Acoustic analysis of the English pronunciation of Japanese high school teachers and university students

Ian Wilson, Junichi Fujinuma, Naoya Horiguchi & Kazuaki Yamauchi
Center for Language Research, University of Aizu, Japan

Objectives

▪ Provide benchmark acoustic data on the L2 English pronunciation of Japanese junior and senior high school teachers and university students
▪ Compare acoustic measures of pronunciation of university students before and after a 14-week EFL pronunciation course
▪ Compare acoustic measures of English pronunciation of junior and senior high school EFL teachers with university-level EFL students and with native English speakers (all reading the same paragraph)

Background

As Japan attempts to meet the demands of the Ministry of Education by introducing English-as-a-Foreign-Language (EFL) classes in all elementary schools, there has been a shortage of qualified native Japanese EFL teachers at all levels. The English communicative ability (and pronunciation, in particular) of Japanese EFL teachers varies across educational levels, throughout Japan – including within individual prefectures.

Hinofotis & Bailey (1981, cited in Ueno 1995) stated that American undergraduates list pronunciation as the single most important factor in their evaluation of an international teaching assistant's ability. Shimizu (1995) reported that Japanese undergraduates also feel pronunciation is important – the results of her survey of 1,088 Japanese undergraduate students showed that of the qualities and attributes that students feel are important in Japanese teachers of English, *pronunciation* ranks second, even ahead of such things as intelligence, ability to explain things clearly, and treating students fairly. With such pressure on non-native teachers, how is their pronunciation?

Although many accounts exist of problem areas for Japanese learners of English pronunciation, these are primarily based on a contrastive analysis of English and Japanese, and large-scale phonetic analyses of actual pronunciation are sparse.

Method

Subjects

Group 1: STD – 40 Japanese 3rd year undergraduate students
Group 2: JHS – 20 Japanese junior high school EFL teachers
Group 3: SHS – 20 Japanese senior high school EFL teachers
Group 4: NAT – 10 native speakers of American English

Gender balancing was only possible with Group 4, which had 5 males and 5 females. Group 1 had 32 males and 8 females, a typical ratio in the University of Aizu's Computer Science and Engineering program. Group 2 had 5 males and 15 females, and Group 3 had 7 males and 13 females.

Much more data were recorded, but were not analyzed due to time constraints. Total subjects recorded in each group was 133 STD, 43 JHS, 34 SHS. Data from over 340 native English speakers exists in the Speech Accent Archive; the 10 speakers in Group 4 were chosen mainly on the basis of recording clarity.

Reading Stimulus

"Please call Stella. Ask her to bring these things with her from the store: six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station."
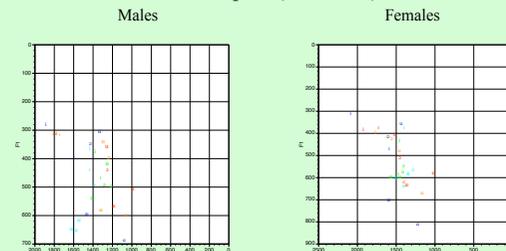
Procedure

Data Collection

Teachers recorded themselves only once after listening to the first author read the paragraph aloud twice. Recordings were made using Praat software running on Solaris-equipped Sun workstations in a classroom at the University of Aizu. Sony headset microphones were used. Students recorded themselves once during the first class and once during the last class of a 14-class (90-mins/class) semester. In the Results section, STD_ini and STD_fin refer to students in the initial and final (respectively) class of the semester.

Data Analysis

Several Praat scripts were written to facilitate automatic acoustic measurements of the data. All sound files and Praat *textgrid* files were hand-checked for accuracy. The following acoustic measurements were made: overall intensity range, overall pitch range, overall duration, duration of individual sentences and phonemes, voice onset time (VOT) of /p/, /t/, and /k/, and the first two formants (F1 & F2) of vowels of interest. Mean F1 and mean F2 were measured from the 25% point to the 75% point of the vowel with a 10 ms frame interval. Extreme outliers of formant analyses were automatically eliminated. Many studies (e.g., Clopper et al., 2005) make formant measurements at a single point in each vowel. However, Clopper et al. point out that different speakers manipulate spectral change differently, and we have taken the mean of formant measurements made every 10 ms between the 25% and 75% vowel duration points. Spectra were measured for /s/ and /ʃ/ to determine the frequency at peak power.

Results

Duration

Duration of paragraph

Duration of schwa

Pitch

Pitch Range over paragraph

Intensity

Mean intensity of paragraph

*NAT group is not included here because recording conditions (mic type, etc.) vary

Spectral Peaks of /s/ and /ʃ/

/s/ in "six" and /ʃ/ in "she"

Vowel Space (F1 and F2)
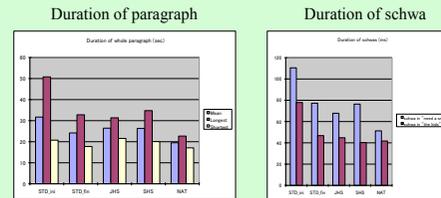
Males

Females

Groups are indicated by color:
red = STU_ini
orange = STU_fin
green = JHS
light blue = SHS
dark blue = NAT

Mean formants of vowels in these words:
/i/ - she, cheese
/ɪ/ - six, kids
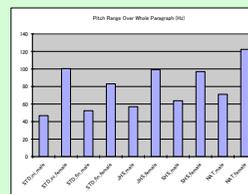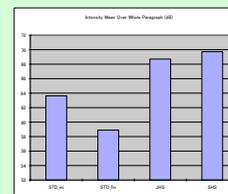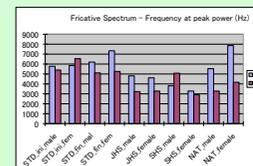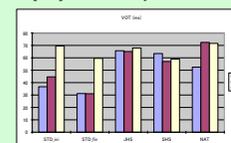/æ/ - ask, bags
/ə/ - a, the
/ɑ/ - Bob
/u/ - scoop

VOT

/p/-"peas" /t/-"toy" /k/-"kids"

Discussion and Conclusions

Using the duration of the entire reading as a measure of reading fluency, it is no surprise that the native speakers have the lowest duration – a mean of about 20 seconds. As expected, the students in their first class had the longest duration, but it is interesting to note that they increased reading speed by the final lesson to be faster than both groups of teachers. As for the duration of schwa, it is a common problem among Japanese ESL speakers that schwa is pronounced too much like a full vowel. A huge reduction in duration can be seen in students' pronunciation from the initial lesson to the final lesson. In fact, the students reach the level of Japanese teachers of English by their last lesson.

Females use a greater range of pitch than males, with native speakers using a greater range than non-native speakers. Pitch range generally increased from group to group as the *expected* proficiency increased.

Teachers spoke louder than students, probably to do with their confidence in English or maturity. Interestingly, the students spoke louder during the first class when they presumably had less confidence. For some, it is possible that they were self-conscious of speaking with good pronunciation and thus spoke more quietly.

In Japanese, when an /s/ occurs before a high front vowel, it is palatalized. Thus, some Japanese learners of English have difficulty pronouncing words like "sea" and "sit", pronouncing them like "she" and "shit" instead. In native speaker speech, English /s/ has a much higher spectral peak than /ʃ/. The students showed great improvement in this regard, from STD_ini to STD_fin. The senior high school teachers have a problem in this area.

For native English speakers, VOT decreases as speaking rate increases (e.g., Theodore et al., 2009). Results here show that the same is true for low-intermediate L2 speakers. Because of the choice of reading passage for data collection, our VOT results are based on only a single token per stop per subject. More tokens would give a better generalization, but our means are calculated over a fairly large number of subjects. Joto et al. (2007) found that native listeners misunderstood Japanese speakers' English voiceless stops /p/, /t/, /k/ if the VOT was less than 30 ms, 50 ms, and 55 ms respectively. From our data, it is unlikely that the teachers' /p/, /t/, and /k/ would be misunderstood, but the students' /t/ would be misunderstood as /d/.

As for vowel space, one clear problem is the lack of separation between /i/ and /ɪ/, for both males and females. This is not so much a problem with the teachers' speech, but they seem to separate these two vowels based on height, but not backness. The F1 value of the low back vowel is too low for most groups, both male and female. One puzzling point is that the female teachers' vowel space is extremely compact – except for /i/. This will be further investigated.

One disadvantage of using read speech is that it is not spontaneous. However, having all subjects read the same paragraph gave us identical data with which to make comparisons. One disadvantage of using the "Stella" paragraph for acoustic analysis is that many of the words contain liquids, nasals or glides. This makes it very difficult to automatically detect the beginning and end of vowels. However, an advantage of the paragraph is that there exist many examples of first and second language speakers' recordings on the internet.

Future Research

What acoustic measures best correlate with pronunciation proficiency as perceived by a native listener?
Does intense work on a particular passage of speech result in the improvement of an L2 learner's speech intelligibility in free speech?

References

Boersma, P (2001). "Praat, a system for doing phonetics by computer," Glot International **5**, 341–345. [software available at <http://www.praat.org>]
Clopper, C. G., Pisoni, D. B., & de Jong, K. (2005). "Acoustic characteristics of the vowel systems of six regional varieties of American English," J. Acoust. Soc. Am. **118(3), Pt. 1**, 1661-1676.
Hinofotis, F. & Bailey, K. (1981). "American undergraduates' reactions to the communication skills of foreign teaching assistants." In J. Fisher, M. Clarke, & J. Schacter (Eds.), *On TESOL '80: Building bridges* (pp. 120-133). Washington, DC: TESOL.
Joto, A., Nagase, Y. & Funatsu, S. (2007). "The effect of VOT on the intelligibility of English voiceless stops produced by native speakers of Japanese," Online Proceedings of the Phonetics Teaching &Learning Conference, Aug. 24-26, 2007, University College London.
Shimizu, K. (1995). "Japanese college student attitudes towards English teachers: A survey," The Language Teacher **19(10)**, 5-8.
Speech Accent Archive. (accessed 2009 at <http://accent.gmu.edu>). Linguistics program, Dept. of English, George Mason University.
Theodore, R. M., Miller, J. L., & DeSteno, D. (2009). "Individual talker differences in voice-onset-time: Contextual influences," J. Acoust. Soc. Am. **125(6)**, 3974-3982.
Ueno, N. (1995). "Teaching English pronunciation to Japanese English majors: A comparison of a suprasegmental-oriented and a segmental-oriented teaching approach," PhD dissertation, Temple University.

Acknowledgements