Visualizing Contrastive Prosody in Second Language English*

Yuji Shuhama

1 Introduction

This short article reports a survey of English learners to investigate the prosodic features of their speech. Prosody can be interpreted as the melodic aspect of speech.[1)] For example, speakers can convey various emotions and intentions when they utter the same sentence "You're taking a linguistics class this semester" by adjusting speech rate and intonation. Among such prosodic features, this report primarily presents pitch ranges to consider their relationship with a contrastive structure.

I believe that prosodic features, including pitch control, are vital for humanistic, face-to-face interactions because communication research shows that vocal and non-verbal cues convey speakers' messages to listeners more effectively than verbal cues (Mehrabian 1981). This should be true even when communicating in foreign languages; however, acquiring the target language prosody appears more challenging than the vocabulary and grammar for second/foreign language learners when the author has been teaching them in college for more than 10 years.

Some course books designed for communicative purposes address the mechanism of prosody with a link to downloadable audio files in a limited space. Acquiring prosodic features is difficult because they cannot be observed easily by learners since their characteristics are not typically presented in a visually intuitive manner. For example, the following notation (1), cited from an old but valuable source, allows one to easily understand how an utterance can be pronounced with an emotional accent in the middle (i.e., *tér-ri-bly*). The pitch height is directly represented by the line height of (a syllable of) each expression.

(1)
```
                                  tér
        wánt you to know
I                             how
                                      ribly sad we are for you!
```
(Bolinger 1986: 83)

For prosodic visualization, a survey was conducted to obtain second language learners' speech data and analyze their pitch changes. After reviewing related studies (see Section 2), an outline of the prosody survey is described in Section 3. Section 4 presents the data from a native speaker's model speech and learners' recordings with visualized pitch changes. Finally, Section 5 summarizes the report and discusses future research directions.

2 Related Studies

Behrens (2018) describes prosody as a melody of speech characterized by multiple features including stress, emphasis, intonation, pitch, loudness, and speech rate. These features are important, particularly in spoken communication, in expressing speakers' emotional information such as willingness, disappointment, and surprise. Furthermore, they convey linguistic information, for example, to differentiate *éxport*, *décrease*, *récord* as nouns, and *expórt*, *decréase*, *recórd* as verbs through accent placement.

To a commonly posed question regarding prosody, 'How can teachers help students sound less robotic during oral presentations?' Behrens answers from acoustic and psychological perspectives that fear and nervousness in public speaking cause tensing of the muscles surrounding the vocal folds, thus resulting in a monotone, robotic-sounding voice. Her comment suggests that instructors should consider calming students' nerves when planning interview-style exams and speech recordings.

Regarding theoretical studies of second language pitch, Archibald and Croteau (2021) investigated whether non-native speakers of Japanese were able to acquire a natural pitch pattern for *wh*-questions. Sixteen proficient English speakers learning Japanese were instructed to read aloud *wh*-sentences as in (2), which have single or multiple *wh*-phrases and a question particle. As reported in the literature, Archibald and Croteau discovered a pitch boost in *wh*-words and a final rising intonation in the control group of native speakers.

(2) Naoya-wa nani-o nomiya-de nonda no ?
Naoya-TOP what-ACC bar-LOC drank +Q
"What did Naoya drink at the bar?"

(Archibald 2024: 188)

Interestingly, Archibald and Croteau discovered that the same pitch pattern was observed in second language learners' data. Figure 1 illustrates a pitch boost on *nani* "what," a low-pitch plateau, and a final rising tone. With no remarkable pitch change or prosodic break in between, this indicates the prosodic contiguity of *wh*-words and question particles in Japanese and further supports Richards' (2016) contiguity theory.

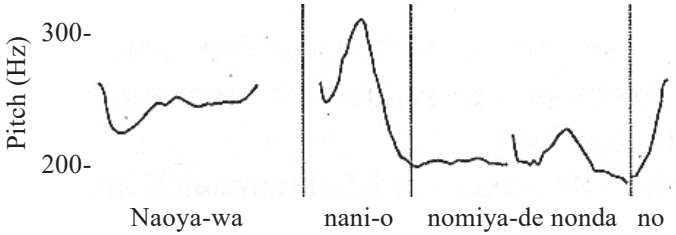

Figure 1. Visualization of pitch patterns of non-native speaker
(Archibald 2024: 191)

Their study shows that non-native speakers can acquire natural pitch patterns. This suggests the existence of pitch and intonation in Japanese. Although Japanese speech is typically characterized by continuous equal-timed syllables along a flat flow of intonation (e.g., Shizuka 2019), Archibald and Croteau discovered clear pitch changes in *wh*-sentences, regardless of whether they were interrogative utterances. Since Japanese speech does have pitch boosts and intonation changes, evaluators should not merely associate learners' unnatural prosody in English with the tendency for flat intonation in their first language, Japanese.

3 Methods

3.1 Participants

Seventeen participants participated in this study. All participants were in their first year of college in Japan, aged–18 and 19 years old. They spoke Japanese as their native language and continued to learn English for at least eight years in an EFL classroom setting. One week before the spring semester began in April, 2024, they had taken the visualizing English language competence (VELC) test, which is a 70-min measurement of

listening and reading skills, such that they can be assigned to the appropriate classes based on their proficiency levels.[2)] Table 1 shows their age and reading/listening scores. The mean values of the reading and listening scores were 521.9 and 541.4, respectively, which correspond approximately to 230/270 in TOEIC reading/listening and indicates upper A2 English proficiency in the CEFR scale.

Table 1. Age and reading/listening proficiency of participants

	Mean	**SD**	**Min.**	**Max.**
Age	18.8	0.22	18.3	19.1
VELC (R)	521.9	44.24	412	592
VELC (L)	541.4	39.46	428	606

(n=17)

3.2 Procedures

Nineteen students were initially enrolled in the author's reading-based skill-integrated English class based on their VELC score. However, the enrollment of two students was canceled; thus, 17 students underwent written and oral exams for a final evaluation to complete the course. Since the class required the application of reading to other skills, the students were informed in advance that they must sit for not only a written exam but also an interview test using reading materials from the course book.

After regular weekly meetings for three months from April to June 2024 and two-week oral review sessions, an interview test was held in the middle of July. For the test material, the following excerpt extracted from a section titled "Innovating the future" was selected:

> **Turning the L.A. streets into food**
>
> For Ron Finley, gardening isn't just about growing plants, it's about growing people. He lives in South Los Angeles, U.S., where quality fresh food can be hard to find. In 2010, he planted fruit and vegetables on some land between his house and the street. When the local authorities said what he was doing was illegal, he got the law changed. Ten years later, he has helped to create community gardens in unused spaces all over the city.
>
> (*Voices* 5, Unit 3, p. 37)

Before the interview, all students understood the contents of the passage and learned the power of a citizen in influencing the community. Additionally, the students were advised by their instructor (the author) to practice reading the passage aloud in advance and listen to an audio file of the passage narrated by a female speaker of American English uploaded to the Google Classroom course site.

The interview test comprised three aspects: pronunciation, skimming, and opinion.[3)] For pronunciation, each student was instructed to read the passage silently, and then read it aloud. Their readings were video and audio recorded, and the recorded sound files in MP3 format were analyzed for prosodic features using *Praat* (version 6.1.50). The students' pronunciation was graded in terms of clarity and fluency, and for comparison, the data were segregated into better speech samples and others, labeled A and B, respectively.

4 Data Analysis

4.1 Model speech by native speaker

The first sentence of the passage, "… gardening isn't just about growing plants, it's about growing people," has a negative construction "not just *x*, (but) *y*" to contrast two pieces of information. Although we naturally say that people take care of plants in gardening, this topic tells us more about gardening than how it can change people for the better. Owing to this contrast, *plants* and *people* are expected to be pronounced remarkably.

Figure 2 illustrates the waveform and pitch of a sentence pronounced by a female American English speaker. As expected, the pitch curve showed a sharp rise and fall for contrasting words, *plants* and *people*. More importantly, a clear V-shaped rise–fall–rise on *isn't just* was observed. Based on the native speaker's model speech, the contrastive construction features the following pitch change in (3), as indicated by different line heights, similarly to (1).

The target sentence was read in 5.76 s and the entire passage in 41.5 s at a steady, moderate pace. Additionally, the remainder of the passage includes more words read with remarkable pitch changes: *fóod*, *hárd*, *find*; *végetables*, *lánd*, *hóuse*, *stréet*; *illégal*, *chánged*; *áll óver the cíty*. The passage sounds rhythmic to the author because of these words, which are

centered in prosodic units, such as *quality fresh fóod*, *fruit and végetables*, and *on some lánd*.

(3) Gardening is n't just about pla growing nts,
it's a bout grow ing peo ple

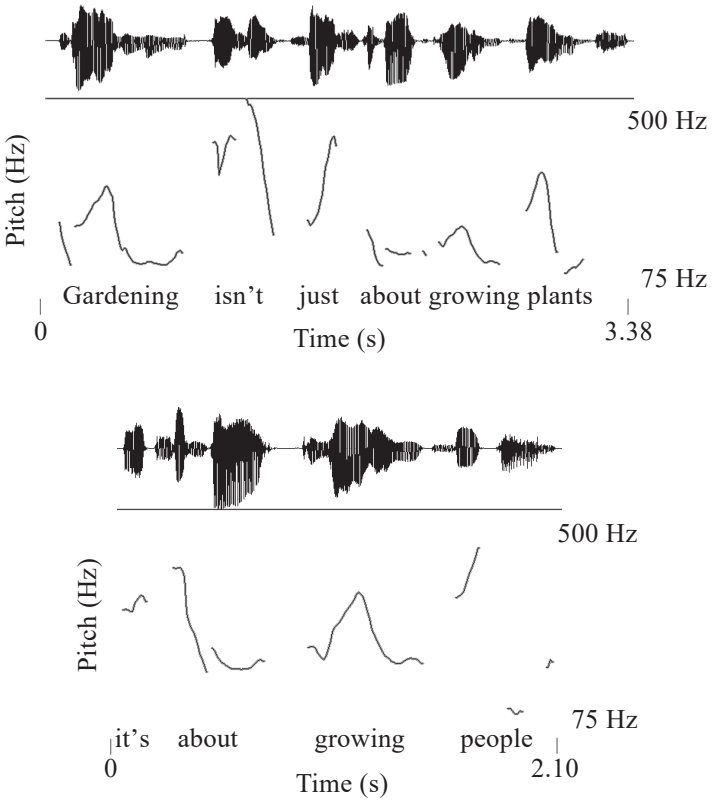

Figure 2. Waveform and pitch of "Gardening isn't just …"

4.2 A samples

Seventeen speech samples were grouped into A and B (nine and eight samples, respectively) based on clarity and fluency graded by the author. Clarity was graded based on the correct pronunciation of each word and the appropriate loudness, and fluency was graded based on the speech rate and natural flow, without unnecessary pauses. Each of the two features was rated on a scale of 1 to 5, with higher scores indicating better performance. Table 2 shows the clarity and fluency data of speech samples A and B, as well as those of all the samples.

Table 2. Graded clarity and fluency of all, A, and, B samples

	Mean	**SD**	**Min.**	**Max.**
All (n = 17)				
Clarity	3.6	0.77	2.0	5.0
Fluency	3.2	0.80	2.0	5.0
A samples (n = 9)				
Clarity	4.2	0.64	4.0	5.0
Fluency	3.8	0.62	3.0	5.0
B samples (n = 8)				
Clarity	2.8	0.57	2.0	3.0
Fluency	2.6	0.48	2.0	3.0

The A samples performed better than the B samples in terms of clarity and fluency. The clarity of the former was particularly good, and as indicated by the high average score (4.2 points), the passage was read almost correctly in the right volume. Figure 3 shows the acoustic data of the target contrastive sentence read by a female participant with a grade of 4.0 in terms of clarity and fluency. The pitch curve shows some rises and falls; however, the pitch range was significantly narrower than that of the model speech. In the A samples, the average pitch of the first clause, “Gardening isn’t … plants,” ranged from 87.5 to 216.6 Hz, which was significantly narrower than the model speaker’s pitch ranging from 142.8 to 497.5 Hz.[4)]

Example (4) is a visual representation of the target sentence read by the abovementioned participant. In the A samples, the contrast between *plants* and *people* tends to be slightly emphasized within the limited range of pitch

change. The sentence was read in 5.39 s on average and the entire passage in 40.9 s, which implies that the students in the A group can read out as smoothly as the model speaker. Learners at this proficiency level should be advised on how to use wider pitch ranges naturally and more expressively.

(4) Gardening isn't growing plants,
just about
it's about growing peo
ple

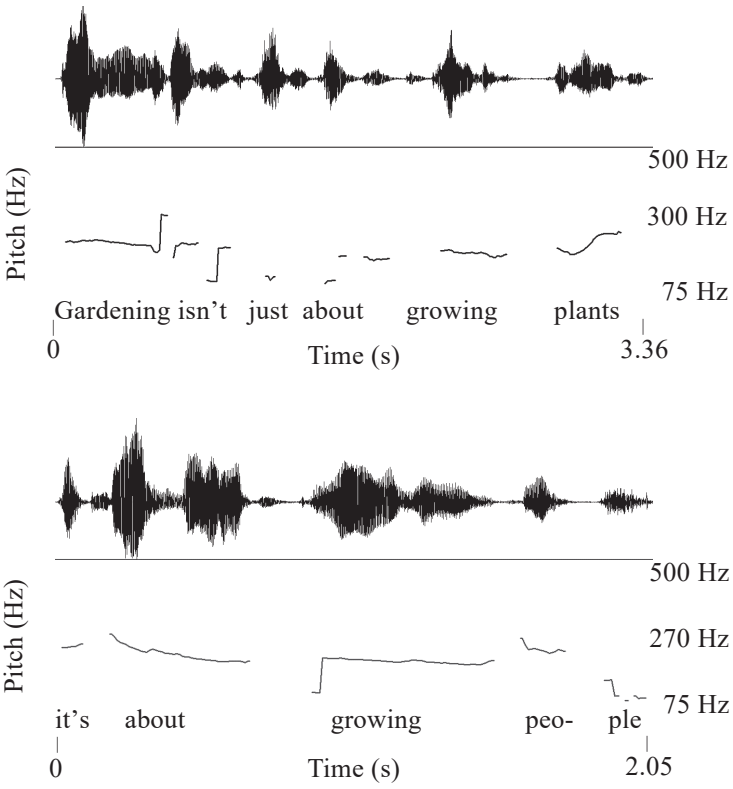

Figure 3. Speech data of female participant in A samples

4.3 B samples

The B samples were graded below 3.0 points (rated as “not so good”) on average in terms of clarity and fluency. The average fluency score was 2.6, which indicates that the passages were not read smoothly. In fact, the speeches were frequently interrupted by specific words that appeared difficult or impossible to pronounce and understand by the students in the B group: *quality*, *authorities*, *illegal*, *unused*, and unexpectedly, *Finley*.

A person’s family name *Finley* /fɪnli:/ was frequently mispronounced in various ways such as /faɪnli:/, /fi:nleɪ/ or as unnatural pronunciations after a short pause, which indicated the speakers’ uncertainty. Therefore, the passage was read as, “For Ron, Finley /faɪnli:/, gardening isn’t …” Owing to several pauses of hesitation in the B samples, the average speech rate was 44.8 s, which was 3.9 s longer than that of the A samples.

Regarding pitch, a flat intonation was generally observed in the B samples. For example, Figure 4 shows the acoustic data of the first clause of the target sentence pronounced by a female student who was graded 3.0 in terms of clarity and fluency.[5)] Here, the speaker’s pitch range was from 156.9 to 253.4 Hz. The average pitch of the entire target sentence in the B samples ranged from 113.3 to 191.3 Hz, which was much narrower than that in the A samples. This indicates that learners with lower oral skills could not use the pitch freely and expressively.

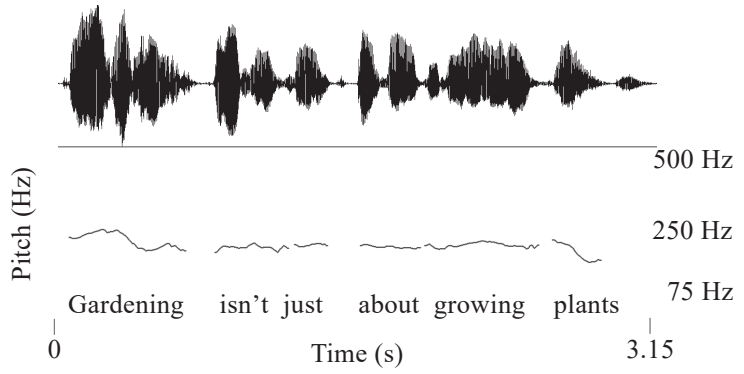

Figure 4. Speech data of female participant in B samples

In summary, the pitch data of the A and B samples show that the pitch range of upper A2 learners of English was narrower than that of the native speakers, and that their pitch-use ability varied. Learners with relatively high prosodic skills appeared to be able to manipulate pitch to express contrastive information, whereas those with low prosodic skills could neither control their pitch nor read aloud smoothly because of uncertain and mispronounced words.

5. Conclusion

The survey results of English learners' prosodic features, primarily pitch and intonation, are reported herein. Unlike native speakers' natural pitch control along with contrastive information, upper A2 learners appeared to have acquired limited prosodic skills to emphasize contrastive words by pitch if they were able to read aloud with some accuracy and fluency. Some A2 learners with low oral skills have yet to acquire pitch control, thus resulting in steady flat intonation.

One implication for better instruction of prosodic features is that it can be helpful for intermediate learners, such as the A-group students in this study, to visually present grammatical and prosodic information simultaneously. (5) presents an example where a contrastive construction "not just *x*, (but) *y*" and the location to boost pitch are presented simultaneously. This presentation method is simpler than that of (3) and allows learners to understand more easily the manner by which sound is related to sentence structures.

(5) Gardening is**n't just** about growing **pla-nts**, it's about growing **peo-ple**.

Because the present survey involved a limited number of second language learners addressing a specific construction, future studies should be conducted with more participants and a broader focus on prosodically prominent constructions. The focus-related pitch phenomena observed in Section 4 must be related to the *wh*-prosody in a second language, as investigated by Archibald and Croteau (2021) and Archibald (2024). Further investigations involving their phonological theory shall be conducted using experimental methods for theoretical contributions.

Notes

* This research was supported by Grants-in-Aid for Scientific Research (C) (JSPS KAKENHI, grant no. 23K00687).

1) Section 2 introduces prosody in more detail.

2) VELC is the abbreviation for visualizing English language competence. It is a standard English proficiency test available in Japan and has been developed specifically for Japanese university students based on language testing theories. For more details, please refer to the following website: https://www.velctest.org/whatsvelc/

3) After the pronunciation check, questions were posed to each student for skimming and reflection, such as “What did Ron do around his house?” and “What will you do if you have a space in the community garden?”

4) The average pitch of the second clause was narrower, i.e., 79.8–174.6 Hz compared with the model speaker’s pitch range of 85.9–425.4 Hz.

5) The acoustic data of the latter half of the target sentence were omitted because they showed a pitch curve similar to that presented in Figure 4.

References

Archibald, John. 2024. *Phonology in Multilingual Grammars: Representational Complexity and Linguistic Interfaces*. Oxford: Oxford University Press.

Archibald, John, and Nicole Croteau. 2021. Acquisition of L2 Japanese WH questions: Evidence of phonological contiguity and non-shallow structures. *Second Language Research* 37(4), 649-679.

Barber, Daniel, and Marek Kiczkowiak. 2022. *Voices*. American English, Level 5, Student Book with Online Practice. Cengage Learning.

Behrens, Susan J. 2018. *Understanding Language Use in the Classroom: Including Teaching Materials for College Educators*. Bristol: Multilingual Matters.

Boersma, Paul, and David Weenink. 2021. *Praat: Doing Phonetics by Computer*. An open computer program for speech analysis available at: https://www.fon.hum.uva.nl/praat/

Bolinger, Dwight. 1986. *Intonation and Its Parts: Melody in Spoken English*. Stanford University Press.

Mehrabian, Albert. 1981. *Silent Messages: Implicit Communication of Emotions and Attitudes*. Belmont, CA: Wadsworth.

Richards, Norvin. 2016. *Contiguity Theory*. Cambridge, MA: MIT Press.

Shizuka, Tetsuhito. 2019. *Pronunciation Textbook: the best for native Japanese speakers to learn English sounds and rhythm they are not good at*. Tokyo: TS-Planning. [in Japanese]