Phonetics Applied: An Acoustic Analysis of Japanese L2 English

Anthony M. Diaz
Miyazaki International College

Reference Data:

Diaz, A. M. (2023). Phonetics applied: An acoustic analysis of Japanese L2 English. In P. Ferguson, B. Lacy, & R. Derrah (Eds.), *Learning from Students, Educating Teachers–Research and Practice.* JALT. https://doi.org/10.37546/JALTPCP2022-10

While experienced English instructors may possess an intuitive knowledge of the difficulties their students face regarding the mastery of English pronunciation, without training in phonetics or sufficient knowledge of the differences between Japanese and English phonology, it may not be apparent how phonetic differences between Japanese and English contribute to the acquisition of intelligible speech by Japanese learners. This may cause instructors to be hesitant to teach the skill of pronunciation due to lacking the confidence or understanding to teach the skill (Baker, 2011). This lack of confidence may be especially applicable to the Japanese EFL context, where non-native instructors are common (Koike, 2014). With reference to this pedagogical issue, this study presents an acoustic analysis of the English speech of seven freshmen Japanese university students conducted with the aim of providing quantitative data regarding Japanese L2 English pronunciation, and to discuss the pedagogical implications that arise from its analysis.

経験豊富な英語のインストラクターは、英語の発音習得の際に学生が直面する困難について直感的な知識を持っているかもしれないが、音声学のトレーニングや日本語と英語の音韻論の違いについての十分な知識がなければ、日本人学習者のわかりやすい発話の習得に貢献できないおそれがある。インストラクターは、発音のスキルを教える自信や理解の不足から、発音指導をためらう可能性がある(Baker, 2011)。この自信の欠如は、ネイティブではないインストラクターが一般的である日本の EFL の状況に特に当てはまる(Koike, 2014)。このような教育学的問題に関して、本研究では、日本人の L2 英語発音に関する定量的データを提供することを目的として実施された、日本人の大学1年生7人の英語音声の音響分析を提示し、その分析から生じる教育学的含意について議論する。

Despite its important role in communication, the skill of pronunciation is often overlooked in the field of applied linguistics (Derwing & Munro, 2005). This is especially true in contexts where English is used as a Lingua Franca between speakers, with research suggesting that a majority of communication breakdowns between non-native speakers (NNS) and non-native listeners are a result of pronunciation errors (Jenkins, 2000). Since most Japanese learners are likely to use English with other NNS, accurate pronunciation is a relevant goal for the Japanese EFL context, yet there still does not seem to be much emphasis on the skill of pronunciation in educational contexts in Japan.

The aim of this paper is to practically apply the field of phonetics by presenting the results of an acoustic analysis of the vowels, stop consonants, and sentence stress of the L2 English pronunciation of Japanese university students. In addition, recordings of two American English speakers, a Canadian English speaker, and a British English speaker were made in order to determine how similar or different the Japanese participants' pronunciation is from the L1 targets. These comparisons will serve as an indication of whether or not participants have acquired a sufficient control of English pronunciation prior to the commencement of their studies at the university level of instruction.

Participants

The participants in this study were seven (3 males, 4 females) freshman students in their first semester of study at a small private university in Southern Kyūshū. All participants were recruited on a volunteer basis, and after receiving clearance for the study from the university's institutional review board, participants received a document written in Japanese thoroughly explaining the research and signed a document of informed consent.

Extent of Participants' Pronunciation Experiences

In order to determine the extent of their experience of explicit pronunciation instruction prior to entering the university, a brief survey was given to all seven participants. Three of the four female participants stated that they had explicitly studied pronunciation, with two indicating that they had practiced pronunciation with their Assistant Language Teachers (ALTs) in high school. Female participant 3 said that she had studied pronunciation extensively in preparation for competing in English speech contests. An interesting detail that this student shared was that in her preparation for the speech contests, she had practiced her English pronunciation with her Japanese English teacher rather than with an ALT or native speaker. This information was very encouraging not only because this student had the most accurate pronunciation across each of the areas of acoustic analysis but also because it indicates that there are Japanese teachers who are concerned with teaching the skill of pronunciation, even if it is only for the sake of preparing students for speech competitions. Female participant 4 explained that she had never explicitly studied pronunciation prior to entering the university, and interestingly, she exhibited some of the least L1-like pronunciation samples. All male participants stated that they had not explicitly studied English pronunciation. In total, three out of the seven participants (roughly 43%) in this study had studied English pronunciation before entering the university.

Methods

Participants were recorded with a high-quality head mounted condenser microphone (Shure WBH53) onto a handheld digital recorder (Zoom H5) with a sampling rate of 44,100 Hz. Recordings were made in a quiet office environment to ensure minimal background noise. The recording script used to elicit the phonology of English was taken from the chapter by Peter Ladefoged (1999) on American English in the Handbook of the International Phonetic Association (see Appendix A). The rest of the data consisted of short phrases and sentences for eliciting the suprasegmental features of English.

Data Collection: Vowels and Consonants

To elicit all of the vowels and consonants of North American English, participants listened to recordings of the native Canadian English speaker while wearing high-quality closed back headphones, and repeated each word they heard after being prompted by a beep. This listen and repeat task design was done to separate the participants' actual abilities to perceive and produce English vowels from any influence that English orthography may have on their speech.

Data Collection: Sentences

In addition to the recordings from the speech perception experiment for vowels and consonants, participants were recorded reading a number of sentences for the purpose of quantifying the suprasegmental features of their speech (Appendix A).

Analysis of Vowels

Acoustically, vowels are distinguished by their differing patterns of acoustic energy that are concentrated around specific frequencies. These concentrations of acoustic energy are known as *formant frequencies* and result from the effect of the movements of the mouth's articulators on its resonance. Regarding the role of vowel formants in the perception of speech, the first three formants of a vowel are specifically important (Reetz & Jongman, 2020, p.p. 207-208).

Unlike the complex vowel system of English, Japanese is a language with a comparatively simple five-vowel system (Labrune, 2012). This type of vowel system is cross-linguistically common, with many languages such as Spanish having similar systems (two front vowels, one mid vowel, and two back vowels) (Maddieson, 2013). In contrast, languages like English, which are classified as having large vowel inventories (7 to 14 vowels), are comparatively rarer (Maddieson, 2013). As a result, English learners from language backgrounds with comparatively simpler vowel systems tend to struggle with acquiring accurate English vowel phonemes. This is related to the theoretical concept that speakers' perceptual abilities are shaped by the phonetic categories of their native languages, and as a result of this process, listeners tend to hear foreign speech sounds in terms of native sounds (Johnson, 2012, p. 107).

In order to investigate the acoustic properties of the Japanese participants' vowels, the first two formant frequencies of each vowel were measured at their approximate midpoint in HVD contexts (Appendix A) using a plugin called Fast Track (Barreda, 2021) for the phonetic analysis program Praat (Boersma & Weenink, 2022). Measurements at the midpoint of each vowel were taken to minimize the influence from coarticulation which occurs due to transitions between different segments. The vowel formant data were used to create two-dimensional vowel plots using the program RStudio (RStudio Team, 2022). These vowel plots are often referred to as *vowel spaces*. It is important

to note that these vowel spaces represent acoustic or auditory properties rather than articulatory ones (Reetz & Jongman, 2020, p. 192). This is because without utilizing imaging technology, we cannot make any definite claims about what the tongue is doing inside the mouth or its effect on its resonance due to anatomical or idiosyncratic differences from speaker to speaker.

Figure 1

Vowel Plots of Japanese Male Participants (left) and Female Participants (Right)(Raw Data)

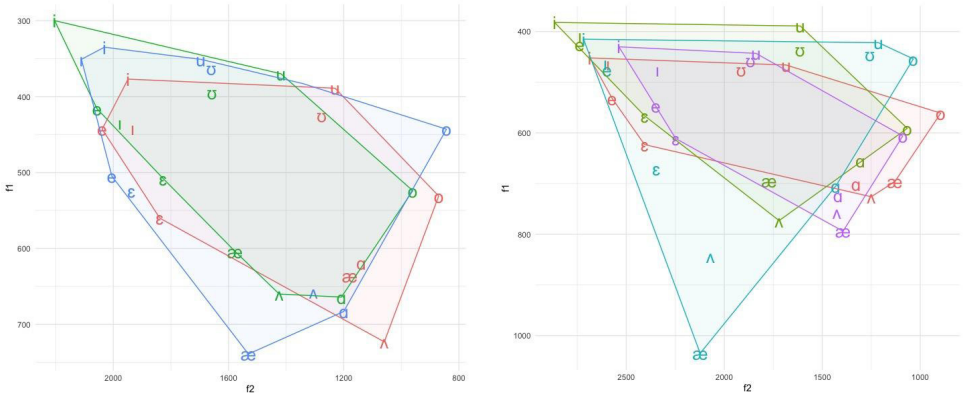
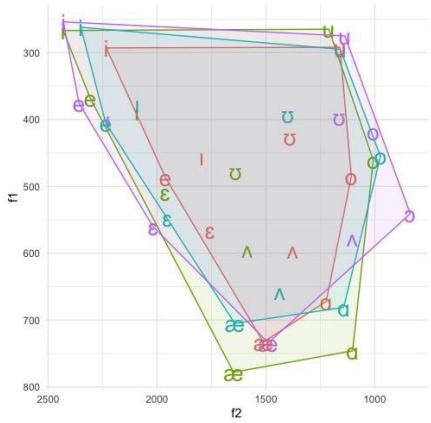


Figure 1 shows plots of the vowel spaces of the Japanese male and female participants with individual participants indicated by color. Aside from the vowel space of female participant 3 (aqua-colored participant on the right plot), who was mentioned previously and happens to be the speaker who exhibited the most L1-like pronunciation, the vowel spaces of the Japanese participants appear to have a compressed shape when compared to the native vowel spaces. In these vowel space plots, the y-axis (F_1 formant frequency) roughly corresponds with the height of the tongue; that is, low F_1 frequencies roughly correlate with close vowels and high F_1 with open vowels. The x-axis (F_2 formant frequency) corresponds with the backness or horizontal movement of the tongue, with high F_2 frequencies being observed in front vowels and low F_2 in back vowels. This could indicate that the Japanese speakers are not moving their tongues in configurations that approximate L1 vowel targets. However, as mentioned previously, this is not evidence of any articulatory realities but rather acoustic ones.

Figure 2

Vowel Plot of Native English Speakers (Raw Data)

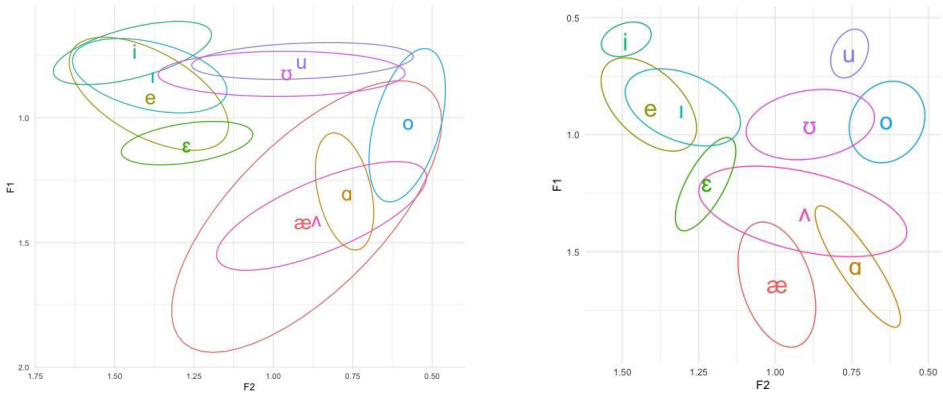


The plots in Figure 2 show the raw formant data, and they have been plotted separately because of formant differences that occur between female and male speakers. In order to compare the vowels of speakers of different biological sexes or dialects, vowel formant measurements must be transformed using the phonetic process of normalization (Barreda, & Nearey, 2018). There are numerous formulas for vowel normalization, but for this study the individual log-mean method of vowel normalization was chosen due to the benefits it provides (Nearey, 1978).

Figure 3 displays a simplified plot of all speakers' normalized vowels. The ellipses were calculated in RStudio and indicate the mean area that each vowel occupies in the perceptual space. There are several things to be said about the Japanese participants' vowel space. First, there are many ellipses that overlap with each other. This indicates that overlapping vowels are not accurately produced by the Japanese speakers. Second, the vowels that overlap the most are the ones that do not exist in Japanese, while the native Japanese vowels /i e a u o/ appear to all be confined to distinct areas. However, some overlap is observed for the /i/ and /e/ vowels, which may have resulted due to participants repeating L1 English stimuli rather than L1 tokens.

Figure 3

Summary of English Vowel Spaces of Japanese Speakers (left) and Native English Speakers (right)



Regarding the native vowel space, because the British English speaker does not pronounce 'hod' with an /ɑ/ vowel but rather a /ɔ/ vowel, his data for /ɑ/ was excluded from the analysis. Also, while some overlap is observed in the native vowel plot, this overlap could be a result of grouping different dialects of English together. Nevertheless, the native vowel space appears to be more organized than the Japanese one.

Analysis of Consonants

While recordings of all of the consonants of English were made for this study, due to the length constraints of this paper, the discussion of consonants will only focus on the *stop consonants* of English /b p t d k g/

Voice Onset Time (VOT)

The phonetic property of *voice onset time* (VOT) is a convenient measure for distinguishing between *voiced* and *voiceless* stop consonants. Cross-linguistically stop consonants tend to fall into three categories: voiced, voiceless, and *voiceless aspirated*. While some languages such as Thai have a three-way contrast (voiced, voiceless, voiceless aspirated), a majority of languages feature a two-way contrast (Nasukawa, 2005). VOT is a measurement of the number of milliseconds that transpire between the release of a stop burst and the initiation of vocal fold oscillation. Voiced stops exhibit a negative VOT. In contrast, voiceless stops exhibit a positive near zero VOT, and a voiceless aspirated stop has a longer positive VOT, which is generally around 30 ms (Johnson, 2012, p. 101)

Truly voiced stops are when phonation begins before the release of the stop resulting in a negative VOT, i.e., vocal fold oscillation commences before the stop is released. It has been said that rather than a truly voiced/voiceless distinction between stops, English exhibits a voiceless aspirated /p t k/ and voiceless unaspirated /b d g/ contrast rather than a voiced-voiceless one (Nasukawa, 2005). This, like many phonetic features of language, is not always descriptive of all varieties of English, with some sources noting that certain varieties of English exhibit stop consonants with negative VOTs (Hunnicutt & Morris, 2016). Likewise, this phenomenon was evident in the native speaker data collected for the voiced stops /b d g/, with the British English speaker producing no negative VOTs, the American speakers producing all negative VOTs, and the Canadian speaker producing two stops with negative VOTs /b g/ and one without /d/.

Aside from these inconsistencies, the two most significant differences between the stop consonants elicited from the Japanese participants and those from the native speakers is first that even though there was a range of negative and positive VOTs for the native English speakers, the average VOT for the group as a whole was negative. The second notable difference is that the differences between voiced and voiceless stop consonants (voiced length – voiceless length in milliseconds) is quite pronounced, with the Japanese group having an average difference of 39 milliseconds compared with the English difference of 139 milliseconds. This means that on average the difference between voiced and voiceless stop consonants between the native speakers and Japanese participants was 100 milliseconds. One way that this data can be interpreted is that this difference in VOTs indicates that it is likely that native English speakers, or speakers of any language with longer distinctions between voiced and voiceless stops for that matter, may have trouble perceiving the difference between voiced and voiceless stops in the English speech of native Japanese speakers. Table 1 presents a summary of the VOT data for all participants.

Table 1

A Summary of VOTs for All Participants

Japanese (n=7)			English (n=3)		
Voiced Average	Voiceless Average	Difference	Voiced English	Voiceless English	Difference
28 ms	67 ms	39 ms	-72 ms	67 ms	139 ms

Note. All times were rounded to the nearest tenth of a millisecond

One unexpected result that was observed from the listen and repeat style of data collection for this study is the apparent misperception of the English word-initial voiceless bilabial stop /p/. After hearing the word 'pie' [pʰaɪ], three of the seven Japanese participants said the word 'high [haɪ] instead. This is likely related to the lack of aspiration for the phoneme /p/ in Japanese, and the acoustic similarity between the aspiration of the /p/ and the frication of the glottal fricative /h/ could explain why some students incorrectly perceived this phoneme. This finding supports a case for instruction focused on contrasting the English /p/ and /h/ sounds, since it is likely that some Japanese learners confuse the two sounds.

Suprasegmental Analysis

Participants were recorded reading a total of six sentences in order to investigate the suprasegmental features of the Japanese participants' speech. This data collection was conducted primarily to analyze the *isochronous* stress patterns of the participants' speech. This feature of fluent English speech is often referred to as *sentence stress* in which stressed and unstressed words or syllables are combined to form a rhythmic pattern across whole utterances. Since English is a language which utilizes stress, syllables in polysyllabic words are pronounced with different levels of prominence. This variation between stressed and unstressed syllables contributes to the overall phonetic quality of English and its *isochrony*. It has been argued by a number of researchers that alterations to the stress timing of L2 accented speech, rather than the mastery of segments, should receive a higher degree of attention in EFL and ESL pedagogy, particularly for pronunciation instruction to L1 Japanese speakers (Nakashima, 2006; Nakamura, 2010; Koike, 2014). This is because it is reasoned that the stress-timing of English is a more salient quality of intelligible English speech than vowel or consonant segments (Kang, et al., 2010; Hahn, 2004, Munro & Derwing, 1995). Specific research regarding Japanese L2 English learners has identified that they often do not utilize stress, nor do they reduce unstressed words, which is a major contributor to their perceived foreign accent and overall intelligibility. Since Japanese is a *mora-timed* language, each syllable of an utterance features similar intensities and lengths with pitch being the primary contrastive suprasegmental feature (Labrune, 2012). The following section presents an acoustic analysis of the stress patterns of the seven Japanese participants and compares them with the four native English speakers.

Phonetic Analysis of Sentence Stress

The three major phonetic correlates of English stress are *duration*, *intensity*, and *pitch* (Rogerson-Revell, 2011). In order to measure these features, Japanese L1 participants were recorded reading several sentences ranging from three to 11 words. Duration, intensity, and pitch for each vowel were measured in Praat, and the means for duration, intensity and pitch for all vowels occurring in stressed syllables were calculated and compared with those which occurred in unstressed syllables across all utterances for each participant. (Refer to Appendix A for the stimuli).

Because duration can vary according to a participant's rate of speech, the analysis of the difference between vowel duration in stressed syllables in contrast to unstressed syllables will primarily focus on the mean ratios of participants; that is, mean stressed vowel duration divided by mean unstressed vowel duration. The measurement of intensity also faces a similar challenge, since amplitude can vary greatly according to how loudly each participant speaks or if the gain settings of the recording equipment happen to be adjusted. Again, the same method used for the analysis of duration was applied to the analysis of intensity. In this way, the expected between-speaker differences of duration and intensity can be controlled for. Means and ratios were also calculated for the entire group of L1 Japanese participants. Table 2 presents the means, differences and ratios for the vowel durations of all participants with all times rounded to the nearest millisecond.

Table 2

Average Vowel Durations, Differences and Stressed/ Unstressed Ratios

Participant	Stressed Mean	Unstressed Mean	Difference	Ratio
Male 1	156 ms	157 ms	-1 ms	0.99
Male 2	150 ms	129 ms	21 ms	1.16
Male 3	143 ms	152 ms	-9 ms	0.94
Female 1	148 ms	131 ms	17 ms	1.13
Female 2	151 ms	157 ms	-6 ms	0.96
Female 3	133 ms	109 ms	24 ms	1.22
Female 4	163 ms	171 ms	-8 ms	0.96
Native 1 (UK)	137 ms	82 ms	55 ms	1.68
Native 2 (CAN)	141 ms	108 ms	33 ms	1.30
Native 3 (US)	167 ms	129 ms	38 ms	1.30
Native 4 (US)	146 ms	97 ms	49 ms	1.5

Table 3

Summary of Duration Statistics

Japanese (n = 7)				English (n = 3)			
Stressed Mean	Unstressed Mean	Difference	Ratio	Stressed Mean	Unstressed mean	Difference	Ratio
149 ms	144 ms	5 ms	1.04	148 ms	104 ms	44 ms	1.42

When examining the data presented in Table 2, four of the seven Japanese participants had ratios within 0.06 of 1, which means that the duration of stressed vowel segments and unstressed vowel segments would be indistinguishable from each other. In fact, various perceptual studies have found that a difference of less than 20 ms is imperceptible to human listeners (Pisoni, 1977; Pastore, & Farrington, 1996), so factoring this into the analysis, six out of seven participants are not utilizing perceptible durational contrasts between stressed vowels and unstressed vowels. The only Japanese participant that produces a duration contrast that is likely to be perceptible is female participant 3 with a difference of 24 ms and a ratio of 1.22. This is the same participant discussed earlier who had the most accurate pronunciation out of the group of Japanese participants. In contrast, the native speakers all have a ratio of 1.3 or more, which means that stressed vowels are on average 30% longer than unstressed vowels. If we look at the differences, we also notice that they are all at least 30 ms, which would be a perceptible contrast for human listeners.

When measurements from each participant group are averaged together (Table 3), this difference becomes even more striking, with native speakers having stressed vowels that are 42% longer on average compared with the 4% of the Japanese speakers. As was mentioned earlier, an average difference of 5 ms in duration would not be perceptible by a human listener, thus as a population, the Japanese speakers do not utilize duration sufficiently to contrast between stressed and unstressed vowels.

Intensity

Table 4 shows the mean statistics for vowel intensity rounded to the nearest decibel. This acoustic correlate of stress appears to be utilized more than duration by the Japanese participants, with almost all speakers exhibiting a difference of 3 decibels or more. This is significant because due to the logarithmic nature of the decibel as a unit of measurement, an increase of 3db is equivalent to a doubling of sound energy (Plack, 2018, p.14). In other words, if stressed segments are being produced 3db louder than unstressed segments by the Japanese participants then it may be likely that they would be perceived in contrast with the unstressed syllables. However, due to the nature of the human auditory system, it is difficult to quantify how loud certain sounds are to individual listeners because the perception of 'loudness' is subjective.

Table 4

Mean Vowel Intensities, Differences and Stressed/Unstressed Ratios

Participant	Stressed Mean	Unstressed Mean	Difference	Ratio
Male 1	68 dB	65 dB	3 dB	1.05
Male 2	77 dB	73 dB	4 dB	1.06
Male 3	68 dB	65 dB	3 dB	1.06
Female 1	68 dB	65 dB	3 dB	1.04
Female 2	67 dB	65 dB	2 dB	1.04
Female 3	66 dB	63 dB	3 dB	1.05
Female 4	68 dB	67 dB	1 dB	1.01
Native 1 (UK)	71 dB	64 dB	7 dB	1.10
Native 2 (CAN)	79 dB	74 dB	5 dB	1.07
Native 3 (US)	72 dB	73 dB	-1 dB	0.98
Native 4 (US)	64 dB	60 dB	4 dB	1.07

Table 5

Summary of Intensity Statistics

Japanese (n = 7)				English (n = 3)			
Stressed Mean	Unstressed Mean	Difference	Ratio	Stressed Mean	Unstressed Mean	Difference	Ratio
69 dB	66 dB	3 dB	1.04	71 dB	68 dB	3 dB	1.05

Pitch

Table 6 shows the measurements of pitch rounded to the nearest hertz. This acoustic correlate of stress is probably the most difficult to analyze for several reasons. While hertz is a measurement of how many cycles of a sound occur each second, there is not a linear relationship between the number of cycles and the perceived pitch. This is because the human auditory system is more sensitive to small changes at the low end of the audible range and less so for higher frequencies (Johnson, 2012, p.p. 88-90). This can be illustrated by the frequency intervals between the keys on a piano, in that the higher the octave, the larger the intervals are between notes. Since female speakers tend to have higher pitched voices, it should be expected for them to vary their pitch more than male speakers would. Evidence for this can be noted in table 6, with two of the female participants varying their pitch between stressed and unstressed vowels to a greater degree than the native speakers do (25 & 21 Hz compared with -4 to 12 Hz). Again, due to these inconsistencies, a more thorough analysis of pitch would need to be conducted, which goes beyond the scope of the current project. Table 7 summarizes the pitch statistics, and we can observe that the population ratios are different between the Japanese and native English speakers.

Table 6

Mean Vowel Pitches, Differences and Stressed/Unstressed Ratios

Participant	Stressed Mean	Unstressed Mean	Difference	Ratio
Male 1	102 Hz	100 Hz	2 Hz	1.01
Male 2	122 Hz	127 Hz	-5 Hz	0.96
Male 3	128 Hz	126 Hz	2 Hz	1.02
Female 1	207 Hz	202 Hz	5 Hz	1.03
Female 2	208 Hz	187 Hz	21 Hz	1.11
Female 3	247 Hz	222 Hz	25 Hz	1.11
Female 4	211 Hz	206 Hz	5 Hz	1.03
Native 1 (UK)	102 Hz	92 Hz	10 Hz	1.12
Native 2 (CAN)	137 Hz	125 Hz	12 Hz	1.10
Native 3 (US)	122 Hz	128 Hz	-6 Hz	0.95
Native 4 (US)	102 Hz	101 Hz	1 Hz	1.01

Table 7
Summary of Pitch Statistics

Japanese (n = 7)				English (n = 3)			
Stressed Mean	Unstressed Mean	Difference	Ratio	Stressed Mean	Unstressed Mean	Difference	Ratio
175 Hz	167 Hz	8 Hz	1.04	116 Hz	112 Hz	4 Hz	1.03

Limitations

While this study sought to investigate the acoustic qualities of Japanese L2 English and generalize findings to a broader population, there were a number of limitations to the current study. The first is that this study was conducted with a relatively small sample size (n = 7), so findings could be made weightier with a larger sample size. Other limitations relate to the way sentences were recorded for the analysis of stress. Because recordings were made from read passages, it is not clear how this may have contributed to the participants' control of English stress and rhythm in spontaneous speech.

Another issue that was not factored into the analysis of duration for this study was the difference in length between monophthongal English vowels which occur as a consequence of tense and lax distinctions. Also, diphthongal vowels were not treated differently from monophthongal ones. Pitch encounters the same kind of issue since certain vowel segments have higher pitches than others, and the phrasal intonation pattern that a vowel occurs in can also affect its pitch. Finally, the number of stressed and unstressed syllables were not balanced across all sentences that were recorded, with a total of 21 stressed syllables and 20 unstressed ones. Nevertheless, the findings of this study offer some compelling evidence of quantifiable phonetic differences in the English L2 speech of Japanese learners.

Conclusion

While a thorough discussion of all the phonetic properties of Japanese L2 English could possibly fill an entire book, this study presented a general overview of the pronunciation of Japanese L2 English across the dimensions of vowels, stop consonants, and sentence stress, with the aim of providing a reference for instructors who work with or might work with Japanese learners in the future, or for any party interested in Japanese L2 English.

In terms of vowels, overall, the Japanese speakers seem to individually exhibit smaller vowel spaces with a limited range of F_1 frequencies when compared with the vowels of native speakers, and as a group, exhibit many vowels that overlap in the perceptual space. While it is impossible for this study to inform us of whether or not the participants were able to perceive all of the different qualities of the English vowels, it *is* able to inform us of how their perceptions of the English stimuli were produced by their vocal tracts.

Stop consonants produced by the Japanese participants did not exhibit any negative VOTs for the English voiced stops, and the VOTs of voiceless stops were significantly shorter than those of the English speakers, which indicates that it is likely that native English speakers might misperceive the difference between voiced and voiceless word initial stops when spoken by native Japanese speakers. It is also apparent that the English voiceless bilabial stop /p/ may be easily confused with the voiceless glottal fricative /h/ by native Japanese speakers.

Further research on this point might consist of a speech perception experiment where English speakers from a variety of L1 backgrounds are made to listen to recordings of Japanese speakers saying voiced and voiceless stop consonant contrasts in ambiguous contexts to identify if the English speakers would have difficulty in understanding the contrasts. Also, a similar experiment could be conducted that investigates how often /p/ is perceived as /h/ by Japanese speakers and if an increase or reduction in aspiration has any effect on the participants' judgments.

Regarding the phonetic correlates of stress, it seems that intensity is used by the Japanese speakers more than length to contrast between stressed and unstressed syllables in a sentence, but the differences in intensities were near the perceptible threshold for human listeners. In contrast, the native speakers that were analyzed used duration the most to signal contrasts between stressed and unstressed parts of speech. Pitch was also analyzed but is probably the least straightforward phonetic correlate of stress due to biological differences and the non-linear nature of the human perception of pitch.

The most interesting statistic for the Japanese group was that ratios for all three correlates of stress across all participants were 1.04, which is strong evidence that Japanese speakers do not treat stressed segments differently from unstressed segments. Based on these data, it is likely that Japanese learners could greatly benefit from pronunciation instruction that focuses on word and sentence stress and how those features contribute to the timing of fluent English. To these ends, it might be fruitful to further investigate how suprasegmental-focused pronunciation instruction might impact learners' use of stress in spontaneous speech.

Bio Data

Anthony M. Diaz is currently a PhD student in Linguistics at the University of California, Davis. He has extensive experience teaching ESL and EFL in a variety of institutions and to a broad range of students in the United States and Japan. His research focuses on phonetics, speech perception, and pronunciation teaching. antdiaz@ucdavis.edu

References

Baker, A. A. (2011). Discourse prosody and teachers' stated beliefs and practices. *TESOL Journal*, 2, 263-292.

Barreda, S. (2021). Fast track: fast (nearly) automatic formant-tracking using Praat. *Linguistics Vanguard*, 7(1)

Barreda, S., & Nearey, T. M. (2018). A regression approach to vowel normalization for missing and unbalanced data. *The Journal of the Acoustical Society of America*, *144*(1), 500–520. https://doi.org/10.1121/1.5047742

Boersma, P., & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.2.18, retrieved 2 September 2022 from http://www.praat.org/

Celce-Murcia, M., Brinton, D., & Goodwin, J. M. (2010). *Teaching pronunciation: A course book and reference guide* (2nd ed.). Cambridge University Press.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379. https://doi.org/10.2307/3588486

Gallacher, L. (2004). English sentence stress. Retrieved December 08, 2016, from https://www.teachingenglish.org.uk/article/english-sentence-stress

Hahn, L. D. (2004). Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly*, 38(2), 201. doi:10.2307/3588378

Hunnicutt, L., & Morris, P. A. (2016). *Prevoicing and aspiration in Southern American English* (University of Pennsylvania Working Papers in Linguistics: Vol. 22 : Iss. 1 , Article 24). https://repository.upenn.edu/pwpl/vol22/iss1/24

Jenkins, J. (2000). *The phonology of English as an international language: New models, new norms, new goals*. Oxford University Press.

Johnson, K. (2012). *Acoustic and auditory phonetics*. Wiley-Blackwell.

Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *The Modern Language Journal*, 94(4), 554– 566. doi: https://doi.org/10.1111/j.1540- 4781.2010.01091.x

Koike, Y. (2014). Explicit pronunciation instruction: Teaching suprasegmentals to Japanese learners of English. In N. Sonda & A. Krause (Eds.), *JALT2013 Conference Proceedings*, Japan, 361-374. https://jalt-publications.org/sites/default/files/pdf-article/jalt2013_036.pdf

Labrune, L. (2012). *The phonology of Japanese*. Oxford University Press.

Ladefoged, P. (1999). American English. In *Handbook of the International Phonetic Association: A guide of the use of the international phonetic alphabet*. Cambridge.

Maddieson, I. (2013). *Vowel quality inventories*. Wals online - chapter vowel quality inventories. Retrieved January 6, 2023, from https://wals.info/chapter/2

Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97. doi:10.1111/j.1467-1770.1995.tb00963.x

Nakamura, S. (2010). Analysis of relationship between duration characteristics and subjective evaluation of English speech by Japanese learners with regard to contrast of the stressed to the unstressed. *Pan-Pacific Association of Applied Linguistics*, 14(1), 1-14.

Nakashima, T., (2006). Intelligibility, suprasegmentals, and L2 pronunciation instruction for EFL Japanese learners. 福岡教育大学紀要 (*Fukuoka University of Education Journal*), 55(1), 27-42

Nasukawa, K. (2005). The representation of laryngeal-source contrasts in Japanese. In J. van de Weijer, K. Nanjo, & T. Nishihara (Eds.), *Voicing in Japanese.* essay, Mouton de Gruyter.

Nearey, T. M. (1978). *Phonetic Feature Systems for Vowels.* Indiana University Linguistics Club.

Pastore, R. E., & Farrington, S. M. (1996). Measuring the difference limen for identification of order of onset for complex auditory stimuli. *Perception & Psychophysics*, *58*(4), 510–526. https://doi.org/10.3758/bf03213087

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *The Journal of the Acoustical Society of America*, *61*(5), 1352–1361. https://doi.org/10.1121/1.381409

Plack, C. J. (2018). *The sense of hearing*. Routledge.

Reetz, H., & Jongman, A. (2020). *Phonetics: Transcription, production, acoustics, and perception*. Wiley.

Rogerson-Revell, P. (2011). English phonology and pronunciation teaching. Bloomsbury Academic.

RStudio Team (2022). RStudio: Integrated development environment for R. RStudio, PBC, Boston, MA URL http://www.rstudio.com/.

Yates, J. (2005). Pronounce it perfectly in English (2nd ed.). Barron's Educational Series.

Appendix A

Recording script

Words for eliciting consonants

/p/ Pie	/b/ Buy	/m/ My	/f/ Fie	/v/ Vie	/w/ Why	/t/ Tie	/d/ Die	/n/ Nigh	/θ/ Thigh
/ð/ Thy	/s/ Sigh	/z/ Zoo	/ɹ/ Rye	/l/ Lie	/k/ Kite	/g/ Guy	/h/ Hang	/h/ High	/ʧ/ Chin
/ʤ/ Gin	/ʃ/ Shy	/ʒ/ Azure	/j/ You						

Words for eliciting vowels

i	ɪ	e	ɛ	æ	ɑ	o	ʊ	u	ʌ
Heed	Hid	Hayed	Head	Had	Hod	Hoed	hood	who'd	hud

Sentences for eliciting word and sentence stress

1. ***Time*** is ***mo***ney (4 syllables: 2 stressed, 2 unstressed)
2. A ***pe***nny **saved** is a ***pe***nny ***earned*** (8 syllables: 4 stressed, 4 unstressed)
3. ***Be***tter ***late*** than ***ne***ver (6 syllables: 3 stressed, 3 unstressed)
4. The ***boy ate*** an ***a***pple. (6 syllables: 3 stressed, 3 unstressed)
5. I ***asked*** you to ***buy*** me a ***bunch*** of ***red rose***s. (12 syllables: 5 stressed, 7 unstressed)
6. He ***ate*** a ***big grape***. (5 syllables: 3 stressed, 2 unstressed)

This script was adapted from:
(Celce-Murcia, M., et al., (2010), (Yates, J., 2005), (Gallacher, L., 2004), (Ladefoged, 1999)