Effects of Hyper-Pronunciation Training Method on Japanese University Students’ Pronunciation

Toshinobu Nagamine
Prefectural University of Kumamoto, Japan

Bio Data:
Toshinobu Nagamine is currently Associate Professor in the Department of English Language & Literature at Prefectural University of Kumamoto, Japan, where he teaches English phonetics, English linguistics, and EFL teacher education courses. He holds a Ph.D. in Composition & TESOL from Indiana University of Pennsylvania, USA. His research interests include teacher development, pronunciation pedagogy and theory, and grammar pedagogy and theory.

Abstract
Mutual intelligibility or overall comprehensibility of L2 speech has been regarded as a crucial goal in recent ESL pronunciation pedagogy. In other words, native-like accuracy has received less pedagogical attention. It is not necessarily reasonable, however, to underestimate native-like accuracy in pronunciation teaching targeting student-teachers in ESL/EFL teacher-education settings. This study, therefore, examined the efficacy of Hyper-Pronunciation Training Method, which “initially exaggerates pitch contours and the duration of stressed syllables in English” (Celce-Murcia, Brinton, & Goodwin, 1996, p. 26), in an EFL teacher-education contexts. Subjects of the study were Japanese university students (N=30), and they all took part in a year-long pronunciation training. At the time of the investigation, the subjects had a plan to become EFL teachers after graduation, but their level of confidence in their English pronunciation was extremely low. Before and after the pronunciation training sessions, subjects’ speech was collected and analyzed using Praat (acoustic-analysis software; Boersma & Weenink, 2008). In the acoustic analysis, a prime focus was put on Voice Onset Time (VOT) of voiceless bilabial, alveolar, and velar stops/plosives and pitch range. Thus, both segmental and suprasegmental aspects of L2 speech were analyzed in this study. This paper reports on the results of acoustic analysis and discusses the applicability of Hyper-Pronunciation Training Method in EFL/ESL teacher-education settings.

Keywords: Hyper-Pronunciation Training Method, Pronunciation pedagogy, Voice Onset Time, Pitch range, Japanese university students

1. Introduction
Mutual intelligibility or overall comprehensibility of L2 speech has been regarded as a primary goal in recent ESL pronunciation pedagogy. Since
suprasegmental features play a crucial role in communication (see Munro & Derwing, 1995), such features as sentence-stress, rhythm, pitch, and intonation receive much pedagogical attention in current pedagogy of English pronunciation (Binghadeer, 2008; Celce-Murcia, Brinton, & Goodwin, 1996; Nagamine, 2002). In other words, the native-like accuracy of segmental features (i.e., consonants and vowels) tends to receive less pedagogical attention. It is not necessarily reasonable, however, to underestimate native-like accuracy in pronunciation teaching targeting student-teachers in ESL/EFL settings. As Ashby (2002) claims, a modern language teacher is expected to have sufficient knowledge of articulatory phonetics, a well-trained ear, and skills to analyze as well as remedy learners’ articulatory errors.

As an EFL teacher-educator myself, I have encountered such dilemmas as EFL student-teachers’ non-native-like pronunciation and their self-efficacy, self-confidence in model reading, or students’ perceptions of EFL student-teachers’ non-native-like pronunciation. Thus, ESL/EFL teacher-educators may consider both native-like accuracy and mutual intelligibility or overall comprehensibility a major goal when teaching English pronunciation to student-teachers. Based on such observations, this study examined the efficacy of Hyper-Pronunciation Training Method, which “initially exaggerates pitch contours and the duration of stressed syllables in English” (Celce-Murcia et al., 1996, p. 26) to instruct segmental as well as suprasegmental features of the English language to EFL student-teachers. Acoustic data was collected before and after pronunciation training sessions and analyzed in terms of two acoustic properties of L2 production: Voice Onset Time (henceforth VOT) in word-initial voiceless bilabial, alveolar, and velar stops/plosives and pitch range. Therefore, segmental as well as suprasegmental aspects of L2 speech were investigated.

1. Literature Review

It has been reported that native speakers of Japanese have unique characteristics of vocalization, and that they tend to show such characteristics in L2 production (i.e., the speech production of English)
(see Todaka, 1995, 1996). Those phenomena have been known as L1 transfer or L1 interference. Regarding the unique characteristics of vocalization of native speakers of Japanese, Tateno (1984) identified the following characteristics of voicing: (a) A tendency to tighten the throat so that the root of the tongue is raised. As a result, the pharynx is narrowed; (b) The opening of the mouth is narrow; (c) A tendency to produce rather strained voice; (d) Bad resonance; (e) Bad glottal efficiency; (f) More inspiratory noise; (g) When uttering a loud voice, they tend to yell and cannot project the voices appropriately; and (h) Less expiratory pressure. As for (h), Ishiki and Matsui (1993) add that Japanese speakers do not effectively use pulmonic air pressure in L2 production, referring to the importance of the effective use of abdominal respiration. It can thus be speculated that especially in initial stage of learning, Japanese L2 speakers produce English words, phrases, and sentences with less abdominal respiration than native speakers of English (cf., Ishiki & Matsui, 1993). Hence, English consonants such as word-initial voiceless bilabial, alveolar, and velar stops/plosives (i.e., /p/, /t/, /k/) are likely to be produced with insufficient aspiration noise (Todaka & Nagamine, 1996). Regarding (b), (d), and (e), it can be postulated that Japanese L2 speakers, especially in initial stage of learning, do not skillfully use the speech organs (i.e., they do not fully maneuver, for instance, the muscles of the mouth in order to produce English sounds). The characteristics of voicing reported by Tateno may indicate the need for Japanese L2 speakers to acquire the overall maneuvering skills of the vocal and nasal tracts, as well as phonetic realization rules specific to the English language (cf., Schmidt & Flege, 1995). Thus, it can also be speculated that the lack of speech motor control and phonetic realization rules causes Japanese L2 speakers to fail in generating appropriate acoustic properties in their L2 production (e.g., intonation contours/patterns, pitch range, etc) (see Todaka, 1993).

VOT is defined as “[t]he interval between the release of a closure and the start of the voicing” (Ladefoged, 1999, p. 125). VOT signifies the time between the beginning of “the articulatory gesture responsible for the release of a closure” and the beginning of “the laryngeal gesture responsible for vocal fold vibration” (Cho & Ladefoged, 1999, p. 225).
Namely, VOT can be considered laryngeal timing. As a crucial acoustic feature of phonation contrasts, VOT serves to separate stop categories in both production and perception. VOT has been widely used to differentiate phonetic, as well as phonemic, categories across languages (Cho & Ladefoged, 1999; Keating, Linker, & Huffman, 1983; Lisker & Abramson, 1964; Riney, Takagi, Ota, & Uchida, 2007). In addition, the ability to discriminate differences in VOT has been investigated in infants. For instance, Eimas, Sinqueland, Jusczyk, and Vigorito (1971) studied one month old infants. They found that the subjects could distinguish the differences in VOT in a categorical fashion, which implies that the ability to perceive VOT differences may be innate. The infants’ ability to detect cross-language VOT differences was also examined by Lasky, Syrdal-Lasky, and Klein (1975). Subjects (four to six months old infants) who were raised in Spanish and English-spoken environments could detect VOT differences across the two different languages which have differing VOT categorical boundaries. Adult learners of English who have not been raised in English-spoken environments have much difficulty in detecting or producing VOT differences due possibly to the lateralization of the brain or fossilization (either temporary or permanent) of their interlanguage system (Acton, 1984; Selinker, 1972; Todaka, 1996). These studies indicate the necessity for Japanese learners of English to learn language-specific VOT characteristics through formal instruction and/or training.

Intonation plays a critical role in communication. It is used to express speaker’s personal attitude or emotion along with other prosodic as well as paralinguistic features (Nagamine, 2002). Todaka (1993) maintains that Japanese speakers of English have a tendency to produce narrower pitch excursions than native speakers of English. In other words, Japanese speakers of English tend to produce monotonous utterances. Binghadeer (2008) investigated Saudi EFL learners’ pitch ranges and compared the results to those of native English speakers. She collected and analyzed both American and British native speakers’ pitch ranges as well. Significant differences in mean pitch range were observed between Saudi EFL learners and native English speakers for utterances with falling intonation. No significant differences were observed between Saudi EFL
learners’ pitch ranges and those of American native speakers for utterances with rising intonation, while significant differences were observed between Saudi EFL learners’ mean pitch range and that of British native speakers. Like Japanese speakers of English, Saudi speakers of English have similar problems in pitch range, which might be attributed to L1 influence. Although pronunciation aspects of L2 speech have been known to be highly resistant to change when fossilization occurs (Selinker, 1992), nonnative speakers “can be taught to broaden their range of pitch to carry more dramatic changes characteristic of English intonation” (Binghadeer, 2008, p.111; Nagamine, 2002).

2. Method

3.1. Subjects

Subjects of this study were Japanese university students (freshmen; 18-19 years old; all females; N=30). None of them had experienced studying abroad and had received formal pronunciation training before participating in this study. They had a plan to become EFL teachers after graduation (i.e., student-teachers). Their level of confidence in English pronunciation was extremely low, and they voluntarily decided to participate in this study to improve their pronunciation. One year-long pronunciation training was carried out applying Hyper-Pronunciation Training Method. L2 speech samples (acoustic data) were collected before and after the pronunciation training sessions.

2.2. Research Questions

Research questions addressed in this study are as follows:

(a) What effects can be observed in EFL student-teachers’ VOT values after participating in pronunciation training sessions?
(b) What effects can be observed in EFL student-teachers’ pitch ranges after participating in pronunciation training sessions?
(c) What is the efficacy of Hyper-Pronunciation Training Method?

2.3. Speech Material & Experimental Procedures

2.3.1. VOT
The following carrier sentence was used to elicit VOT values (msec.) of word-initial voiceless bilabial, alveolar, and velar stops/plosives (/p/, /t/, /k/).

“Say _____ again.”

/p/ --- pit, pat, put
/t/ --- tick, tap, took
/k/ --- kick, cap, cook

Three words for each phoneme (3x3) were prepared before the investigation. Before and after the pronunciation training sessions, individual subject’s speech was recorded in a sound-proof recording studio at PUK for later acoustic analysis (see Figure 1). The subjects were asked to read the carrier sentence, inserting each word in the sentence three times. Mean VOT values were calculated for each phoneme.

Figure 1. Measurement of VOT values (“pit”)

In this study, previously reported native speakers’ (both Japanese and English) VOT values were used as a reference in order to compare subjects’ VOT values with native speakers of English (see Table 1). As indicated in Table 1, the data of native speakers of English and Japanese were adopted from the studies conducted by Lisker and Abramson (1964)
and Riney, Takagi, Ota, and Uchida (2007) respectively.

Table 1.

Previously Reported VOT Values of Native Speakers of English and Japanese

<table>
<thead>
<tr>
<th></th>
<th>/p/</th>
<th>/t/</th>
<th>/k/</th>
</tr>
</thead>
<tbody>
<tr>
<td>English* (American)</td>
<td>58.00</td>
<td>70.00</td>
<td>80.00</td>
</tr>
<tr>
<td>Japanese**</td>
<td>30.00</td>
<td>28.50</td>
<td>56.70</td>
</tr>
</tbody>
</table>

*English data was taken from Lisker and Abramson (1964).

**Japanese data was taken from Riney, Takagi, Ota, & Uchida (2007).

A graphic representation of the data presented in Table 1 is shown in Figure 2. As you can see, obvious differences in VOT values can be observed between the native speakers of Japanese speaking Japanese and the native speakers of English speaking English.
2.3.2. Pitch Range

Scripts from a TV show (formal MC of a tabloid TV show) and a movie (informal conversational English) were used to elicit subjects’ pitch ranges. Before and after the pronunciation training sessions, an individual subject was asked to read entire scripts, and the recording was made in a soundproof recording studio at PUK. The following randomly selected utterances from the entire scripts were analyzed in terms of pitch range (Hz; see Figure 3). The original speech data of native speakers of English was analyzed for later data comparison.

[MC] “Two very large twins and the man who squeezes himself between them.”

(Exaggerated Speech)

[Movie] “Why? You don’t think I’m capable?”

(Natural Conversation Speech)

Figure 3. Measurement of Pitch Range (MC: Exaggerated Speech)
2.4. Instructional Procedures

2.4.1. Pronunciation Training: Hyper-Pronunciation Training Method

Pronunciation training carried out in this study was composed of three phases: Phases 1, 2, and 3. In Phase 1, the subjects were instructed to exaggerate their English pronunciation/articulation (exaggerated speech production exercises), to listen primarily to exaggerated speech (recognition exercises), and to become aware of English-specific sound features. In Phase 2, adjustment exercises were done to shift their exaggerated speech production to more natural conversational level production. In Phase 3, only natural conversational level production was repeatedly practiced retaining English-specific sound features. These three phases are essential components of Hyper-Pronunciation Training Method (cf., Todaka & Nagamine, 1996). The method is designed to have L2 learners produce and practice L2 features of English pronunciation in an exaggerated manner, which allows them to raise their awareness of English-specific sound features. The method is known to be very effective when teaching such suprasegmental features as sentence-stress, rhythm, and intonation, because as indicated before, it “initially exaggerates pitch contours and the duration of stressed syllables in English” (Celece-Mrcia et al., 1996, p.26).

Hyper-Pronunciation Training Method focuses on learners’ understanding of basic speech production mechanism as well: aerodynamic, voice projection, psychomotor, and pitch/loudness control. The method focuses on both recognition and production practices (a dual-focus approach). The target features include, but not restricted to, aspiration and frication noises or utterances as a whole (e.g., intonation) rather than discrete articulatory points. The method takes into account learners’ self-esteem, self-confidence, risk-taking, and motivation. The ultimate goal is to promote learners’ monitoring of their own articulation, active learning, and participation. The method includes diagnostic evaluation, ongoing evaluation with feedback (e.g., instructor’s feedback and peer feedback), and final evaluation done by the instructor. A sample lesson plan used in this study is shown in Appendix A.
3. Results and Discussion

4.1. VOT

A paired-sample t-test was implemented to see if there was a significant difference in the mean scores between the pre- and post-training VOT data. The results are presented in Table 2. There was a significant difference in the VOT values of word-initial voiceless bilabial, alveolar, and velar stops/plosives in the pre-training data (/p/: $M=34.73$, $SD=14.92$; /t/: $M=40.68$, $SD=16.73$; /k/: $M=52.31$, $SD=18.53$) and the post-training data (/p/: $M=53.43$, $SD=11.79$; /t/: $M=67.77$, $SD=10.22$; /k/: $M=80.86$, $SD=7.39$); /p/: $t(29)=-7.65$, $p<.001$; /t/: $t(29)=-9.16$, $p<.001$; /k/: $t(29)=-7.79$, $p<.001$.

As Table 2 shows, subjects’ VOT values of word-initial voiceless bilabial, alveolar, and velar stops/plosives (/p/, /t/, /k/) to a great extent improved through the pronunciation training sessions. It should also be noted that subjects’ mean scores of VOT became very close to native English speakers’ VOT values, though individual differences were observed. In addition, the standard deviations ($SD$) of VOT values observed in the pre-training data decreased in the post-training data. This implies that the subjects of this study as a whole came to produce VOT values more constantly then before. Thus, it can be said that the pronunciation training sessions conducted in this study affected the subjects’ VOT values in a positive manner.

<table>
<thead>
<tr>
<th>Table 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Paired-Sample t-Test for the Pre- and Post-Training Mean Scores:</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VOT</th>
<th>$M$</th>
<th>$SD$</th>
<th>$t$ value</th>
<th>df</th>
<th>sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Training /p</td>
<td>34.73</td>
<td>14.92</td>
<td>-7.65</td>
<td>29</td>
<td>(p &lt;.001)</td>
</tr>
<tr>
<td>Post-Training /p</td>
<td>53.43</td>
<td>11.79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Training /t</td>
<td>40.68</td>
<td>16.73</td>
<td>-9.16</td>
<td>29</td>
<td>(p &lt;.001)</td>
</tr>
<tr>
<td>Post-Training /t</td>
<td>67.77</td>
<td>10.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Training /k</td>
<td>52.31</td>
<td>18.53</td>
<td>-7.79</td>
<td>29</td>
<td>(p &lt;.001)</td>
</tr>
<tr>
<td>Post-Training /k</td>
<td>80.86</td>
<td>7.39</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.2. Pitch Range

Paired-sample *t*-tests were carried out to see if there was a significant difference in the mean scores between the pre- and post-training pitch range data for both exaggerated level and natural conversation level speech. The results are shown in Tables 3 and 4.

**Table 3.**
*Paired-Sample t-Test for the Pre- and Post-Training Mean Scores: Pitch Range (Exaggerated Speech)*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>t value</th>
<th>df</th>
<th>sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Training</td>
<td>256.14</td>
<td>87.65</td>
<td>-2.44</td>
<td>29</td>
<td>.02</td>
</tr>
<tr>
<td>Post-Training</td>
<td>309.22</td>
<td>69.75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As for the exaggerated speech, there was a significant difference in the pre-training data (*M*=256.14, *SD*=87.65) and the post-training data (*M*=309.22, *SD*=69.75); *t*(29)=−2.44, *p*=.02. Native English speaker’s pitch range for the exaggerated level speech was 376.06 Hz (minimum pitch 82.93 Hz; maximum pitch 458.99 Hz; mean pitch 239.46 Hz). Though individual differences were observed, subjects’ pitch ranges of the exaggerated level speech showed drastic improvement through the pronunciation training sessions. The decrease of SD in the post-training data signifies that the subjects as a whole came to produce more constant pitch ranges than before.

**Table 4.**
*Paired-Sample t-Test for the Pre- and Post-Training Mean Scores: Pitch Range (Natural Conversation Speech)*

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>t value</th>
<th>df</th>
<th>sig. (2-tailed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Training</td>
<td>238.29</td>
<td>76.58</td>
<td>-.0009</td>
<td>29</td>
<td>ns</td>
</tr>
<tr>
<td>Post-Training</td>
<td>238.30</td>
<td>57.57</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Regarding the natural conversation speech (see Table 4), there was no
significant difference in the pre-training data ($M=238.29$, $SD=76.58$) and the post-training data ($M=238.30$, $SD=57.57$); $t(29)=-.0009$, $ns$. Native English speaker’s pitch range for the natural conversation speech was 226.29 Hz (minimum pitch 159.34; maximum pitch 385.63 Hz; mean pitch 272.49 Hz). To reiterate, there was no significant difference in the pre- and post-training mean scores of pitch range of the natural conversation speech. Possible explanations for this finding are: (a) the subjects, while looking at the entire script of the movie, might have realized that they were expected to adjust their speech to natural conversation level production; and thus (b) the subjects could successfully adjust their speech production to suit the situation/context of the scene reflected in the movie script. The decrease of $SD$ in the post-training data implies that the subjects as a whole could produce more constant pitch range than before. It should be noted here, however, that 15 out of 30 subjects produced greater pitch ranges in the post-training data than those of the pre-training data, which, in turn, signifies that adjusting pitch ranges (from exaggerated level to natural conversation level speech) might have been a challenging task for most of the subjects. Accordingly, though speculative, more than one year might have been needed for every subject to complete such a task. All in all, however, it can be said that the pronunciation training sessions affected the subjects’ pitch range production in a positive fashion.

4. Concluding Remarks
A positive impact of the pronunciation training sessions was observed in the subjects’ post-training data. The subjects’ VOT values and pitch ranges improved greatly and became very close to those of the native speakers of English. In other words, the efficacy of Hyper-Pronunciation Training Method was confirmed at segmental and suprasegmental levels. When producing high VOT values in particular, learners are required to effectively use their abdominal air pressure as well as their pulmonic air pressure. Learners are also required to know L2 specific features as well. However, knowing those features and actually maneuvering speech organs to properly produce L2 speech are two different things. In this regard, the subjects who took part in the year-long pronunciation training sessions
demonstrated their knowledge as well as the maneuvering skills of speech organs appropriately. As mentioned earlier, Japanese learners of English tend to produce less aspirated voiceless consonants (i.e., low VOT values) and monotonous speech with less rising and falling intonation contours (i.e., narrow pitch range). Such tendencies were also observed in the subjects’ pre-training data, but the drastic improvement in those acoustic properties was observed in the post-training data. Accordingly, Hyper-Pronunciation Training Method can be said to have high applicability or feasibility, especially for teacher-education settings.

Finally, there are limitations of the study. The subjects were all would-be EFL teachers. Even though their confidence level in English pronunciation was extremely low, they might have been highly motivated to learn and improve their English pronunciation. The subjects were also restricted to only Japanese learners of English. Therefore, the findings reported in this article might not extend to other English learners from differing L1, socio-cultural, and educational backgrounds. Furthermore, the subjects were all females. In other words, sex-related issues or differences were not taken into consideration in this study. In addition, although two different scripts (exaggerated speech and natural conversation level speech) were prepared and investigated in terms of pitch range in this study, the experimental condition might not reflect real-life communication situations; this holds true for the analysis of VOT values in this study. Nevertheless, on the basis of this study, it can be said that Hyper-Pronunciation Training Method can be a reasonable option for ESL/EFL teacher-educators to help would-be EFL/ESL teachers who wish to raise both native-like accuracy and intelligibility or comprehensibility levels of their L2 speech production.

References


Phonetics, 35(3), 439-443.

Appendix A: A Sample Lesson Plan for June 16, 2009
1. Greetings
2. Kinesthetic training & voice projection exercises
3. Description of lesson focus: awareness and recognition/production of a target feature
4. Mini-Lecture: Voiceless Consonants /p/, /t/, /k/
   (a) tongue placement (/p/, /t/, /k/)
   (b) tongue movement between articulatory settings (/p/, /t/, /k/)
   (c) group discussions among students: trying to raise awareness of L1 and L2 sound systems
5. Listening practices
   (a) instructional audio prepared for recognition exercises
   (b) pair recognition practice

6. Production practices
   (a) target feature
   (b) word-level exercises
   (c) phrase-level exercises
   (d) sentence-level exercises
   (e) choral practice