

An experimental study on English aspiration by Japanese students

Yuichi Todaka
Toshinobu Nagamine*

1. Introduction

It has been reported that most adult learners speak L2 with an accent, and that foreign accent usually becomes apparent between 5 and 8 years of age, becoming stronger beyond puberty (Flege & Fletcher, 1992). In addition, adult L2 pronunciation frequently fossilizes (Selinker, 1972 cited in Flege & Fletcher, 1992). For instance, Adult native Spanish speakers who had learned English as a second language produced significantly longer VOT (i.e., voice onset time) in English than in Spanish words, but their English VOT values were nonetheless shorter than VOT values produced by native speakers of English. On the other hand, Spanish children between the ages of 5 and 6 produced Spanish stop consonant, /t/ with appropriate short-lag VOT values and English /t/ with appropriate long-lag VOT values (Flege, 1991a). In addition, it's been reported that the acoustic differences resulting from articulatory differences (ex. dental vs. alveolar stops) may be detectable (Flege and Hammond, 1982; Flege, 1984; Flege, 1990c cited in Flege, 1991a); however, adult listeners nonetheless seem to identify the realizations of those sounds as identical at the phonological level. For instance, Spanish monolinguals consistently identified long-lag English [th] tokens as /t/ in a perception test (Flege, 1991b). Thus, it is difficult for nonnative adult speakers of English to acquire the subtle phonetic differences generated by the articulatory gestural differences.

Although it seems that adults have difficulties in acquiring the target pronunciation, this doesn't mean that they are not capable of learning any feature of the target pronunciation effectively. It's been reported that a significant improvement in adults' L2 pronunciation can be made by means of visual and auditory input (Matsui, 1995; Kellerman, 1990 cited in Matsui, 1995; see Nicoll et al., 1995 for motivation factor). Matsui (1995) reports that Japanese students' production of word-initial English labio-velar semi-vowel /w/ improved drastically by means of visual and auditory input. In addition, Murakawa (through personal communication) reports that Japanese students have made significant progress in the production of English /r/ by means of visual and auditory input. Therefore, we are interested in investigating the learnability of another English phonetic feature - stop aspiration by Japanese learners of English in the present study.

The general findings of cross-language VOT studies are that most ESL adult speakers of languages with short-lag stops produce English long-lag stops either with VOT values in their first language, or with VOT values that are intermediate to those in the L1 and their L2 (Schmidt and Flege, 1995).

A recent study (Schmidt and Flege, 1995) reports that non-proficient Spanish (NS) subjects of English (20 subjects), compared to relatively proficient Spanish subjects of English (20

*Toshinobu Nagamine, a member of English Phonetics Seminar, is a senior at Miyazaki Municipal University.

subjects), produced an English /p/ with VOT values close to their L1 VOT values more often, which indicates that the nonproficient NS might have substituted Spanish /p/ for English /p/. In addition, they found that the native English subjects' mean VOT value at a normal speaking rate was significantly smaller when compared to that at a fast speaking rate. On the other hand, a mixed result was found for the proficient and nonproficient Spanish learners of English. Although the proficient NS subjects showed a significant speaking rate effect on VOT as native English (NE) subjects, the nonproficient NS subjects did not. The nonsignificant mean VOT difference found for the nonproficient NS subjects between the two speaking rates implies, as mentioned earlier, that the subjects might not have established English motor realization rules necessary for the production of the appropriate target sound though inter-speaker variability was observed. Since a global foreign accent is known to be significantly correlated with VOT in stops spoken by nonnative speakers (Flege & Eefting, 1987a; Major, 1987 cited in Flege, 1991a) and the perceptual value for distinction between voiced and voiceless stops by Japanese was found to be similar to that by Spanish, it is interesting to see how Japanese learners of English produce English /p/.

Lisker and Abramson (1964) describe voicing and aspiration in stops in terms of VOT, and indicate that VOT values range from 0 to +25 ms for voiceless unaspirated stops and from +60 to +100 ms for voiceless aspirated stops. Although researchers report different mean VOT values for voiceless aspirated stops, these differences are due to differences in experimental conditions, stress/unstressed (Vance, 1987), word-initial/word-medial (Keating et al., 1983) and neighboring vowel effects (Shimizu, 1990). Nonetheless, VOT values for English voiceless aspirated stops reported previously fall within the long-lag VOT range while VOT values for Japanese voiceless stops can be categorized as having a short to medium range of voicing lag (Itoh et al., 1979; Homma, 1981; Shimizu, 1990).

Based upon the previous studies mentioned above, we can speculate several hypotheses regarding VOT values of Japanese learners of English: (1) they would produce English long-lag stops with VOT values that are characteristic of the short-to-medium lag values in their first language at normal and fast speaking rates; or (2) they would produce English long-lag stops with VOT values that are intermediate to those in the L1 and their L2 at a normal speaking rate. In the case of hypothesis (1), we can assume that Japanese speakers of English substituted English /p/ with Japanese-like /p/. In terms of hypothesis (2), we can speculate that Japanese speakers of English, having studied English for more than six years, have learned some of the acoustic phonetic properties of the target sound; however, they haven't established the necessary English motor realization rules, yet. In such a case, we would expect their production of English stops at a fast speaking rate to exhibit the same VOT values found at the normal speaking rate or to show various degrees of influence from changes in speaking rate.

The purposes of the present study are, therefore, threefold. First, we investigate VOT of Japanese learners of English at a normal speaking rate condition to see if their production is influenced by their L1. Second, we investigate VOT of Japanese learners of English at a fast speaking rate to see the effect of rate changes on their production. Finally, we investigate the learnability of the English-specific long-lag stop feature using audio-visual aids and the hyper-pronunciation training method.

2. Methods

2.1 subjects

We asked Miyazaki Municipal university students to participate in the pilot study, and 10 of them (one male and 9 female) volunteered to take part in the present study. All of the participants were first-year students (except 3), and have studied English for 7-8 years at school. None of them (except one) had prior English pronunciation training.

2.2 Study Period/Time Table

The study was conducted from October 2nd to October 30th, 1995 during the fall semester. Each session lasted thirty to forty minutes, twelve sessions in total. We decided to teach 30 - 40 minutes a session because it is the maximum time we thought we could assign to teach the pronunciation aspect of the target language in a regular ninety minutes class. In addition, twelve sessions were conducted because the semester system at Miyazaki Municipal university lasts 13 weeks including the final exam week.

2.3 Speech Material

The production of monosyllabic words, "pea" and "key" was elicited using the carrier sentence given below:

carrier sentence: Say ----- again

2.4 Experimental Procedures

Recordings were made in a recording room using a Panasonic microphone before and after the study. The spoken material was preamplified and recorded on a Panasonic tape recorder. The 10 subjects read the target words in a carrier sentence 10 times each at two speaking rates: normal, and fast (i.e., at a speaking rate that the subjects felt twice as fast as their normal speaking rate). Before the actual recording, the subjects practiced their readings. They read the target words in the carrier sentence 10 times in the following order: normal, fast, normal, fast.

The first normal rate set and the last fast rate set were discarded. The middle two productions were digitized using the Kay Computerized Speech Lab (CSL). Productions were digitized at a 10-kHz sampling rate which automatically set the low-pass filter to a cutoff frequency of 4 kHz. Thus, a total of 800 utterances (10 subjects x 10 repetitions x 2 target words x 2 speaking rates for the pre-training recording and 10 subjects x 10 repetitions x 2 target words x 2 speaking rates for post-training recording) was measured for each recording). The acoustic measurements were only done for the target word, "pea" in the present study.

Two temporal measurements were made in each utterance: VOT and vowel duration. VOT was measured for the interval from the release (R) to the onset of voicing (OV), and vowel duration was measured to examine the speaking rate changes. Onset of vowel was defined as the onset of visible energy in the region of F1 and higher formants on the spectrogram, with supplementary use of information in the waveform, and offset of the vowel was defined as the offset of energy in the region of F1 in the spectrogram as shown in Figure 1.

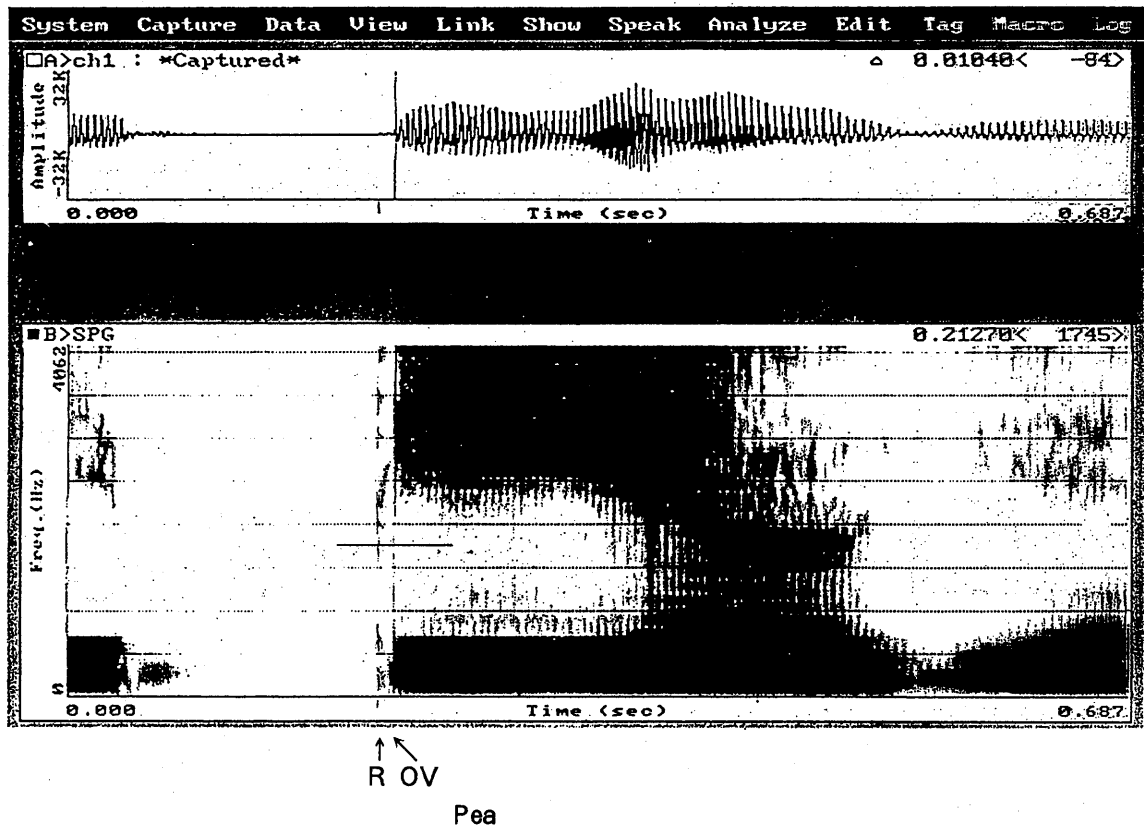


Fig. 1 sample waveform and spectrogram

3. Instructional guidelines

The process of speech production is comprised of three stages (refer to Figure 2): (1) air flows in and out of the lungs; (2) the vocal cords controlling the air flow; and (3) the vocal tract and nasal tract configurations affecting sounds produced (Wang, 1989). Various abdominal and laryngeal muscles work in harmony to flow air in and out of the lungs. The vocal cords within the larynx act as valves to control the air flow, and they provide voicing, control pitch, and voice quality, working together with the respiratory muscles. Finally, the acoustical properties of the sounds generated are determined by the vocal tract and the nasal tract configurations.

The primary source of acoustic energy depends, therefore, upon streams of airflow. The effective use of the pulmonic air is essential for learners to produce various sounds of a language. Especially, pulmonic air pressure created at various articulatory points in the vocal tract and the nasal tract affects the way in which acoustic properties of individual sounds are made. It's been reported that pulmonic air pressure is, in general, lower in the production of Japanese than that of English (cf. Ishiki et al., 1993). Since it is important for learners to understand the effective use of the source of acoustic energy when they produce English sounds, the hyper-training method proposed by Todaka (1995, 1996) was used in the present study.

The following four aspects constitute the structural basis of the hyper-pronunciation training method: (1) kinesthetic; (2) target feature; (3) affective; and (4) assessment. The kinesthetic aspect focuses on learners' understanding of basic speech production mechanisms (i.e., aerodynamic, voice projection, psychomotor, and pitch/loudness control practice). The

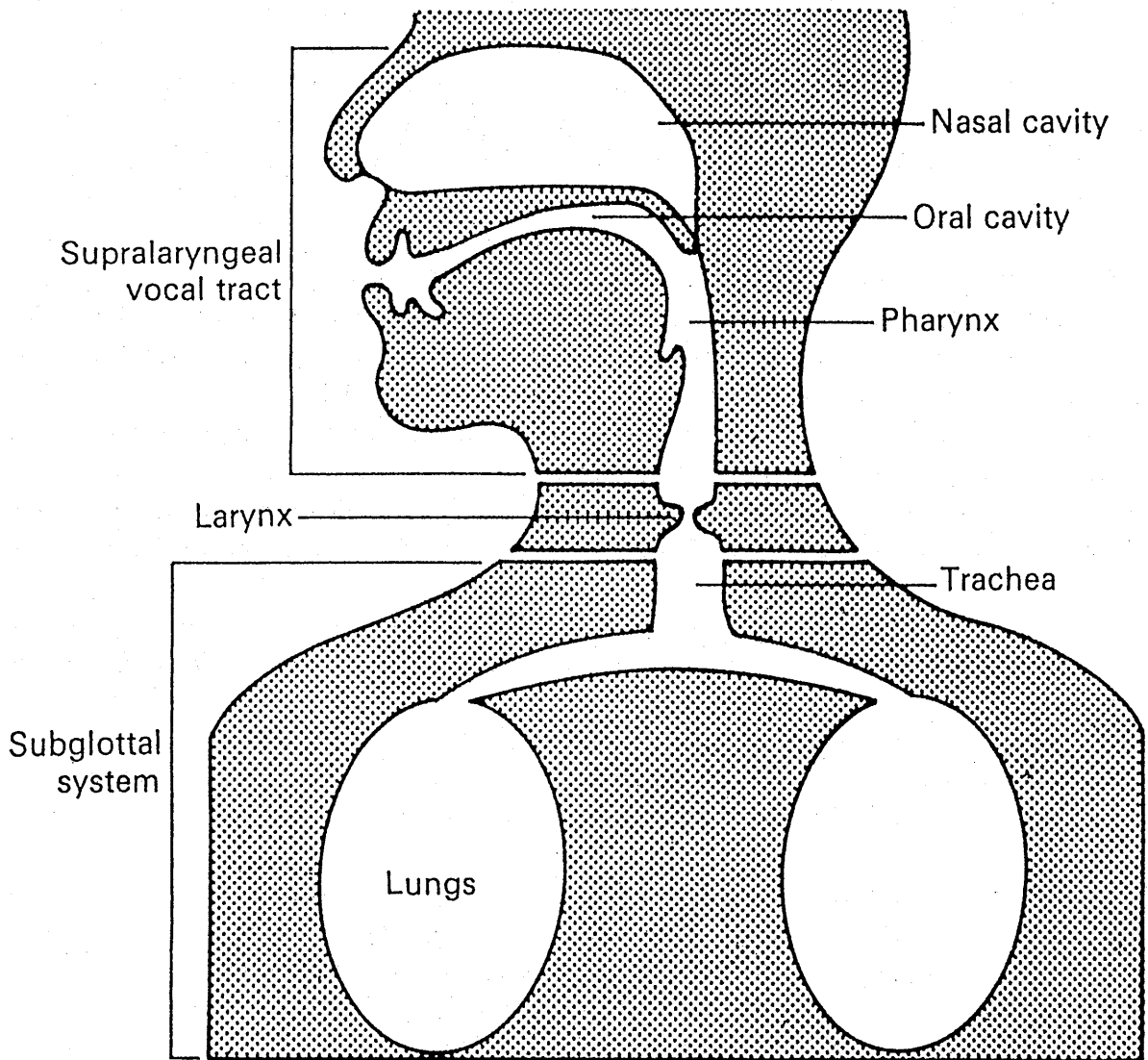


Fig.2 Human Speech Production

target feature considerations - articulatory settings, and three levels of recognition/production practice (i.e., from sound to discourse level)- focus on the actual production/recognition of the target feature. The target features are those that affect sets of sounds (e.g., aspiration and frication noises) or utterances as a whole (e.g., intonation) rather than discrete articulatory points. The affective considerations - self-esteem, self-confidence, risk-taking, and motivation - promote active learning and participation. Finally, assessment considerations - diagnostic evaluation, ongoing evaluation with feedback/peer feedback, and final evaluation - are the core of the evaluation of the learners in the present method.

A Macintosh computer connected to a video converter for use with the LL projection system to provide visual reinforcement for the production of the English stops, and an audio tape

created under the guideline of the hyper-training method to help learners recognize the target feature were also utilized during the sessions. Since it is crucial for learners to be able to apply the principles learned in the training sessions to real situations, a multi-level approach (i.e., from the most local-level, word-level, to discourse-level) to pronunciation teaching was also incorporated into the latter part of the training sessions. This global approach to teaching pronunciation has recently been advocated by many researchers (cf. Morley, 1994).

The following is the outline of the present method we followed in teaching the target feature in the present study.

Outline of the teaching method (taken from Todaka, 1996)

I. Kinesthetic Considerations (focusing on a feeling of articulation and speech production)

a. aerodynamic practice (effective use of abdominal muscles) - the fundamental principle of this activity lies in the realization of the differences of L1 and L2 productions. It is very important for learners to understand the effective use of the acoustic energy to be able to maintain the energy level up in spoken English.

b. voice projection (effective use of energy) - lack of confidence hinders learners from fully participating in class activities. It is, therefore, intended that learners are instructed to speak out in class (no mumbling).

c. psychomotor relaxation (e.g., lip spreading/rounding, mouth opening) - the fundamental principle of this activity lies in the understanding of the functions of various articulators while practicing pronunciation.

d. pitch/loudness control - the intent is to reinforce the learning of effective use of acoustic energy while focusing on pitch/loudness features, which in turn help learners facilitate the mastery of suprasegmental features.

II. Target Feature Considerations (e.g., aspiration/frication noises)

a. articulatory settings (visual aids) - the intent is to provide learners with the various settings of the target feature visually in order (1) to help them understand the L1 and L2 setting differences and (2) to help them develop study awareness.

b. recognition practice I (audio-visual aids, tape recorder) - the intent is (1) to help learners recognize the differences auditorily and (2) to develop self-monitoring skills. A slightly exaggerated version of the pronunciation of the target feature is utilized first to recognize the primary feature.

c. production practice I (audio-visual aids, tape recorder) - the intent is (1) to help learners produce the target feature effectively and (2) to develop self-monitoring skills. Silent production of the target feature is done several times to focus on the articulatory settings prior to actual production practice. In addition, learners are asked to produce the target sounds several times by varying articulatory settings to learn to hear the subtle differences as well as to produce them appropriately.

d. recognition practice II (audio-visual aids, tape recorder) - the intent is to help learners recognize the target sounds produced in words. learners listen to the target sounds produced in word-initial, word-medial, and word-final positions.

e. production practice II (audio-visual aids, tape recorder) - the intent is to help learners produce the target sounds effectively at the word level. The focus is on the target sounds and the articulatory settings. Silent production practice is reinforced.

f. recognition practice III (audio-visual aids, tape recorder) - the intent is to help learners recognize the target sounds produced in sentences.

g. production practice III (audio-visual aids, tape recorder) - the intent is to help learners produce the target sounds effectively at the word level. The focus is on the target sounds and the articulatory settings. Therefore, only the target sounds are first produced without the productions of vowels in each sentence. The focus is on the overall articulatory settings, and self-pronunciation adjustment strategies. As an assignment, learners are asked to read English passages aloud for thirty minutes a day to reinforce the learning.

h. recognition/production practice IV (tape recorder) - the intent is to help learners produce the target sounds effectively in semi-actual situations. This practice is done in the latter half of the sessions after students having developed confidence in producing the target sounds. Several speech tasks (e.g., self-introduction) are given and learners are asked to talk about the topics given. Each speech is recorded on a tape, and learners are asked to listen to their own speech to self-monitor it. Later learners work in pairs to give each other suggestions about the partner's recorded speech. Instructor and assistants guide learners in recognizing speech changes in themselves and in their classmates, paying attention to the positive changes rather than negative ones.

III. Affective Considerations

a. self-esteem and self-confidence - the intent is to encourage learners to gain self-esteem, self-confidence, knowledge of themselves and belief in their own capabilities for pronunciation. Thus, constant encouragements are given, and a comfortable classroom atmosphere is maintained.

b. risk-taking - the intent is to help learners be willing to try out hunches about pronunciation learning and take the risk of being wrong (let learners understand that it's OK to make mistakes in learning).

c. motivation - the intent is to help learner's motivational level kept high by having them take responsibilities in their own learning (let learners be active participants rather than passive recipients).

IV. Assessment (based on Goodwin et al's (1994) and Morley's (1994) suggestions)

a. diagnostic evaluation - The standardized diagnostic passage (Prator and Robinett, 1985) was used to gain information about our learners' difficulties at the outset of the training sessions.

b. ongoing evaluation with feedback/peer feedback (tape recorder) - The intent is to help learners develop monitoring skills and learning strategies as the instructor and teaching assistants evaluate learners' progress. In other words, learners are constantly guided in recognizing pronunciation changes taken place during the sessions. A comfortable classroom atmosphere is, therefore, a must to promote maximum teacher-student and student-student interactions.

c. final assessment (individually) - The same diagnostic passage was utilized to assess the learners' progress at the end of the sessions. The learners' progress is assessed auditorily by the instructor and teaching assistants. Assessment measures based on acoustic and auditory (i.e., native speakers' judgments) analyses are also frequently conducted to obtain the

information about learners' progress on particular features objectively. In such a case, utterances containing the target features are selected and analyzed to reconfirm our auditory judgments on learners' achievement. In addition, learners are asked to assess their own progress by having them compare their pre-training data with the post-training data. Teaching materials were made based upon the above pedagogical outline. Following are lesson plans given on October 2nd, 1995.

LESSON PLAN FOR OCTOBER 2, 1995

Target feature: aspiration

1. greetings,
2. kinesthetic training
3. description of lesson focus: awareness and recognition/production of target feature:
 - a. tongue placement (/p/, /t/, /k/)
 - b. tongue movement between articulatory settings (/p/, /t/, /k/)
 - c. discussion with students of sound system differences: trying to elicit awareness of L1 and L2 differences
4. Quick English, CD-ROM instructional software for pronunciation : consonants /p/, /t/, /k/, animations and descriptions of articulatory settings, choral practice
5. Listening practice
 - a. Quick English, CD-ROM instructional software
 - b. audio tape specifically designed for hyper-training
6. Production Practice
 - a. target feature
 - b. words
 - c. sentences

4. Results and Discussion

We examined VOT values of ten subjects in order to investigate the following aspects: First, we investigated VOT of Japanese learners of English at a normal speaking rate condition to see if their production is influenced by their L1. Second, we investigated VOT of Japanese learners of English at a fast speaking rate to see the effect of rate changes on their production. Finally, we investigated the learnability of the English-specific long-lag stop feature using audio-visual aids and the hyper-pronunciation training method mentioned earlier.

One of the ten subjects was often absent from the pronunciation training sessions; therefore, the results presented here are based upon the data we gathered from the nine subjects.

4.1 Pre-training data

The mean vowel durations (229 ms at normal and 163ms at fast) for all the subjects were shorter at the fast speaking rate than at the normal speaking rate ($p < .0001$). This indicates that all the subjects followed the instructions to change their speaking rates. Second, our subjects as a whole produced English /p/ with short to medium VOT values during the pre-training sessions. The mean VOT values at the normal and the fast speaking rates were

19.89 ms and 22.12 ms respectively ($p < .294$). Lisker and Abramson (1964) report that VOT values of short-lag and long-lag stops are in the ranges of 0 to 25 ms and of 35 to 100ms respectively.

As mentioned earlier, English voiceless stops in initial position have a long-lag VOT, while Japanese voiceless stops have a short to medium-lag VOT (Itoh et al., 1979; Shimizu, 1990). If the subjects had already established English motor realization rules, their mean VOT for English/p/ should have been in the range of 35 to 100ms. This means that the subjects produced English/p/ with Japanese-like short-lag VOT values. In other words, all the subjects in general merely substituted Japanese/p/ for English/p/, and the establishment of Motor Realization rules for the target sound hadn't taken place before they participated in the training sessions. This speculation is further supported by the fact that no significant difference between the mean VOT values at the normal and fast speaking rates was found.

Next, we examined the data to see if inter-speaker variability existed. The mean VOT of each subject measured before the training can be found in Table 1.

Table 1 Mean VOT obtained before training

Group 1			
subject/	Mean VOT (ms)		
speaking rate	NORMAL	FAST	
subject 3	12.4	12.3	
subject 4	34.6	25.2	
subject 5	8.0	12.1	
subject 6	18.1	15.3	
subject 7	42.1	59.0	
subject 8	13.0	11.1	
Group 2			
subject/	Mean VOT		
speaking rate	NORMAL	FAST	
subject 1	14.6	19.5*	
subject 9	16.0	28.4*	
subject 2	20.1	16.1*	

*indicates that the mean VOT values difference between the two speaking rates was significant at the 1% level.

We can see from Table 1 that the changes in speaking rate had little effect on the mean VOT in six of the nine subjects (henceforth, group 1), but did affect the mean VOT in three subjects (hereafter, group 2) at the 1% significant level. Concerning the six subjects in group 1, their mean VOT values, with one exception, was greatly different from that of native English speakers reported previously, and the values were in the range of Japanese short-lag VOT. It is, therefore, clear from the data that they hadn't established Motor Realization Rules for the target sound prior to the training sessions. Though speculative, one subject (subject 7) might have accurately assessed the acoustic phonetic properties of English/p/ because her mean VOT (42.17ms) at the normal speaking rate roughly matched the mean VOT (57.3ms) of native English speakers found in Schmidt and Flege (1995). However, it can be assumed that subject 7

had not established an appropriate phonetic category nor necessary English-specific motor realization rules for English/p/ since the mean VOT observed at the fast speaking rate was actually longer than the one found at the normal speaking rate. It has been reported, as mentioned earlier, that native speakers change VOT in word-initial tokens systematically when speaking at different rates. Their mean VOT value at the fast speaking rate was significantly shorter than the one at the normal speaking rate.

Regarding the three subjects in group 2, two subjects had longer mean VOT values at the fast speaking rate while the other subject had shorter mean VOT value at the fast speaking rate when compared to their mean VOT values at the normal speaking rate. These two subjects actually displayed a reversed effect of speaking rate changes on their mean VOT values. In addition, the mean VOT of these three subjects did not match that of Native English speakers, either. They remained in the range of Japanese short-to-medium lag VOT. Therefore, it can be said that these subjects had not established Motor Realization Rules for the target sound. And we can assume that they merely exaggerated their production of English/p/ at the fast speaking rate. In addition, the mean VOT of the other subject (i.e., subject 2) at both the speaking rates remained in the range of Japanese short-to-medium lag VOT though the mean VOT difference between the normal and the fast speaking rates was not only significant but also was in the right direction, as found in Schmidt and Flege (1995).

In summary, we can say that all the subjects hadn't established motor realization rules necessary for the production of English voiceless bilabial stop prior to the training sessions, and none of them had established a pertinent phonetic category for the target word before they participated in the training sessions.

4.2 Post-training data

The mean vowel durations measured in all subjects were significantly shorter ($p < .0001$) at the fast speaking rate (230.73ms) than at the normal speaking rate (178.76ms). This indicates that all the subjects followed the instructions to change their speaking rates. The mean VOT values at the normal and the fast speaking rates were 58.7 ms and 45ms respectively, and the difference observed was significant ($p < .0009$).

Schmidt and Flege (1995) reported native speakers' mean VOT for English/p/ at the normal and the fast speaking rates to be 57.3 ms and 42.9 ms respectively. It should be noted that the mean VOT values we observed after the training at two speaking rates clearly matched those of native speakers reported by the researchers.

Let us now turn to our discussion of the data we examined.

4.3 Comparison

As mentioned earlier, we found that our subjects had substituted Japanese/p/ for English/p/ before the training sessions. By comparing the pre-training mean VOT value with the post-training mean VOT value, a statistically significant difference was found ($P < .0001$). In other words, the post-training mean VOT value was significantly longer than the pre-training mean VOT value.

Now, let's turn to the mean VOT values for each subject, which are shown in Table 2.

Table 2 after training Mean VOT (ms)
in comparison with pre-training mean VOT (ms)

group 1 :

subject/ speaking rate	Mean VOT	
	NORMAL	FAST
subject 2	76.3 (20.1)	37.0 (16.1)*
subject 6	58.0 (18.1)	26.0 (15.3)*
subject 7	103.2 (42.1)	85.0 (59.0)*
subject 4	69.8 (34.6)	57.5 (25.2)*
subject 5	63.3 (8.0)	45.6 (12.1)*
subject 8	74.7 (13.0)	84.2(11.1)

group 2 :

subject/ speaking rate	Mean VOT	
	NORMAL	FAST
subject 1	31.6 (14.6)	43.3 (19.5)
subject 3	17.9 (12.4)	16.9 (12.3)
subject 9	23.8 (16.0)	19.2 (28.4)

() denotes the pre-training Mean VOT values of each subject.

*indicates that the mean VOT difference between the two speaking rates was significant at the 1% level.

First, the post-training mean VOT of all the subjects increased significantly when compared to their pre-training mean VOT. However, only six of the subjects' (i.e., subjects 2, 4, 5, 6, 7, 8) mean VOT at the normal speaking rate was in the range of long-lag VOT. Of these six, five subjects' mean VOT values at the normal speaking rate was significantly longer than their mean VOT values at the fast speaking rate. Their mean VOT values (except subject 6's mean VOT value at the fast rate) at both speaking rates fell within the range of long-lag VOT mentioned earlier, and they roughly matched those of native speakers found in Schmidt and Flege (1995). Thus, we can speculate that these subjects (i.e., subjects 2, 4, 5, 7) might have established motor realization rules and the pertinent phonetic category for English/p/ during the training sessions. On the other hand, the other subject (i.e., subject 8) had no significant mean VOT difference between the two speaking rates and the difference was in the reversed direction (i.e., longer mean VOT at the fast speaking rate) although the subject's post-training mean VOT was significantly longer than the pre-training mean VOT at the two speaking rates. In addition, subject 6's mean VOT (26.0ms) at the fast speaking rate fell within the range of short-to-medium lag VOT. It is, however, interesting to note that subject 6's post-training mean VOT (26.0ms) was significantly longer than the subject's pre-training VOT (15.3ms) at the fast speaking rate. This means that the subject 6 as well as subject 8 might have accurately assessed the acoustic phonetic properties of English/p/, but they were unable to establish the pertinent phonetic category for the target sound during the term of our investigation. However, it's possible to speculate that the phonetic category needed to distinguish Japanese from English sound is under way and subsequent improvement will come at a later stage of their pronunciation learning. These findings are in line with those of Yule and Macdonald (1994),

recognizing, however, that confirmation of this hypothesis is dependent on the completion of significant longitudinal studies. Nonetheless, the subjects have made significant improvement in their production of the word-initial voiceless bilabial stop.

The post-training mean VOT values of the remaining three subjects (i.e. subjects 1, 3, 9), though they increased significantly at both speaking rates when compared to their pre-training VOT values, remained in the range of short-to-medium lag VOT. This means that although they understood the English-specific aspiration feature, they haven't developed pertinent motor skills to produce the target sound appropriately. Nonetheless, these subjects have made statistically significant progress in their production when compared to their pre-training production. Therefore, it can be said that all the subjects have made significant progress in their production of English/p/though individual differences have been observed. It is, however, interesting to note that the post-training mean VOT value of subject 9 at the fast speaking rate was actually shorter than the pre-training mean VOT value. However, it is also important to notice that the subject's mean VOT values at the two speaking rates before the training sessions showed the reversed effect of speaking rates on VOT values that had been reported previously. This suggests, albeit speculatively, that the subject is moving in the direction of mastering the target feature, and further improvement is under way.

5. Conclusions

In the present study, an attempt was made to assess the degree of Japanese learners' foreign accent before and after the training sessions. First, our data confirmed the general findings of previous studies. That is, EFL/ESL adult learners of languages with short-lag stops produce the stops of the target language either with VOT values in their L1, or with VOT values that are intermediate to those in the L1 and in the L2. Our Japanese learners of English produced English/p/with VOT values in their L1. Second, our subjects before the training sessions did not show a significant speaking rate effect on VOT, which in turn confirmed Schmidt and Flege's (1995) finding. In other words, they found that their nonproficient Spanish subjects of English showed a nonsignificant rate effect on VOT, while their proficient Spanish subjects and their native English subjects did. If we consider VOT a barometer of a global foreign accent, we can say that our subjects as a whole are classified as nonproficient EFL learners. Third, we can infer from the results of the present study that our pronunciation training sessions helped the subjects improve their production of the target sound. It must also be noted that none of the subjects showed deterioration after the training sessions. However, the context was limited to one word, and the acoustic analysis was done in a laboratory setting. Additional evidence for their establishment of English motor realization rules must be gathered. In addition, the number of subjects and the term of our investigation were limited. Further investigations are, therefore, needed. Nonetheless, our auditory judgment on their overall L2 pronunciation before and after the training sessions shows that they have made distinct improvements. They seem to have gained confidence in their pronunciation, and many of them informed us of their determination to be more active in their own L2 learning.

Finally, we are currently conducting a series of studies to improve the teaching and learning of English pronunciation at Miyazaki Municipal University. Although we are still at the initial stages of designing an appropriate program for our students to facilitate their mastery of overall communicative skills, we firmly believe that we are heading in the right

direction. We hope that the present study may serve as a foundation on which subsequent studies can be based.

References

- Celce-Murcia, M. (1987). Teaching pronunciation as communication. In J. Morley (ed.), *Current Perspectives on Pronunciation*, 1-12. Washington D. C.: Teachers of English to Speakers of Other Languages.
- Collins, B. and Mees, I. M.(1995). Approaches to articulatory setting in foreign-language teaching. In J. W. Lewis (ed.), *Studies in general English phonetics: essays in honour of professor J.D.O'Connor*, 415-424, London and New York: Routledge.
- Esling, J. and Wong, R. (1983). Voice quality settings and the teaching of pronunciation. *Tesol Quarterly* , Vol.17(1), 89-95.
- Flege, J. (1984). The detection of French accent by American listeners, *J. Acoust. Soc. Am.* 76, 692 -707.
- Flege, J. (1990c). The production of cognate English words by native speakers of Spanish: more evidence for the distinction between phonetic implementation and realization, submitted to *J. Phon.*
- Flege, J. (1991a). Age of learning affects the authenticity of voice-onset-time(VOT) in stop consonants produced in a second language. *J.Acoust.Soc.Am.* 89, 395-411.
- Flege, J. (1991b). Perception and production: the relevance of phonetic input to L2 phonological learning, in T. Heubner and C. Ferguson (eds.), *Crosscurrents in second language acquisition and linguistic theories*. Amsterdam: Benjamins.
- Flege, J. and Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: evidence for phonetic category formation. *J. Acous. Soc. Am.* 83(2), 729 - 740.
- Flege, J. and Eefting, W.(1987a). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics* 15, 67-83.
- Flege, J. and Fletcher, K.(1992). Talker and listener effects of degree of perceived foreign accent. *Journal of Acoustical Society of America* 91(1), 370-389.
- Flege, J. and Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties, *Stud. Sec. Lang. Acquis.* 5, 1 - 18.
- Goodwin, J., Brinton, D., and Celce-Murcia, M.(1994). Pronunciation assessment in the ESL/EFL curriculum. In Morlry(ed.), *Pronunciation Pedagogy and Theory*, 3-16. VA: Teachers of English to speakers of other languages.

- Homma, Y.(1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9, 273-281.
- Honikman, B. (1964). Articulatory settings. In D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott and J. L. Trim (eds), *In Honour of Daniel Jones*, London: Longman, 73-84.
- Ishiki, M. and Matsui, C. (1993). *Eigo onseigaku: nihongotono hikakunite*. Tokyo: Kenkyusha.
- Itoh, M., Sasanuma, S., Tatsumi, I. F. and Kobayashi, Y.(1979). Voice onset time characteristics of apraxia of speech. *Ann.Bll.RILP* 13, 123-132.
- Keating, P., Linker, W. and Huffman, M. (1983). Patterns in allophonic distribution for voiced and voiceless stops, *J. Phonetics* 11, 277 - 290.
- Kellerman, S. (1990). Lip service: the contribution of the visual modality to speech perception and its relevance to the teaching and testing of foreign language listening comprehension, *Appl. Ling.* , vol. 11, no. 3, 272 - 280.
- Lisker, L. and Abramson, A. (1964). A cross language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.
- Major, R.(1987). English voiceless stop production by speakers of Brazilian Portuguese. *J.Phon* .15, 197-202.
- Matsui, J. (1995). The use of Audio-Visual Aids in the Language Laboratory to Teach Lip-Rounding. *Language Laboratory* ,Vol.32, 169-176.
- Morley, J.(1994). A Multidimensional Curriculum Design for Speech-Pronunciation Instruction. *Pronunciation Pedagogy and Theory*, 66-91. Teachers of English to Speakers of Other Languages.
- Nicoll, H. and Todaka, Y. (1995). A pilot study on computer-assisted pronunciation teaching. *Bulletin of Miyazaki Municipal University*, vol. 3., no. 1, 1 -16.
- Prator, C.H. and Robinett, B.W.(1985). *Manual of American English pronunciation*. FL: Harcourt Brace Jovanovich.
- Schmidt, A. and Flege, J. (1995). Effect of Speaking Rate Changes on Native and Nonnative Speech Production. *Phonetica* 52, 41-54.
- Selinker, L. (1972). Interlanguage, *Int. Rev. Appl. Ling.* 10, 209 - 231.
- Shimizu, K. (1990). *A cross language study of Voicing Consonants of Stop Consonants in Asian Languages*. Unpublished Ph.D. dissertation, University of Edinburgh, .
- Todaka, Y. (1995). A preliminary study of voice quality differences between Japanese and

America English: some pedagogical suggestions. *JALT Journal* , Vol. 17, 261-268.

Todaka, Y. (1996). The use of audio-visual aids and hyper-pronunciation method in teaching English consonants to Japanese college students. *Proceedings of the First Seoul International Conference on Phonetics Sciences*, 149-154. Seoul: The Phonetic Society of Korea.

Vance, T. (1987). *An introduction to Japanese phonology*. Albany: State University of New York Press.

Wang, W.(1989). *The basis of speech*. Tokyo: Kenkyusha.

Yule, G. and Macdonald, D.(1994). The Effects of Pronunciation Teaching. *Pronunciation Pedagogy and Theory*, 111-118. Teachers of English to Speakers of Other Languages.

