiter, M. J. (2009). The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition*, 31(4), 533–557. <http://doi.org/10.1017/S0272263109990015>
- Fujimori, A., Nakayama, M., Yoshimura, N., Yamane, N., Teaman, T., & Yoneyama, K. (in press). Development of L2 prosody: The case of information focus. *XV Selected Proceedings of GASLA 15*.
- Gussenhoven, C. (2008). Types of focus in English. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and focus: Cross-Linguistic Perspectives on Meaning and Intonation* (pp. 83–100). Dordrecht: Springer. [https://doi.org/10.1007/978-1-4020-4796-1\\_5](https://doi.org/10.1007/978-1-4020-4796-1_5)
- Hamano, S. (1998). *The sound-symbolic system of Japanese (Studies in Japanese Linguistics 10)*. Tokyo: Kurosio Publishers.
- Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language & Communication*, 10(3), 185–205. [https://doi.org/10.1016/0271-5309\(90\)90002-S](https://doi.org/10.1016/0271-5309(90)90002-S)

- Ishihara, S. (2011). Japanese focus prosody revisited: Freeing focus from prosodic phrasing. *Lingua*, 121(13), 1870-1889. <https://doi.org/10.1016/j.lingua.2011.06.008>
- JASP Team (2022). JASP (Version 0.16.1) [Computer software].
- Kaneko, I. (2010a). An experimental study of English emotional prosody among Japanese college students. *Journal of Health and Sports Science Juntendo*, 1(3), 385–395.
- Kaneko, I. (2010b). Emotional prosody and emotional expression in English: Comparison between Japanese learners and native speakers. *Theory of Information Culture* 9, 11–30.
- Katz, J., & Selkirk, E. (2011). Contrastive focus vs. discourse-new: Evidence from phonetic prominence in English. *Language* 87(4), 771–816. <https://doi.org/10.1353/lan.2011.0076>
- Keating, P., Cho, T., Fougeron, C., & Hsu, C. S. (2004). Domain-initial articulatory strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Phonetic interpretation: Papers in laboratory phonology VI* (pp. 143–161). New York: Cambridge University Press.
- Lee, A., & Xu, Y. (2012). Revisiting focus prosody in Japanese. In Q. Ma, H. Ding, & D. Hirst (Eds.), *Speech Prosody 2012* (pp. 274–277). ISCA Archive. [https://www.isca-speech.org/archive\\_v0/sp2012/papers/sp12\\_274.pdf](https://www.isca-speech.org/archive_v0/sp2012/papers/sp12_274.pdf)
- Levis, J. M. (1999). Intonation in theory and practice, revisited. *TESOL Quarterly*, 33(1), 37–63. <https://doi.org/10.2307/3588190>
- Levis, J. M., & Levis, G. M. (2018). Teaching high-value pronunciation features: Contrastive stress for intermediate learners. *The CATESOL Journal*, 30(1), 139.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory. In E. Delais-Roussarie, M. Avanzi, & S. Herment. (Eds.), *Prosody and language in contact* (pp. 171–188). Dordrecht: Springer.
- Mizuguchi, S., & Tateishi, K. (2020). Prominence in Japanese is not only cued acoustically. *Proceedings of 10th International Conference on Speech Prosody 2020*, 116–120.
- NaturalSoft Ltd. (2020). Text to speech online. <https://www.naturalreaders.com/>.
- Niebuhr, O. (2010). On the phonetics of intensifying emphasis in German. *Phonetica*, 67(3), 170–198. <https://doi.org/10.1159/000321054>
- Pell, M. D. (2001). Influence of emotion and focus location on prosody in matched statements and questions. *Journal of the Acoustical Society of America*, 109(4), 1668–1680. <https://doi.org/10.1121/1.1352088>
- Pierrehumbert, J., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- Quirk, R., & Greenbaum, S. (1973). *A university grammar of English*. London: Longman.

- Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, 35(3), 439–443. <https://doi.org/10.1016/j.wocn.2006.01.002>
- Robinson, C., Obin, N., & Roebel, A. (2019). Sequence-to-sequence modelling of f0 for speech emotion conversion. *ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6830–6834. <https://doi.org/10.1109/ICASSP.2019.8683865>
- Selkirk, E. (2002). Contrastive FOCUS vs. presentational focus: Prosodic evidence from right node raising in English. *Speech Prosody 2002*, 643–646. [https://www.isca-speech.org/archive\\_open/sp2002/sp02\\_643.html](https://www.isca-speech.org/archive_open/sp2002/sp02_643.html)
- Thompson, W. F., & Balkwill, L-L. (2006). Decoding speech prosody in five languages. *Semiotica*, 158, 407–424. <https://doi.org/10.1515/SEM.2006.017>
- Truckenbrodt, H. (1995). *Phonological phrases—their relation to syntax, focus, and prominence* [Doctoral dissertation, Massachusetts Institute of Technology]. Google Scholar. [https://scholar.google.co.jp/citations?view\\_op=view\\_citation&hl=ja&user=ubVt7JEAAAAJ&citation\\_for\\_view=ubVt7JEAAAAJ:u-x6o8ySG0sC](https://scholar.google.co.jp/citations?view_op=view_citation&hl=ja&user=ubVt7JEAAAAJ&citation_for_view=ubVt7JEAAAAJ:u-x6o8ySG0sC)
- Uechi, Y., & Tanizawa, Y. (2004). *Eigono kankaku kanjo hyogen jiten* [Sensory and emotional expressions in English: Tokyodo [Japanese-English dictionary]. Tokyo: Tokyodo.
- Visceglia, T., Su, C. Y., & Tseng, C. Y. (2012). Comparison of English narrow focus production by L1 English, Beijing and Taiwan Mandarin speakers. *2012 International Conference on Speech Database and Assessments* 47–51. <https://doi.org/10.1109/ICSDA.2012.6422454>
- Ward, N. G. (2019). *Prosodic patterns in English conversation*. New York: Cambridge University Press.