Experimental Studies on Prosodic Features in Second Language Acquisition:Training Japanese learners to produce natural English

Iba, Midori

(Degree)
博士（学術）
(Date of Degree)
2010-03-25
(Date of Publication)
2011-12-22
(Resource Type)
doctoral thesis
(Report Number)
甲4837
(URL)
https://hdl.handle.net/20.500.14094/D1004837

Experimental Studies on Prosodic Features
in Second Language Acquisition

MIDORI IBA

Experimental Studies on Prosodic Features in Second Language Acquisition

Training Japanese learners to produce natural English

Shumpusha Publishing Yokohama

Published by
Shumpusha Publishing
Yokohama-shi Kyouikukaikan 3F
53 Momijigaoka, Nishi-ku, Yokohama-shi
220-0044 Japan

Cover Designer: Yoko Goto
Printed in Japan by SHINANO BOOK PRINTING CO., LTD

ISBN-13: 978-4-86110-235-6

Contents

Preface

This book consists in the main of a doctoral dissertation, completed this year, which I began in 2007. My work over the last three years reawakened my interest in an article I had written previously concerning the choice of models for teaching English pronunciation, and this article (Iba, 2002) forms the last chapter of the book. My subject is the teaching of prosody in foreign language acquisition, and in what order the various components are best taught.

Prosody, here, refers not to the study of versification but to the rhythmic and intonational aspects of language. It is impossible, of course, to speak without rhythm and intonation, and these factors affect the quality of speech sounds. However, teaching prosodic features presents particular difficulties because they are not confined to any one segment but occur in the higher levels of an utterance. That is why they are called suprasegmental. More precisely, the suprasegmental features are those which operate over longer stretches of speech, such as rhythm, intonation, stress, pitch, and voice quality as opposed to the segmental features referred to as the individual sounds.

The suprasegmental features are sometimes deemed unteachable. But is this really the case? I have my doubts, and have performed three experiments concerning the teaching of prosodic features which demonstrate that the suprasegmental features are at least learnable. My findings also indicate the importance of the sequence of training, namely, the apparent superiority of training prosodic features over individual sounds. It is my hope that readers will be interested in the contents of this book.

Summer 2010

Midori Iba

Acknowledgements

I am indebted to Prof. Reiko Akahane-Yamada at Kobe University, my supervisor, for providing consistent support throughout the dissertation process. I would also like to thank Dr. Hideki Murakami at Kobe University, my brother, who advised me on the statistics in Study 2. Prof. Shinobu Mizuguchi at Kobe University kindly permitted me to use software of her own design in Study 3. Mr. Ryota Asahi and Mr. Genya Kobayashi, graduate students at Kobe University, contributed their time to the acoustic analysis in Study 3. Prof. Mark Senior at Konan University patiently read my papers and offered helpful suggestions. I am grateful to Mr. Hironori Terachi, editor at Shumpusha Publishing, for his careful management of the production stage. I wish to extend my special thanks to my husband and children for their understanding and support. I would not have completed my dissertation without support from all these quarters. Study 3 was supported by Grants-in-Aid for Scientific

Research [Scientific Research (C) No. 20520440]. This book has been published with the support of Konan University's Chubei Itoh Publication Grants for 2009.

Introduction

SPEECH science dealing with second language[1] (L2) speech learning, along with speech production and speech perception, has emerged as a growing area during the past few decades (e.g., Flege, 1988, 1995; Rvachew, 1994; Rochet, 1995). Many aspects of study have been explored in terms of phonological and phonetic factors relating to the differences between the sound systems of first language (L1) and L2. Since the 1980s, much of this research has attempted to characterize adult perceptions of phonetic distinctions that do not exist in the native language. Considerable work has been carried out to explore the effects of laboratory training on perception of non-native segmental contrasts (Strange and Dittmann, 1984; Best, McRoberts and Sithole, 1988; Yamada, 1993). Much attention in cross-linguistic work has been devoted to non-native segmental contrasts whereas suprasegmental (non-segmental) features have received less attention in empirical research on L2 acquisition (Bohn and Munro, 2007).

Furthermore, the teaching of English to speakers of other languages is treated as a particularly important subject, and Crystal (1997) emphasizes that English has become the major language of international communication. The nature of teaching English to speakers of other languages is changing. The language is now a "global language" (Crystal, 1997) or an "international language" (Jenkins, 2000; Mckay, 2002). The English pedagogy is called Teaching English as a Second Language[2] (TESL) and Teaching English as a Foreign Language (TEFL) in the United States, whereas in the United Kingdom it is called English Language Teaching (ELT).

As an English teacher in a university, I am especially concerned with the above two fields. The former, speech science, is indeed scientific where many empirical experiments have been performed in a laboratory setting. The latter is more practical compared with the former, and embraces a variety of methods, the acquisition of skills, curricular and other subfields (see Spolsky, 1999). The results of experiments in the first field have been applied to the latter, which is called Applied Linguistics. However, in reality, practitioners in the two fields do not seem to interact extensively with each other. We find few well-known scientists referred to in ELT studies. The reverse is also true. It is my view that both fields are indispensable and results from the surveys in each field should be more reflected in the other because both fields, especially L2 study in speech science and ESL/EFL/EIL, address the same subjects, namely language learners.

My goal in this research was to develop new techniques to modify the structure of the learner's phonetic system and to investigate the nature of prosody in speech perception and production. I have been particularly interested in acquisition of prosodic features of L2. Some (e.g., Jenkins, 2000) have claimed that suprasegmentals (prosodic features) are "simply not teachable and can only be acquired over time outside the classroom." Jenkins' claims are understandable because, as Flege (1988) states, interlanguage phonology studies indicate that with increased exposure to the target language, the phonological utterances of speakers will become more like the target language. However, why should they only be acquired outside the classroom? This might have been true three or four decades ago. My contention is that with the development of technology, suprasegmentals are teachable because learners can be exposed to a considerable amount of the target language even in the classroom. Claims of "teachable" or "unteachable" regarding suprasegmentals appeared to me to be anecdotal, and whether prosodic features are teachable or not became one of my research questions.

There was another question: the suprasegmental versus segmental debate. The emphasis in pronunciation instruction was once largely on segmentals, namely vowels and consonants. Then as the Communicative Language Approach to language teaching grew in popularity in the late 1970s, discoursed-based approaches and materials were used relatively. Suprasegmental features of language in discourse context became the main focus in teaching pronunciation. Today the debate seems to have settled down because researches have proven that inability to distinguish non-native sounds carries a high functional load and inability to distinguish prosodic features can have negative impacts on communication. Therefore, as Celce-Murcia, Brinton, and Goodwin (1996) state, "Today's pronunciation curriculum thus seeks to identify the most important aspects of both the segmentals and suprasegmentals, and integrate them appropriately in courses." How "appropriately" can they be integrated in courses? Should they be integrated simultaneously or studied separately in the course of training sessions? If the separate introduction of suprasegmentals and segmentals were proven effective, which one should be introduced first? There have been no prior empirical studies on the order of pronunciation training. I wished to know whether precedence should be given to suprasegmentals or segmentals in the early stages of L2 pronunciation training. These became my main research questions.

In the process of designing an experiment to examine the effect of the order of training, two experiments were designed as pilot studies. Study 1 examined the influence of model sounds on speech production.

The study aimed to compare students' own creative reading sounds with their imitating reading sounds. The results of the study revealed the effects of the model reading on the participants' production. Especially in prosody, the duration of sentences, consonants, and vowels approximated that of the model reading. As for the unfamiliar consonants, there was no significant change.

Study 2 investigated the effectiveness of applying low-pass filters to computer-assisted English pronunciation training. The possibility of applying low-pass filters to pronunciation training was of interest to me. Low-pass filters eliminate certain frequency components of sounds. Eliminating the high frequencies of speech results in muffled sound. Segmental content of speech is no longer intelligible though the prosodic information remains. As it highlights prosodic features, low-pass digital filtering of speech has been applied to various fields such as speech therapy, experiments with learning disability, and neuroscientific experiments. Masking segmental and semantic information by filters may affect auditory processing in the human brain. What would happen if the filtered training were applied to pronunciation teaching? Acoustic analyses by computer and listener judgments were used to determine how accurately the suprasegmentals were produced and to what extent they contributed to foreign accent. The analysis of variance (ANOVA) was used and it revealed a significant difference between the low-pass group and the non-filtered group, especially in the results of acoustic analyses.

As for subjective evaluations, both groups showed improvement and the difference was not significant. The results suggest that English prosody training using low-pass filters contributed to acquire the accuracy of pronunciation of the language. The training without low-pass filters was also effective to improve English prosody according to raters' subjective evaluations.

With the results of Study 1 in mind, model sounds were used when designing the main study, Study 3. Concerning low-pass filters, the results of Study 2 showed little difference especially in the subjective evaluations, and they were not adopted to the main study. Study 3 considered the priority of prosodic features over individual sounds in second language acquisition because theory has not been sufficiently supported by empirical research in this area. Sound data of the pre/mid/post-tests were collected using original software. Sentence or phrase duration and F0 ranges were measured using Praat, acoustic analysis software. Data were also judged by four raters as to whether they sounded natural as English. The results were analyzed in ANOVA. Findings show that the prosody-first group, which practiced prosodic features at the earlier stage of their training sessions, achieved by far the highest results in both acoustic analyses and subjective evaluations.

Chapter 4 discusses the findings of all three studies. Among these, the most significant finding may be that of Study 3 which revealed the importance of the order of training. The apparent priority of training prosodic features with high-interaction feedback over individual sounds was observed. More investigations are needed to support this finding, which will shed further light on our understanding of the process of speech perception and production.

The final chapter concerns the choice of models for teaching English pronunciation from the perspective of a non-native instructor of the English language. The issue may not be related to the three experiments of the book directly. However, choosing models plays a significant part in teaching or designing a piece of software for the pronunciation of English because of the world-wide use of the language. The choice of a proper model is related to the conceptual aspects underlying the teaching of pronunciation. The technical aspects of pronunciation teaching were dealt with in Chapters 1 to 4 in the book. Both aspects are indispensable to successful pronunciation teaching.

Notes

1 In the field of speech science, "second language" means a "non-native foreign language."

2 In this case, "second language" means "taught to those for whom English had an internal function in their county" (Jenkins, 2000).

Chapter 1 (study 1)

The Influence of Model Sounds on the Speech Production of Japanese Learners of English

1.1. Introduction

1.1.1. Stress-timed vs. mora-timed?

ONE of the difficulties in learning a foreign language is in mastering the prosodic aspects of the target language. For example, a native Japanese speaker generally has difficulty with stress and timing when speaking English. On the other hand, it is widely known that for a native English speaker it is hard to learn the mora-timing of the Japanese language. These two are very different languages, using different sounds in different ways. The stress-timed rhythm of the English utterance with the related obscuration of weak syllables is the prime distinguishing feature of the language's pronunciation. As Gimson (1981) noted, "For all learners, accentuation must provide the foundation of which any pronunciation course is built." The two languages differ greatly in prosodic realization, including temporal organization. The typological categorization of timing systems is rooted in the idea that temporal organization is based on some unit of timing, and Japanese is said to be mora-timed, whereas English is stress-timed (Dauer, 1983). In English sentences, the duration is alternately stretched and reduced, interacting with the other two correlates, namely, fundamental frequency and amplitude. The unit of English timing is the stress foot, that is, a string containing a stress accent followed by zero or more unstressed syllables. On the other hand, the unit of Japanese timing is the mora, a syllabification unit. As Ueyama (1996) noted, the duration of each mora is equal, abstracting away from the phrase-final lengthening, and the prosodic distinctions of Japanese are mainly conveyed by fundamental frequency.

However, some researchers are against the concept of stress-timing. Ladefoged (1982) describes it as only a "tendency." Roach (1991) considers that if stress-timing operates at all, it only occurs in very regular, formal speech. Regarding teaching pitch movement, Jenkins (2000) is doubtful. "Even if it were possible to teach pitch in the classroom, I do not believe that the use of "native speaker" pitch movements matters very much for intelligibility in interactions among NBESs[1]." Nevertheless, I take a different view of teaching prosodic features. Even if we don't "teach" them in a strict sense, we should demonstrate them to learners otherwise they may have no other chance to recognize them.

The audio-lingual method is less popular today because of its reliance on drills and habit-formation. Structural patterns in dialogues about everyday situations are imitated and drilled. These monotonous tasks are less interesting to learners who feel the need for more creative work in speech productions. More humanistic approaches are of course welcome but teaching pronunciation has a different story. To acquire the sound system of a different language, imitation and repetition of the target language are necessary.

1.1.2. Purpose of the study

In this research, I have investigated how learners' utterances would differ after they listened to the model sounds. The pre-listening sounds and post-listening sounds were compared with model sounds. The duration of sentences, some selected consonants, and phrase-final vowels, and fundamental frequency were measured by a computer speech analyzer.

The auditory impressions of four professional English teachers were also examined by way of subjective evaluation. The results will be compared with computer assessment and discussed later.

My research questions in this study were as follows:

1 Is there any difference in the speech production of participants after they listened to model sounds and reproduce the same sentences? The following items will be examined.
   - Fundamental frequencies
   - Duration of sentences
   - Duration of some consonants (fricatives)
   - Duration of phrase-final vowels

2 Do the results of subjective evaluation relate to those of the figures obtained through experiment?

1.2. Experiment

1.2.1. Materials

Test materials comprised of eight sentences and two minimal pairs of words. In the pairs, the consonants /f, h, z, dz/ were embedded (See Table 1.1, No. 9 and No. 10). The consonant /r/ and /l/ were inserted in the same context to ascertain whether or not the participants had known the difference between these two consonants before the test, and after the test whether or not their pronunciation changed (See Table 1.1, No. 6 and No. 7). The voiceless dental fricative /θ/ or the voiced form of the same consonant /ð/ were embedded because these two are unique to English sounds (See Table 1.1, No. 1 to No. 4, No. 6, and No. 8). Sentence No. 5

Table 1.1. *The test materials*

No.	Sentence
1	He went over the path.
2	They thought about it.
3	We've fired them.
4	Repeat the word.
5	I knew it was wrong.
6	He didn't collect the papers.
7	He didn't correct the papers.
8	Put all these things in the bag.
9	food/hood
10	cars/cards

was chosen for the test materials to examine the consonant /r/ and word final vowel in "wrong." The main purpose of the study is not to compare the isolated consonants but to focus on the change of prosodic features. Yet, if any changes were found in the above-mentioned consonants, it would be of more interest because no instruction in pronunciation was given to the participants.

1.2.2. Speaker

A male speaker of standard British English recorded the test items.

1.2.3. Participants

The 20 participants (12 female, 8 male) who participated in the experiment were native speakers of Japanese spoken in the Kansai area and students at Konan University in Kobe. They belonged to different faculties of the university. No one had spent more than two months in an English speaking country. Their ages ranged from 19 to 22. They reported normal hearing and vision.

1.2.4. Procedure

At the moment, while our computer-based listening and pronunciation software is still under development, nevertheless a computer-based pretest[2] for this software was made and given to the participants. In this pretest, forty sentences of each participant were recorded. The first twenty sentences were recorded without model sounds and the latter twenty were recorded after listening to model sounds.

The pretest was performed in a CALL room at Konan University. At the beginning of the pretest, I taught the participants how to use the software.

From No. 1 to No. 20, the participant was asked to read the text in the top box and record his/her voice by cricking the icon as seen in Figure 1.1. This pattern lasted until No. 20. Then from No. 21 to No. 40, the participant listened to the model sound first and then recorded his/her voice trying to repeat the sounds in the same way as seen in Figure 1.2.

In the test sentences, No. 11 to 20 and No. 21 to 30 were identical (See Appendix Table 1). These twenty sentences were analyzed in

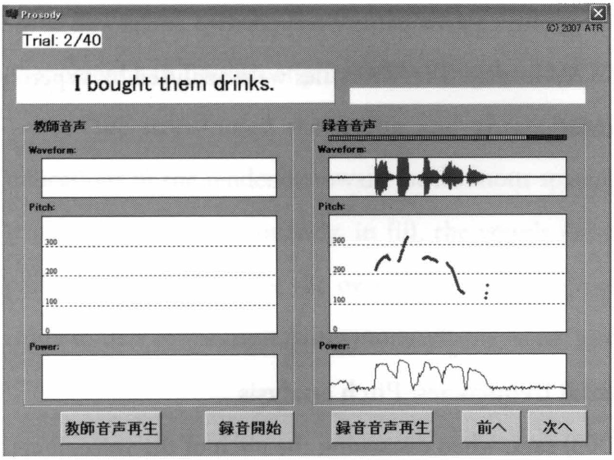


Figure 1.1. *The interface of No. 2. In trial 1 to 20, if "教師音声再生" icon were clicked, no voice would be heard. The participant was asked to read the sentence freely.*

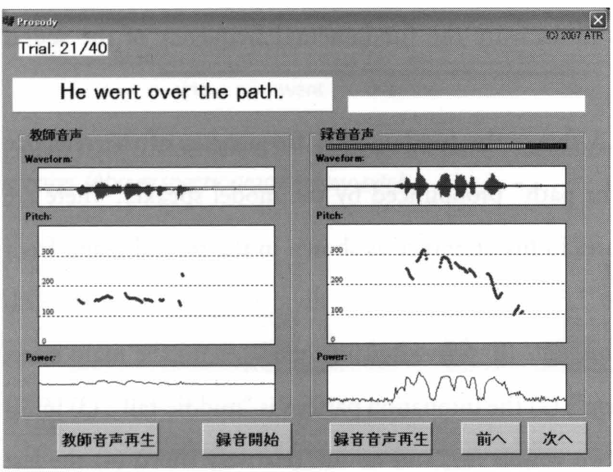


Figure 1.2. *The interface of No. 21*

this study. The voices of the participants were recorded and stored in the computer as the WAV format. The WAV files were analyzed by a speech analyzer, WaveSurfer[3].

1.3. Results

1.3.1. Fundamental frequencies: Pitch analysis

As Ladefoged (2003) says, when discussing the pitch of the voice, it can usually be said to be the rate at which vocal fold pulses recur, and thus the fundamental frequency of the sound wave. Tone and intonation are manifested by pitch. You cannot literally measure the pitch of a recorded sound but you can measure the fundamental frequency of the sound wave.

Figure 1.3 shows the fundamental frequencies of the sentence "He went over the path" pronounced by the model speaker. There are clearly four features in his utterance as shown in the second graph from the top of Figure 1.3: (1) the sentence is divided into three parts with two pauses ((i) /hɪwɛnt/ (ii) /oʊvɚðə/ (iii) /pæθ/), with the main stress occurring on "went"; (2) the intonation pattern is "middle-fall"; (3) in (ii) /ooʊvɚðə/, the two vowels /ɚ/, and /ə/ are relatively short; (4) the last vowel /æ/ is relatively long.

The third feature indicates the stress-timed shortening, a typical effect in English speech. The fourth feature also shows the well-known phenomenon in English of phrase-final lengthening[4].

The stress-timed shortening and phrase-final lengthening are indications of the tendency toward isochronous spacing of prosodically strong syllables. Consequently, in (ii), the vowels /ɚ/ and /ə/ are compressed in order to make the overall duration of the phrase "over the" closer to that of the contrasting monosyllable word "path" in (iii).

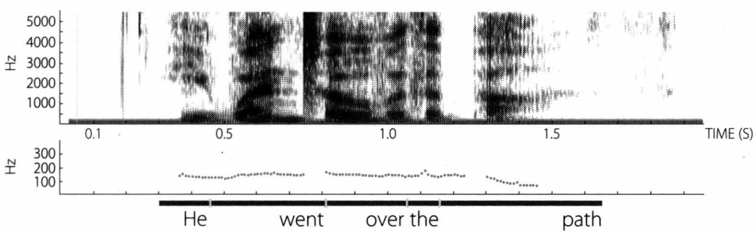


Figure 1.3. *The sentence "He went over the path." Pronounced by a model speaker. (Above: spectrogram. Below: pitch.)*

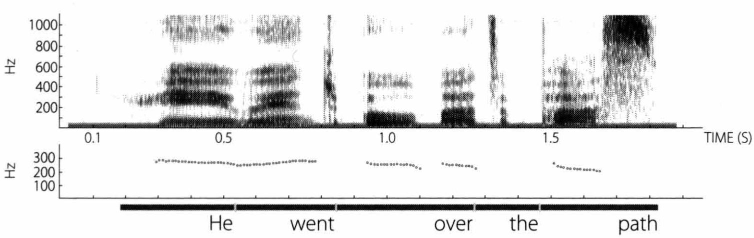


Figure 1.4. *The same sentence as shown in Figure 1.3. It is pronounced by a Japanese participant. (Above: spectrogram. Below: pitch.)*

Figure 1.4 shows a typical example of Japanese students. In her speech, the fundamental frequencies show that (ii) /oʊvɚðə/ is not divided as one segment but as two segments.

As for the comparison of fundamental frequencies between the model sound and 20 participants, I counted the number of segments to look for any difference in segmentation between the before-listening (BL) model sounds and the after-listening (AL) model sounds. If the model (M) speaker's reading of "He went over the path" is divided into three segments (as mentioned before), how are the results of the participants? The following is the results of the segmentation. If a participant pronounced the sentence just as the model speaker, that would be counted as Segment 1 = 1, Segment 2 = 1, Segment 3 = 1. If he pronounced the sentence differently such as //He/went// //over/the// //path//, the segmentation would be counted as Segment 1 = 2, Segment 2 = 2, Segment 3 = 1.

Table 1.2 shows the results of counting the number of segments by 20 participants. As for segmentation, there was no difference between "Before" and "After" in Segment 3. There was not much difference in Segment 1. Yet, in Segment 2, there was some difference. To make the difference more obvious, see Figure 1.5.

As indicated in Figure 1.5, the segmentation in Segment 2 reduced appreciably. A possible reason for this phenomenon is that the Japanese participants tended to separate into two or three parts when they pronounced "over the" because they are affected by the timing system of

Table 1.2. *The results of the segmentation of 20 participants*

	Segment 1	Segment 2	Segment 3
Model	1	1	1

	BL			AL		
Participant	Segment 1	Segment 2	Segment 3	Segment 1	Segment 2	Segment 3
P1	1	2	1	1	1	1
P2	1	3	1	1	2	1
P3	1	2	1	1	3	1
P4	1	3	1	1	3	1
P5	2	3	1	1	3	1
P6	2	3	1	2	3	1
P7	1	3	1	1	3	1
P8	2	3	1	1	2	1
P9	1	2	1	1	1	1
P10	1	3	1	1	1	1
P11	1	2	1	1	1	1
P12	1	3	1	1	1	1
P13	1	3	1	1	1	1
P14	1	3	1	1	2	1
P15	1	3	1	1	2	1
P16	1	1	1	1	1	1
P17	1	3	1	1	1	1
P18	1	3	1	1	3	1
P19	1	2	1	1	2	1
P20	1	1	1	1	1	1
M	1.15	2.55	1	1.05	1.85	1

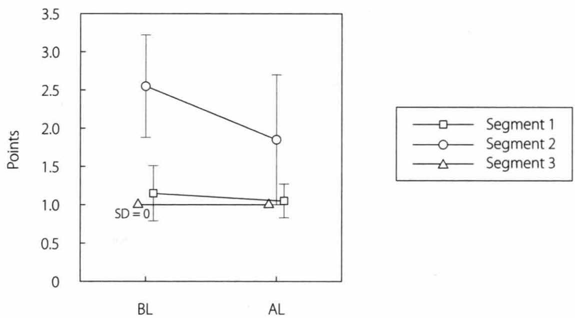


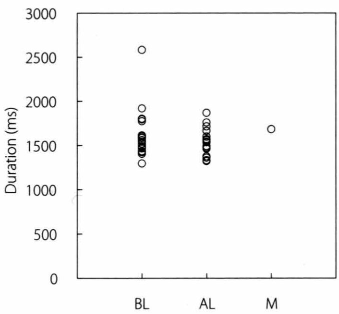
Figure 1.5. *The comparison of mean averages of the number of segments between "BL" and "AL." Error bars indicate ± one standard deviation.*

their first language but they could change their pronunciation after they listened to the English model sounds. If they can imitate model sounds, they will read "over the" in a weak and fast way like the model. I predicted the results of Segment 2 but as for Segment 1, the results were interesting and I had not predicted them. I had thought most participants would pronounce Segment 1 in two parts as "He/went" before they listened to the model reading, yet only three participants out of 20 did. In reality, most students pronounce it as one segment. In this case, we could say that most of the participants are not so strongly influenced by their first language. Comparing Segment 1 with Segment 2, both segments consist of two words but there was a variety of segmentation in Segment 2. Perhaps because the word "over" in Segment 2 is a two-syllable word, the total number of syllables in Segment 2 is three whereas the total number of syllables in Segment 1 is two. There might be a considerable difference between two syllables and three syllables in one rhythm group in perception and production.

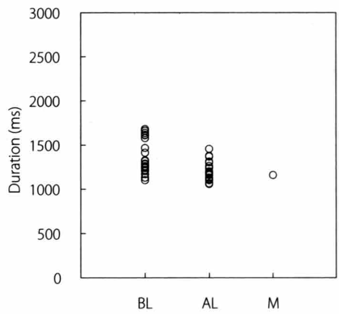
1.3.2. Duration of sentences

Table 2 and 3 in Appendix show the results of measuring the duration of eight sentences pronounced by 20 participants. After listening to model readings, how did the participants change the duration of the sentences? The eight graphs in Figure 1.6 were made according to Table 2 and 3 in Appendix. The graphs show the distribution of the sentence duration before/after listening.

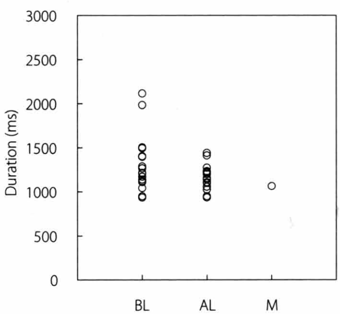
These graphs visually indicate that the duration of the sentences pronounced by the participants were influenced by model sounds. There is a tendency in BL sentences for the pronunciation to be rather longer than the model's. AL sentences tend to be shorter than BL sentences and to approximate to M sentences. There were some sentences in BL which were far longer than the models. For example, the highest dot (duration = 2583 ms) in Duration of Sentence 1 in Figure 1.6. This phenomenon shows the participant's uncertainty in pronunciation. After he listened to the model sound, the duration of the same sentence produced by him became 1604 ms which was similar to the duration of the model sound (1685 ms).



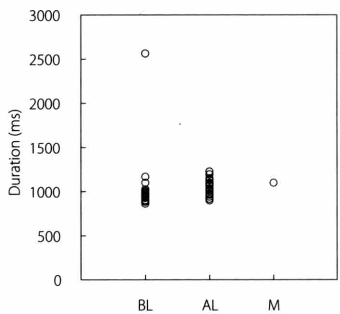
(a) Sentence 1 "He went over the path."



(b) Sentence 2 "They thought about it."

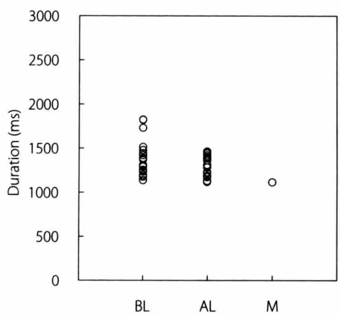


(c) Sentence 3 "We've fired them."

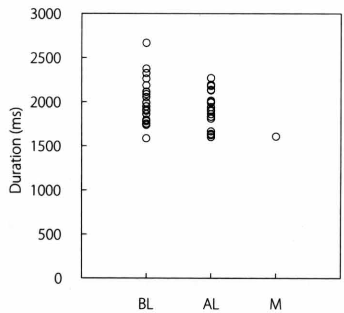


(d) Sentence 4 "Repeat the word."

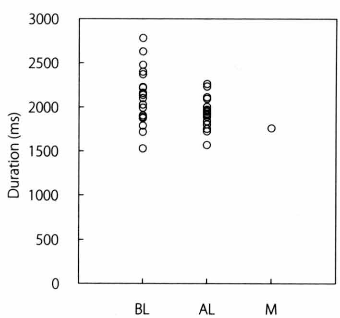
Figure 1.6. *The comparison among the duration of BL sentence, AL sentence and M sentence*



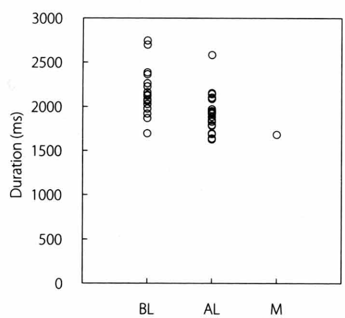
(e) Sentence 5 "I knew it was wrong."



(f) Sentence 6 "He didn't collect the papers."



(g) Sentence 7 "He didn't correct the papers."



(h) Sentence 8 "Put all these things in the bag."

1.3.3. Duration of consonants (fricatives)

The following is the results of measuring the duration of some consonants appeared in the test sentences. Fricatives were chosen for observation because they are conspicuous in the sound spectrogram.

Table 4 in Appendix is the results of measuring the duration of consonants prior to stressed vowels. Table 5 in Appendix shows the results of measuring the duration of consonants prior to unstressed vowels. In the English language, consonants that are placed prior to stressed vowels tend to be longer than when they are prior to unstressed vowels. Table 1.3 is made according to the results of Table 4 and 5 in Appendix to prove this phenomenon.

Table 1.3 shows the following. Even without listening to model sounds, the duration of consonants prior to stressed vowels produced by Japanese participants (mean average = 111.14 ms) is already long compared with that prior to unstressed vowels (mean average = 47.33 ms). After listening to the model reading, long consonants became longer (111.14 ms to 124.67 ms) and short consonants shorter (56.78 ms to 47.38 ms).

To investigate more precisely, the six graphs in Figure 1.7 were made based on Table 4 and 5 in Appendix. These graphs indicate that after listening to the model reading, the duration of consonants produced by Japanese participants tend to approximate to the model sounds.

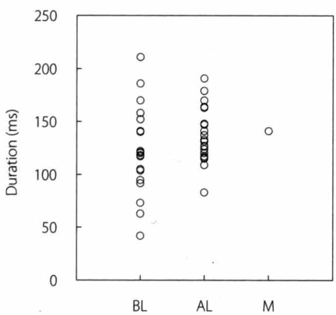
Table 1.3. *Comparison of the duration of consonants between stressed and unstressed*

(a) The duration of stressed consonants (ms)

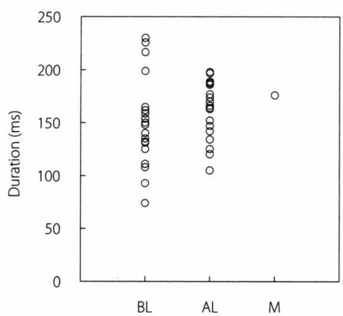
	BL	AL
θ	121.76	137.05
f	149.55	160.45
ð	62.1	76.5
M	111.14	124.67

(b) The duration of unstressed consonants (ms)

	BL	AL
θ	58.6	61.15
θ	65.2	51.4
θ	46.55	29.6
M	56.78	47.38

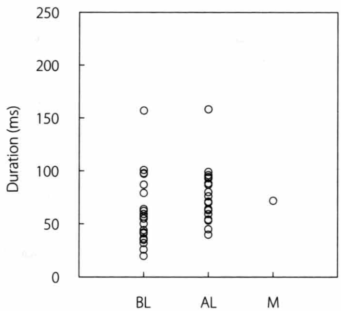


(a) Sentence 2 /θ̣/ in "thought"

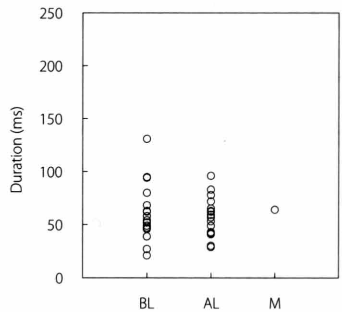


(b) Sentence 3 /f/ in "fired"

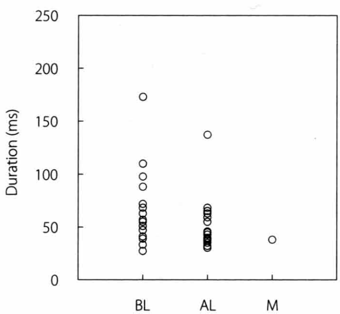
Figure 1.7. *The distribution of the duration of consonants in BL, AL and M 1*



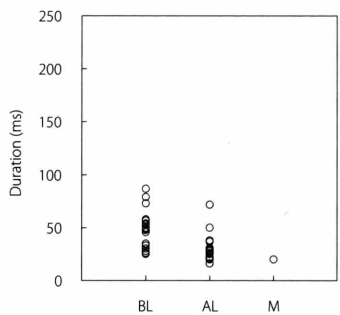
(c) Sentence 8 /ð̣/ in "these"



(d) Sentence 1 /ð/ in "the"



(e) Sentence 4 /ð/ in "the"



(f) Sentence 8 /ð/ in "the"

Figure 1.7. *The distribution of the duration of consonants in BL, AL and M 2*

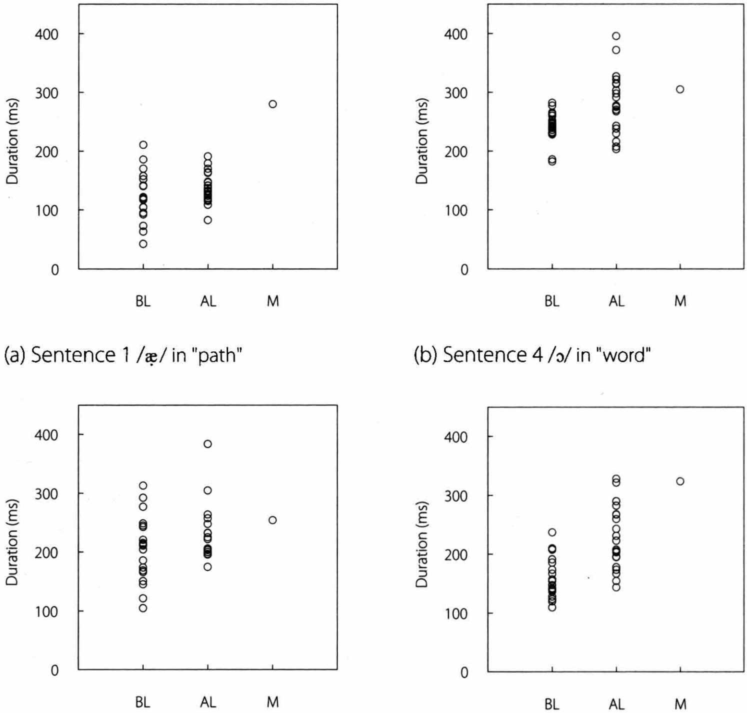
1.3.4. Duration of phrase-final vowels

Table 6 in Appendix shows the results of measuring the duration of phrase-final vowels. As noted earlier, English vowels tend to increase in the duration when they occur at the end of a phrase.

Regarding the mean averages of the duration of vowels in Table 6 in Appendix, the duration of every vowel after listening to the model sounds increases. The reason for this phenomenon can be related to the fact that in most cases, the duration of the model is longer than that of the Japanese participants. There is one exception found in Table 6 in Appendix (See the bold figure **292** ms). Judging by the participant's utterance in this case, she was not very confident about pronouncing the word. After listening to the model reading, the duration became quite similar to the model's. (Participant 11 264 ms, Model 254 ms).

As for Figure 1.8, the four graphs indicate that if the duration of a model vowel is longer than that of participants seen in Graph (a), (b) and (d), AL consonants tend to be longer than BL. If the duration of the model vowel is not so long as in Graph (c), the duration of AL consonant would not increase. In Graph (c), we can observe one exceptionally high dot (identical figure **384** ms in Table 6 in Appendix). Judging from listening the recording sound, the participant seemed to pronounce the word "wrong" exaggeratedly after she listened to the model reading.

In this section, the duration of test sentences, some consonants



(a) Sentence 1 /æ̞/ in "path"

(b) Sentence 4 /ɔ/ in "word"

(c) Sentence 5 /ɔ̞/ in "wrong"

(d) Sentence 8 /æ̞/ in "bag"

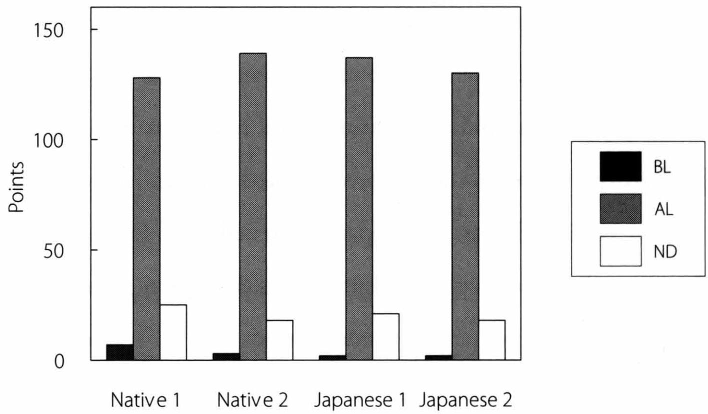
Figure 1.8. *The duration of phrase-final vowels in BL, AL and M*

and vowels has been measured and the comparison was made between BL, AL and M. In consequence, the results demonstrate the significant influence of model sounds. AL sounds tend to approximate model sounds. In this experiment, participants were asked to listen to the model reading at least once and most participants listened once for each sentence before recording their voice. Judging from this, we may conclude that only one or two exposures to model sounds has an influence on speech production of participants. In the next section, auditory impressions by four raters will be stated. I will compare the results of the speech analyzer and human ratings later.

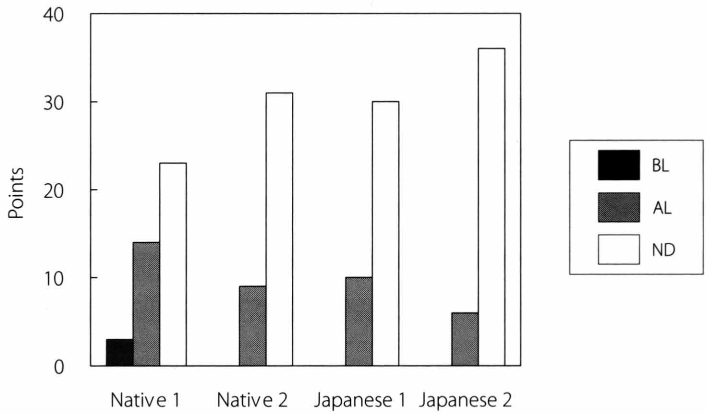
1.3.5. Auditory impressions: subjective evaluations

Participants' recorded productions were evaluated on a three-point scale by a total of four teachers of English. Two of them are native speakers of English and the other two are Japanese. Rating sessions were done individually. Raters listened to the participants' BL sound and AL sound then they were asked to judge which sounded more naturally as English. The three rating scales were BL, AL, no difference (ND). If a rater feels AL sounds to be more natural, one point would be added to AL. As there were 20 participants and each participant's recorded productions were 10 pairs of AL and BL, each rater listened to 200 sound files on computer.

Results of auditory impressions are shown in Table 7 in Appendix and Figure 1.9. Regarding sentences (Sentence 1–8 in Table 7 in



(a) Sentence 1 to 8



(b) Minimal pairs

Figure 1.9. *Auditory impressions by raters*

Appendix and Figure 1.9, Graph (a)), AL sounds were rated quite highly compared to BL. On the other hand, the minimal pairs such as food/hood and cars/cards didn't show much difference in BL and AL. Since no instruction concerning pronunciation was given during the experiment, it must have been difficult for participants to perceive the difference in the unfamiliar consonants of these pairs. As for prosodic features, however, such as rhythm or intonation, all raters felt AL sounds to be more naturally as English.

Table 1.4 and Figure 1.10 show the individual differences among the participants. They tend to be masked by averaging but a variation between individuals is important. There were a group of participants who were fairly good at adjusting their pronunciation to the model reading such as Participant 1, 14, 17, 19, and 20 (whose scores were over 30). On the other hand, there were a few participants whose pronunciation didn't change significantly as shown in Participant 5 and 6. This is because their pronunciation of English was already natural before practicing. In the present experiment, there happened to be no participant whose accent of English is strongly influenced by the first language.

1.4. Discussion and concluding remarks

The results of the study revealed the significant effects of the model read-

Table 1.4. *Individual difference of ratings*

Participant	Sentence 1–8			Sentence 9, 10		
	BL	AL	ND	BL	AL	ND
P1	0	31	1	0	0	8
P2	0	27	5	0	0	8
P3	0	24	8	0	0	8
P4	0	28	4	0	2	6
P5	0	20	12	0	2	6
P6	0	18	14	0	1	7
P7	0	30	2	0	4	4
P8	5	23	4	0	5	3
P9	1	24	7	1	2	5
P10	0	30	2	0	1	7
P11	0	29	3	0	2	6
P12	2	29	0	0	2	6
P13	0	26	6	0	4	4
P14	0	32	0	0	4	4
P15	1	29	2	1	4	3
P16	3	23	6	1	4	3
P17	1	31	0	0	2	6
P18	0	27	5	0	2	6
P19	0	31	1	0	3	5
P20	1	31	0	0	1	7

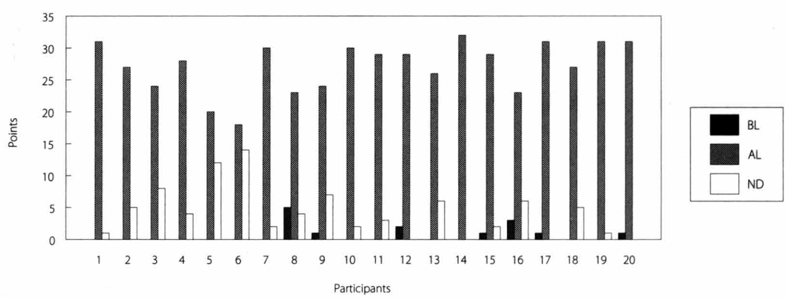


Figure 1.10. *Graphic form of Table 11*

ing on the participants' production. Especially in prosody, the duration of sentences, consonants, and vowels approximated that of the model reading. As for the unfamiliar consonants, there was no significant change between BL and AL. That means the participants could recognize the prosodic features by listening to the model reading only once or twice and produce the sound in a similar way. Yet, regarding unfamiliar consonants, they could not recognize the difference and in consequence, they couldn't produce them. The pedagogical implication of this is that both the suprasegmentals (mainly rhythm, intonation) and segmental features (vowels, consonants) are indispensable in teaching pronunciation of a foreign language.

Regarding the approximation to the model reading that the participants showed in the study, it is too early to conclude that they acquire or improve the prosody of the target language because they were exposed to the model reading only once or twice for each stimulus. How will the participants change their production if they participate in further pronunciation training? More extensive experiments including studies of retention would contribute to our understanding.

In addition to the quantifiable results, I concluded from the experiment that learners' confidence was key to success in learning a foreign language. As I stated earlier, some researchers have doubts about "teaching" suprasegmentals. Even if their contention were right in theory, nevertheless we can still introduce them to learners. Then learners may gain confidence in producing the target language.

I used computer software for this study and I was developing pronunciation training software at the moment. However, this is not to say that computer-assisted training is superior to other approaches. A question might arise as to whether non computer-based training approaches such as traditional teacher-led instruction would be as or more effective. One might attempt to compare these approaches experimentally. However, such a comparison is irrelevant because there are numerous elements that make up an approach. Simply using the same materials for the same period of time would not provide a basis of comparison. In my opinion, they should be complements each of the other.

Notes

1 NBES is an abbreviation that Jenkins created. It stands for "non-bilingual English speaker."

2 The basis of the software was provided by ATR (Advanced Telecommunications Research Institute International). We are permitted to use it on a research basis.

3 WaveSurfer is an open source tool for sound visualization and manipulation created in the School of Computer Science and Communication, The Royal Institute of Technology in Sweden.

4 In speech, slight decelerations and pauses frequently occur at the ends of phrases.

Chapter 2 (study 2)

The Effectiveness of Low-Pass Filters in English Pronunciation Training

2.1. Introduction

2.1.1. Low-pass filters

BY definition, a low-pass filter is a circuit offering easy passage to low-frequency signals and difficult passage to high-frequency signals. It eliminates certain frequency components of sounds. Such a filter was originally used to direct high frequencies to a tweeter speaker for music or speech. Eliminating the high frequencies of speech in a signal that sounds muffled. Segmental content of speech is no longer intelligible though the prosodic information remains. As it highlights prosodic features, the low-pass digital filtering of speech has been applied to various fields such as speech therapy, experiments with learning disability, and neuroscientific experiments.

For instance, low-pass filtered speech effectively separates children with learning problems from those who are normally achieving (Keith & Farrer, 1981). Filtered word testing is one subtest of the SCAN-C auditory test battery (Keith, 2000). Hearing loss and processing problems should be evaluated as separate issues, although they may be closely related. Children with learning disabilities often show signs of auditory processing difficulties. According to Joudry, high-frequency filtered music seems to be highly effective in the treatment of children with hearing problems (Joudry, 2004). A psychophysical experiment was undertaken to investigate whether male and female listeners differed with respect to which frequencies were important in the perception of a male voice as a natural sound object (Hunter, Phang, Lee & Woodruff, 2005).

Children with hypersensitivity suffer from many stressful and disturbing symptoms; they may over-react to common noises, and be distressed by classroom sounds. Filtered sound training for those children has resulted in significant improvements in their hearing ability and behaviors. This suggests that masking segmental and semantic information by filters may affect auditory processing in the human brain. If we adopt filtered training to language teaching for normal-hearing learners, what would happen? This was the first question I had upon reading the papers mentioned above.

2.1.2. Application of low-pass filters to pronunciation training

Perception and production training with low-pass filters has already been adopted as a therapy for hearing-impaired people. It was originally created by Petar Guberina (Guberina, 1972, 1976), a Croatian psycholinguist who was working on problems of perception and production with hearing-impaired people as well as people with normal hearing. His work in this area is based on what he called verbo-tonal theory. This theory provides us with interesting ways of thinking about perception and learning in general. Guberina's notion is that deafness can be thought of not so much as a condition caused by a physical defect but as a way of organizing the world which differs from strategies that have been learnt by people who are not deaf. Guberina's work has subsequently been augmented and reframed through application of the thinking of Jack Derrida (1982), Pierre Bourdieu (1991, 1995), Ann Freadman (1994), and Ania Lian (2003). Lian et al created MMExplore, a system designed to enable the exploration of authentic text in a variety of ways with emphasis on development skills. It enables the use of electronic low-pass digital filtering of speech to highlight intonation patterns.

These attempts at using low-pass filters in language training have been made over the last few decades, but are still not common in the field of language teaching. The emphasis in foreign language teaching is on achieving communicative effectiveness. Many learner-centered communicative approaches aim at enabling learners to successfully communicate in the target language. Pronunciation is an obvious component of communication and serious pronunciation problems are known to hamper communication or put learners at a social and professional disadvantage (Munro & Derwing, 1995). Recent studies have shown that tailor-made training is effective in improving perceptive and productive skills (Akahane-Yamada et al. 1998, Moyer 1999, Hardison 2003). However, in reality, the time that is generally available for pronunciation training in traditional classroom instruction has remained relatively limited in Japan where the grammar-translation method played an important role for a long period of time in importing necessary knowledge and information for modernization. Although the old educational paradigms have shifted to communicative approaches, pronunciation training is still peripheral. Computer-assisted pronunciation training might improve the current situation.

Low-pass training, as mentioned above, already exists in language training but mainly for hearing-impaired people and is highly limited for normal-hearing learners. Regarding the effectiveness of low-pass filters, it has not been empirically proven. Low-pass filters are used for language training and speech therapy on the assumption that they are factually effective. If the system of perception of speech sounds of the hearing-impaired differs from that of normal learners as Guberina mentioned, we should be more prudent about adopting the filter training for the non-disabled group. The research reported in this paper is aimed at examining the effectiveness of using low-pass filters in English pronunciation training for normal-hearing learners. The details of these research questions follow below.

2.1.3. Purpose of the study

This study is intended to investigate the effectiveness of applying low-pass filters to computer-assisted pronunciation training of English. If the application is found to be effective to acquire the prosody of the target language, low-pass filtered sounds can be used for designing efficient pronunciation training programs.

My research questions in this study were as follows:

1 Is there any difference in speech production between the group of participants who attended 10 sessions of training with electric low-pass filtered sounds and the controlled group who did the same training without filtered sounds? Did the low-pass group become more accurate in production than the non-filtered group, or vice versa? Were both improved?

2 How did the participants feel after they finish the 10 training sessions?

2.2. Experiment

2.2.1. Materials

The pretest for this experiment was basically the same as I designed and used in my previous study (Iba, 2007). The total number of stimuli in the pretest was 40 and 10 of them were compared to the same stimuli in the posttest which included 80 stimuli. When I designed the pre/post tests, I made six groups (Group A to Group F) of sentences and words (Table 2.1).

Selection of the stimuli for testing followed these guidelines: 1) familiar vocabulary, 2) structural variety, 3) sustained phonation which may provide a visually obvious display of pitch contour, 4) relatively short sentences to facilitate easy production, and 5) sets of minimal pairs which include consonants that are difficult to produce for Japanese learners of English.

As for the pretest, the four groups of stimuli (A, B, B, D) are selected and set in the software which I used in the previous experiment (study 1). The order of the stimuli is A, B, B, D. The stimuli in the first two groups are displayed only in the text style on the computer display whereas the latter two groups are both in the text style and sounds (See Figure 2.1). The participants were required to read and record the stimuli for the first two, and then for the latter two, they were asked to read the text, listen to the model voice, and record the stimuli.

Table 2.1. *The stimuli of the pre/posttests*

(A)

No.	Sentence
1	Thank you very much for everything.
2	I bought them drinks.
3	What's she saying?
4	They are all afraid.
5	What would you like to do?
6	Give it to him.
7	This is better than that.
8	I know it's true.
9	think/sink
10	right/light

(B)

No.	Sentence
1	He went over the path.
2	They thought about it.
3	We've fired them.
4	Repeat the word.
5	I knew it was wrong.
6	He didn't collect the papers.
7	He didn't correct the papers.
8	Put all these things in the bag.
9	food/hood
10	cars/cards

(C)

No.	Sentence
1	Look at the train.
2	Don't disturb them while they are praying.
3	Humpty Dumpty sat on a wall.
4	Humpty Dumpty had a great fall.
5	All the king's horses and all the king's men.
6	Couldn't put Humpty together again.
7	What a wonderful life he lived!
8	How beautiful you are!
9	ban/van
10	deaf/death

(D)

No.	Sentence
1	Will you read it again?
2	There's a crack in the glass.
3	I think I'll take a bath.
4	Let's keep in touch.
5	That class is easy.
6	He is on vacation.
7	I saw a flash of lighting.
8	Is it true that he is ill?
9	think/sink
10	clothe/close

(E)

No.	Sentence
1	They're leaving next week on a trip around the world.
2	What's the matter with you?
3	Why won't you believe me?
4	Richard and Christine won the state lottery!
5	Is that what you want to say?
6	Are you criticizing me?
7	What are you going to do tomorrow?
8	Don't worry; I'll do it for you?
9	year/ear
10	woos/ooze

(F)

No.	Sentence
1	What kind of person is willing to send his children to wars?
2	What would you like to have for dinner?
3	My breakfast is always bread and butter.
4	I just can't wait.
5	How many times do I have to tell you this?
6	It won't be long.
7	I have to admit that I was a little drunk.
8	I was wondering if you could babysit tomorrow night.
9	I thought you were a normal person.
10	I think I'm coming down with something.

Note. In this experiment, six groups of stimuli (A, B, C, D, E and F) were recorded but only group B was used for analysis. This is principally because groups A, C, D, E and F were recorded for further experiments. In addition, participants would not have focused on pronouncing the target 10 stimuli in B. Consequently, they would read the stimuli naturally.

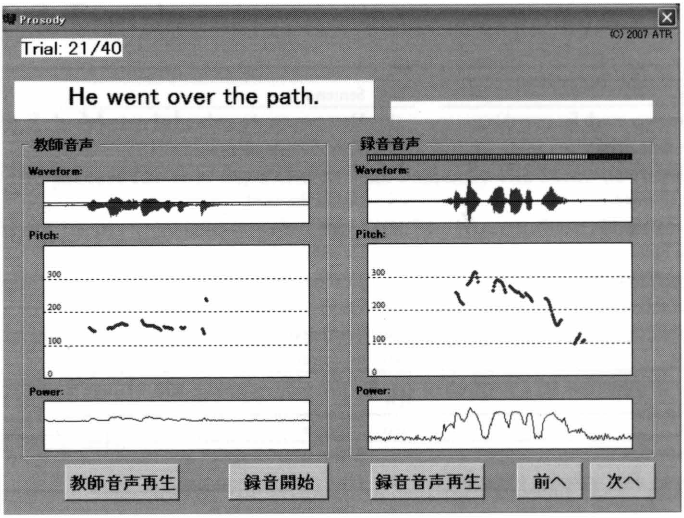


Figure 2.1. *The interface of pre/posttest*

Regarding the posttest, the procedure is the same as the pretest but it includes more groups of sentences (Table 2.2).

I designed the six groups of stimuli in order to compare the results of this experiment with further research. In this experiment, I compared the following.

1 Pretest B under the condition of Read (RD)

2 Pretest B under the condition of Listen and Repeat (LR)

3 Posttest B under the condition of Read (RD)

4 Posttest B under the condition of Listen and Repeat (LR)

Table 2.2. *The order and conditions of stimuli in the pre/posttest A–F are groups of the stimuli (See Table 2.1).*

	Pretest				Posttest							
Order	A→	B→	B→	D	A→	B→	D→	E→	B→	C→	E→	F
Condition	RD	RD	LR	LR	RD	RD	RD	RD	LR	LR	LR	LR

Note. RD for "Read," LR for "Listen and Repeat."

2.2.2. Speakers

In the pre/posttests, a male speaker of standard British English recorded the test items. Between these two tests, participants attended 10 sessions of pronunciation training. As there are 20 stimuli in each session, 200 stimuli were recorded by two male speakers and two female speakers. All of them are professional recorders of standard American English.

2.2.3. Participants

A total number of 13 native speakers of Japanese (10 female, 3 male) volunteered to participate in this study. All of them were undergraduate students at Konan University in Kobe. They belonged to different faculties of the university. None had spent more than two months in an English speaking country. Their ages ranged from 19 to 22. They reported normal hearing and vision. All of them were motivated to improve their production of English.

2.2.4. Procedure

A pretest-posttest design was used to measure the effects of one month's training (10 sessions of about 40 to 50 minutes each) using computerized visual displays of pitch contours and wave forms as feedback (See Figure 2.1). The same software was used for the pretest, 10 training sessions and the posttest. Its basis was provided by ATR (Advanced Telecommunications Research Institute International). Users can customize it by inputting the stimuli. This time 40 stimuli for the pretest, 200 stimuli for the training sessions and 80 stimuli for the posttest were put into the software.

The software was installed into 10 computers in a self-study room at the university. Participants were asked to come to the room at any time during the training period. For the first time, they were asked to read instructions about using the software and took the pretest by computer. Their voices were automatically recorded and stocked in the server. During the training period, some of the participants came daily, finished the training sessions and took the posttest relatively early. Some of them came to the room as regularly as twice a week and others came quite irregularly.

In the training sessions, a group of seven participants were trained with low-pass filtered models while another control group of six participants were given non-filtered examples. Before they participated in this experiment, they took a proficiency test and were divided into nearly homogeneous two groups. For the low-pass filtered (LP) group, each session had 10 stimuli and each stimulus was repeated 10 times. For the first five times, the stimulus was filtered and the second five times, non-filtered. For instance, an LP participant saw the text of stimuli on the display, listened to the ambiguous filtered sound and recorded her voice just as she listened. Thus the first five recorded voices of the LP participant sounded quite indistinct, like humming. Then the second five sounded normal because she was listening to a non-filtered voice. For the non-filtered (NF) group, the same stimuli were used as the LP group but they didn't listen to the filtered sounds at all.

Both groups took the same posttest. Their voices were saved in the computer server as WAV files. In this study, 10 stimuli in the pretest and the posttest were analyzed by computer software called WaveSurfer (See p. 34).

Regarding the questionnaire, participants were asked to complete a questionnaire consisting of the following questions after they took the posttest: 1) How do you feel after finishing the training program? 2) What have you noticed about your own pronunciation in English? 3) (LP participants only) How did you feel when you listened to LP sounds?

2.3. Results

2.3.1. Duration of eight sentences

Table 2.3 shows the duration of eight sentences used in the study. This will be the base of the following analyses.

Table 2.3. *The duration of 8 sentences (ms)*

	Duration of 8 sentences							
	Sen. 1	Sen. 2	Sen. 3	Sen. 4	Sen. 5	Sen. 6	Sen. 7	Sen. 8
Model	1685	1158	1068	1101	1118	1608	1761	1683

Participant No.	Low-pass or Non-filtered	Pretest or Posttest	Read or Listen & Repeat	Duration of 8 sentences							
				Sen. 1	Sen. 2	Sen. 3	Sen. 4	Sen. 5	Sen. 6	Sen. 7	Sen. 8
P1	LP	PRE	RD	1519	1239	937	885	1256	2325	2226	1976
	LP	PRE	LR	1668	1369	1026	1198	1181	2014	1968	1830
	LP	POST	RD	1510	1450	1133	1190	1753	1908	1944	1999
	LP	POST	LR	1680	1357	1056	1129	1360	1864	1780	1818
P2	LP	PRE	RD	1587	1320	1983	886	1419	1861	1993	2065
	LP	PRE	LR	1562	1108	1412	1107	1385	1941	1948	2104
	LP	POST	RD	1284	1112	1202	840	1050	1817	1573	1712
	LP	POST	LR	1192	1140	1083	1110	1115	1618	1760	1660
P3	LP	PRE	RD	1954	1430	1172	1230	1488	2058	2043	2204
	LP	PRE	LR	1686	1129	1100	1112	1447	1902	2175	2087
	LP	POST	RD	1545	1326	1083	1146	1402	1718	1710	1905
	LP	POST	LR	1739	1276	1071	1108	1452	1662	1817	1937
P4	LP	PRE	RD	2489	1913	2739	1289	1820	2804	2035	2954
	LP	PRE	LR	1846	1581	1457	1146	1948	2184	2021	2203
	LP	POST	RD	1744	1574	1607	1122	1609	2032	2051	2235
	LP	POST	LR	1717	1502	1298	1100	1512	1958	1866	1854

Participant No.	Low-pass or Non-filtered	Pretest or Posttest	Read or Listen & Repeat	Duration of 8 sentences							
				Sen. 1	Sen. 2	Sen. 3	Sen. 4	Sen. 5	Sen. 6	Sen. 7	Sen. 8
P5	LP	PRE	RD	1713	1388	1584	1078	1596	1864	2300	2220
	LP	PRE	LR	1722	1375	1319	1099	1705	1776	1947	2126
	LP	POST	RD	1588	1360	1610	930	1408	1855	2004	2143
	LP	POST	LR	1535	1128	1261	1196	1307	1655	1782	1846
P6	LP	PRE	RD	1734	1503	1549	1227	1536	2119	1956	2289
	LP	PRE	LR	1665	1480	1302	1142	1439	2091	1979	2018
	LP	POST	RD	1622	1443	1276	1194	1374	1724	1873	2050
	LP	POST	LR	1411	1399	1164	1100	1350	1623	1780	1701
P7	LP	PRE	RD	1297	1206	1112	1105	1236	2117	1791	1697
	LP	PRE	LR	1763	1117	1217	1133	1189	1940	1912	1694
	LP	POST	RD	1488	1367	1009	1290	1438	1970	2112	2095
	LP	POST	LR	1784	1357	1055	1056	1112	1651	1901	1800
P8	NF	PRE	RD	1492	1280	1218	902	1288	1784	1877	2364
	NF	PRE	LR	1716	1195	1236	1077	1215	1901	1910	1923
	NF	POST	RD	1423	1301	1036	1054	1231	1642	1567	1910
	NF	POST	LR	1303	1233	981	1001	1102	1613	1655	1724
P9	NF	PRE	RD	1424	1252	949	949	1135	1752	1528	2082
	NF	PRE	LR	1479	1059	950	1090	1313	1599	1889	1971
	NF	POST	RD	1413	1116	934	963	1162	1675	1803	1852
	NF	POST	LR	1520	1160	1056	1080	1129	1955	1779	1903
P10	NF	PRE	RD	1764	1470	1359	1237	1724	2122	2609	2835
	NF	PRE	LR	1684	1126	1395	1101	1520	2035	2002	2002
	NF	POST	RD	1186	1003	1202	1179	1401	1910	1872	1813
	NF	POST	LR	1364	1185	1076	1174	1345	1679	1798	1770
P11	NF	PRE	RD	1475	1170	1209	865	1477	2375	2782	3126
	NF	PRE	LR	1379	1200	1057	916	1419	1831	1889	1930
	NF	POST	RD	1426	1178	1073	902	1206	1684	1910	1181
	NF	POST	LR	1368	1106	952	887	1293	1851	1774	1730
P12	NF	PRE	RD	1544	1169	1108	1062	1154	1811	1980	2056
	NF	PRE	LR	1681	1230	1154	1024	1307	1929	2056	1783
	NF	POST	RD	1579	1244	1164	1099	1259	1813	1786	1937
	NF	POST	LR	1737	1359	1134	990	1264	2082	2146	1700
P13	NF	PRE	RD	1605	1326	1407	964	1379	2187	2467	2700
	NF	PRE	LR	1570	1162	1110	1152	1130	2138	2232	2142
	NF	POST	RD	1874	1453	1209	1370	1370	2134	2145	2687
	NF	POST	LR	1652	1362	1192	1164	1164	2082	2157	2004

Analysis No. 1

This experiment classifies participants into two groups, and lets them experience the LP program and the NF program. The purpose of this experiment is to see the difference in participants' achievements between the LP and the NF group, and whether there are differences in reading conditions (RD or LR). I have performed "regression ANOVA", in which I regress "Duration variable" on the following binary variables.

I have created the Duration variable by integrating the observations of duration from Sentence 1 to Sentence 8 and the number of observations is 416. See Table 2.4, 2.5, and 2.6.

The first binary variable is the NF. The elements of the NF are 1 for the participants who experienced the NF program and 0 for the participants who did the LP program.

As for the second binary variable, Sentence 2, its elements are 1 for the participants who read the sentence 2 and 0 for the other participants. Sentence 3 and the other binary variables have similar properties to Sentence 2: for example, the elements of Sentence 8 are 1 for those who read sentence 8 and 0 for the others.

Empirical results

The benchmark of these eight binary variables is the participants who read sentence 1 and did the LP program. Their mean duration is the value of the intercept of estimated equation. The value of intercept and

Table 2.4. *Regression analysis (Analysis No. 1)*

R	0.83
R^2	0.69
Corrected R^2	0.69
SE	232.42
n	416

Table 2.5. *Analysis of variance (Analysis No. 1)*

	df	*SS*	Distribution	Observed variance ratio	*F*
Regression	8	49879245.16	6234905.6	115.42	1.03
Residual	407	21986373.6	54020.57	—	—
Total	415	71865618.76	—	—	—

Table 2.6. *Results of each sentence (Analysis No. 1)*

	Parameter	*SE*	*t*-value	*p*-value	Lower bound 95%	Upper bound 95%
Intercept	1623.56	33.91	47.87	2.76	1556.90	1690.23
NF	-71.81	22.86	-3.14	0.002	-116.74	-26.87
sen.2	-296.42	45.58	-6.50	2.31	-386.03	-206.82
sen.3	-360.08	45.58	-7.90	2.64	-449.68	-270.47
sen.4	-506.77	45.58	-11.12	2.96	-596.37	-417.16
sen.5	-227.46	45.58	-4.99	8.96	-317.07	-137.86
sen.6	325.63	45.58	7.14	4.21	236.03	415.24
sen.7	595.88	45.58	13.07	7.49	506.28	685.49
sen.8	435.48	45.58	9.55	1.20	345.88	525.09

standard error are 1623.56 and 33.91, respectively. Therefore, the confidence interval at 95% level for them ranges from 1556 to 1690.

The NF parameter means the overall difference between the duration of the LP and the NF participants. This is −71.8 with t-statistics equal to −3.14 (*p*-value is 0.018). Therefore, I conclude that the NF participants differ from LP participants in terms of duration for all sentences, and this fact is statistically significant at 1.8% level.

Analysis No. 2

The rest of the experiments are the sub-analysis of Analysis No. 1. In this analysis I perform regression ANOVA to see the effect of Pre/Post effect within LP participants (Table 2.7, 2.8, and 2.9). I am interested in the parameter of the binary variable, Post, the element of which is 1 for the post-participants in the LP group, and 0 for pre-participants. The number of observations is 224. The empirical result is that the parameter of POST is −145.15 with t-statistics and *p*-value 6.87 and 0.00, which means the duration of post-participants is shorter than that of pre participants[1]. This result holds for all the readers from sentence 1 to 8.

Analysis No. 3

This analysis is similar to Analysis No. 2 (Table 2.10, 2.11, and 2.12). My purpose is to see the difference between the Pre/Post participants within the NF group. The number of observation is 192. The empirical result is

Table 2.7. *Regression analysis (Analysis No. 2)*

R	0.83
R^2	0.69
Corrected R^2	0.68
SE	222.77
n	224

Table 2.8. *Analysis of variance (Analysis No. 2)*

	df	*SS*	Distribution	Observed variance ratio	*F*
Regression	8	23590507	2948813.38	59.42	2.49
Residual	215	10669900.39	49627.44	—	—
Total	223	34260407.39	—	—	—

Table 2.9. *Results of each sentence (Analysis No. 2)*

	Parameter	*SE*	*t*-value	*p*-value	Lower bound 95%	Upper bound 95%
Intercept	1717.00	44.65	38.45	2.55	1628.99	1805.02
NF	-145.15	29.77	-4.876	2.10	-203.83	-86.48
sen.2	-289.11	59.54	-4.86	2.31	-406.46	-171.75
sen.3	-329.54	59.54	-5.53	9.02	-446.89	-212.18
sen.4	-532	59.54	-8.94	1.86	-649.35	-414.65
sen.5	-219.89	59.54	-3.69	0.0003	-337.25	-102.54
sen.6	285.96	59.54	4.80	2.93	168.61	403.32
sen.7	512.86	59.54	8.61	1.53	395.50	630.21
sen.8	363.5	59.54	6.11	4.73	246.15	480.85

that the parameter of Post is -121.18 with t-statistics and *p*-value 3.80 and 0.00, which also means the duration of post-participants is shorter than that of pre participants[2]. This result holds for all the readers from sentence 1 to 8.

2.3.2. Pitch comparison

The fundamental frequencies of the sound wave are said to be closely related to the pitch of the voice. The pitch of a recorded sound cannot be literally measured but the fundamental frequencies can. Figure 2.2 shows the fundamental frequencies of the sentence "Put all these things in the bag" pronounced by a model speaker and an LP participant. "Pre-lr" is an abbreviation of "pretest-listen-and-read," and "Post-lr" means "posttest-listen-and-read." They are the names of the sound files. Each participant had two files (Pre-lr and Post-lr) for each sentence and they were compared with the equivalent model sound as shown in Figure 2.2. As there were seven participants and each participant had eight sentences in the LP group, a total number of 56 figures (7 x 8) were visually examined. Regarding the NF group, 48 figures (6 x 8) were compared in the same way as the LP group. In comparison with the pitch of the pre/post files, there was a strong tendency of the pitch contours of "Post-lr" to approximate to the model's. This tendency can be observed both in the LP group and the NF group, but more obviously in the LP group (52 out of 56 figures: approximately 92.86%) than the NF group (29 out of 48 figures: approximately 60.42%).

Table 2.10. *Regression analysis (Analysis No. 3)*

R	0.87
R^2	0.76
Corrected R^2	0.75
SE	219.26
n	192

Table 2.11. *Analysis of variance (Analysis No. 3)*

	df	*SS*	Distribution	Observed variance ratio	*F*
Regression	8	28274680.42	3534335.05	73.52	4.23
Residual	183	8797472.29	48073.62	—	—
Total	191	37072152.7	—	—	—

Table 2.12. *Results of each sentence (Analysis No. 3)*

	Parameter	*SE*	*t*-value	*p*-value	Lower bound 95%	Upper bound 95%
Intercept	1587.51	47.47	33.44	7.07	1493.85	1681.17
NF	-120.18	31.65	-3.80	0.0002	-182.62	-57.74
sen.2	-304.96	63.29	-4.82	3.03	-429.84	-180.08
sen.3	-395.71	63.29	-6.25	2.78	-520.59	-270.84
sen.4	-477.33	63.29	-7.54	2.09	-602.21	-352.45
sen.5	-236.29	63.29	-3.73	0.0003	-361.17	-111.41
sen.6	371.92	63.29	5.88	1.95	247.04	496.80
sen.7	692.75	63.29	10.94	9.07	567.87	817.63
sen.8	519.46	63.29	8.21	3.94	394.58	644.34

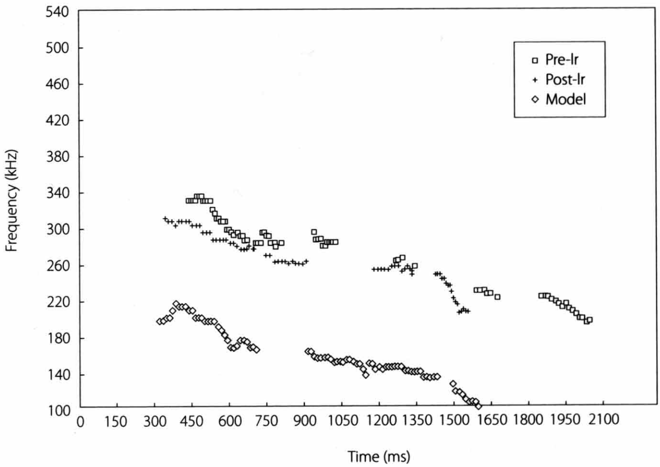


Figure 2.2. *An illustration of the pitch comparison of an LP participant. These are the pitch contours of "Put all these things in the bag."*

2.3.3. Subjective evaluations

The participants' recorded productions were evaluated on a seven-point scale (Figure 2.3) by a total of four teachers of English at a university in Kobe, Japan. Two of them are native speakers of English and the other two are Japanese. Rating sessions were done individually. Raters listened to the participants' four sound files: PRE-RD (read only in pretest), PRE-LR (listen and repeat in pretest), POST-RD (read only in post test), POST-LR (listen and repeat in posttest). Raters were presented with the files of each participant from Sentence 1 to 8 then two minimal pairs, and the order of presenting the four files was random. They were required to judge how natural the utterance sounded as English. If a rater felt an utterance was as natural as English spoken by a native speaker or near-native, seven points would be added to the utterance. As there were 13 participants and each participant's recorded productions were 40, each rater listened to 520 sound files on computer.

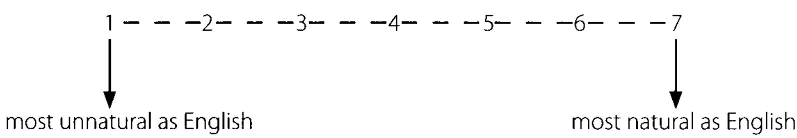


Figure 2.3. *Seven-point scale for subjective evaluations*

Results of subjective evaluations

Statistical tests were conducted by material (Sentence 1 to 8 and Minimal pairs 1 to 2). See Table 8 to 17 in Appendix. Four-factor ANOVA was carried out where "filter" (LP for low-pass-filtered, and NF for non-filtered) was the between-participant variable and "test phase" (PRE for pretest, and POST for posttest), "task" (RD for read and LR for listen and repeat) and "rater" were the within-participant variables.

The results of each material were as follows.

Sentence 1

In Sentence 1, an interaction between phase and task was observed [$F(1, 11) = 5.26, p < .05$]. See Figure 2.8. As seen in Figure 2.8, the interaction is not serious. Regarding main effects, phase [$F(1, 11) = 174.42, p < .05$], task [$F(1, 11) = 39.97, p < .05$], and rater [$F(3, 33) = 3.87, p < .05$] were significant whereas the main effect of filter was not significant [$F(1, 11) = 0.65$, *ns*]. See Figure 2.4, 2.5, 2.6, and 2.7.

Sentence 2

In Sentence 2, no interaction was found between factors. Regarding main effects, phase [$F(1, 11) = 236.22, p < .05$], task [$F(1, 11) = 122.32, p < .05$], and rater [$F(3, 33) = 3.17, p < .05$] were significant whereas the main effect of filter was not significant [$F(1, 11) = 0.09$, *ns*]. See figures 2.9, 2.10, 2.11, and 2.12.

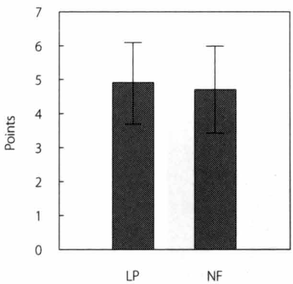


Figure 2.4. *Effect of filter (Sentence 1). Error bars indicate ± one standard deviation.*

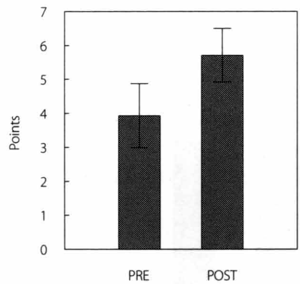


Figure 2.5. *Effect of phase (Sentence 1). Error bars indicate ± one standard deviation.*

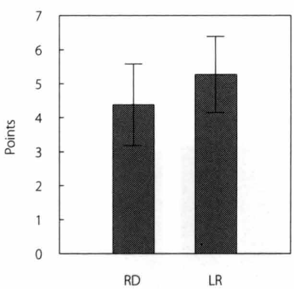


Figure 2.6. *Effect of task (Sentence 1). Error bars indicate ± one standard deviation.*

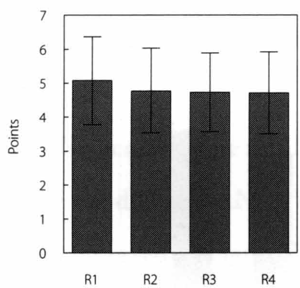


Figure 2.7. *Effect of rater (Sentence 1). Error bars indicate ± one standard deviation.*

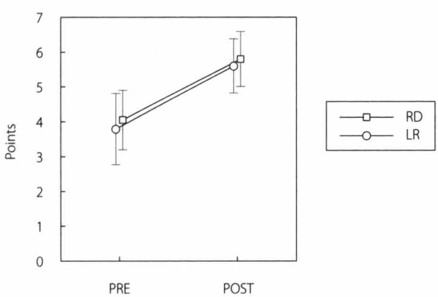


Figure 2.8. *Interaction between phase and task (Sentence 1). Error bars indicate ± one standard deviation.*

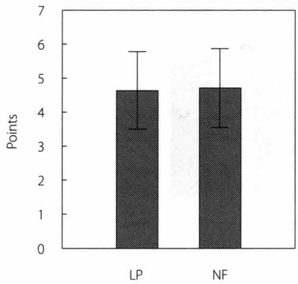


Figure 2.9. *Effect of filter (Sentence 2). Error bars indicate ± one standard deviation.*

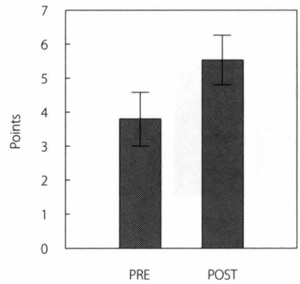


Figure 2.10. *Effect of phase (Sentence 2). Error bars indicate ± one standard deviation.*

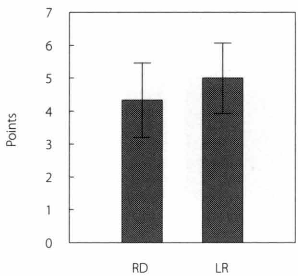


Figure 2.11. *Effect of task (Sentence 2). Error bars indicate ± one standard deviation.*

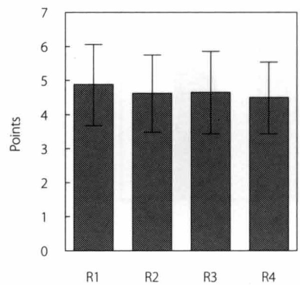


Figure 2.12. *Effect of rater (Sentence 2). Error bars indicate ± one standard deviation.*

Sentence 3

The results of the analysis of variance of Sentence 3 were similar to those of Sentence 1. An interaction between phase and task was observed [$F(1, 11) = 8.67$, $p < .05$]. See Figure 2.17. As seen in Figure 2.17, the interaction is not extensive. Regarding main effects, phase [$F(1, 11) = 309.44$, $p < .05$], task [$F(1, 11) = 108.76$, $p < .05$], and rater [$F(3, 33) = 3.12$, $p < .05$] were significant whereas the main effect of filter was not significant [$F(1, 11) = 0.00$, *ns*]. See Figure 2.13, 2.14, 2.15, and 2.16.

Sentence 4

The results of the analysis of variance of Sentence 4 were similar to those of Sentence 1 and 3, especially some of the digits of ANOVA were the same as Sentence 3 but differences were observed in Figure 2.18 to Figure 2.21. An interaction between phase and task was observed [$F(1, 11) = 8.67$, $p < .05$]. Yet, the interaction is not extensive as shown in Figure 2.22. Regarding main effects, phase [$F(1, 11) = 309.44$, $p < .05$], task [$F(1, 11) = 108.76$, $p < .05$], and rater [$F(3, 33) = 3.12$, $p < .05$] were significant whereas the main effect of filter was not significant [$F(1, 11) = 0.00$, *ns*]. See Figure 2.18, 2.19, 2.20, and 2.21.

Sentence 5

In Sentence 5, an interaction between task and rater was observed [$F(3, 33) = 4.21$, $p < .05$]. However, the interaction is not extensive as in Figure

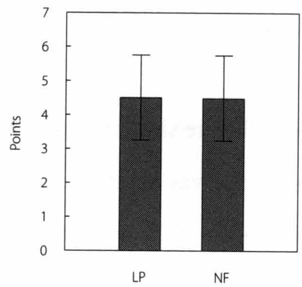


Figure 2.13. *Effect of filter (Sentence 3). Error bars indicate ± one standard deviation.*

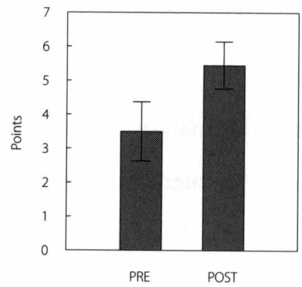


Figure 2.14. *Effect of phase (Sentence 3). Error bars indicate ± one standard deviation.*

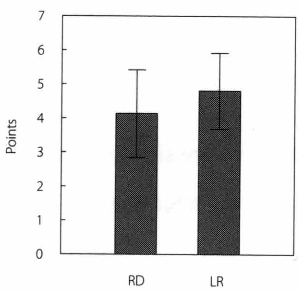


Figure 2.15. *Effect of task (Sentence 3). Error bars indicate ± one standard deviation.*

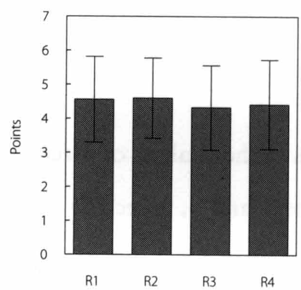


Figure 2.16. *Effect of rater (Sentence 3). Error bars indicate ± one standard deviation.*

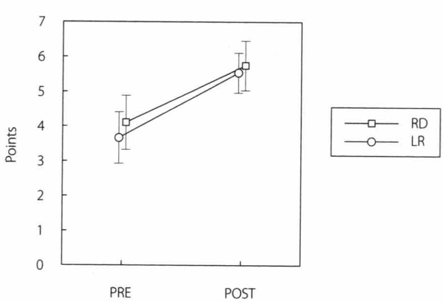


Figure 2.17. *Interaction between phase and task (Sentence 3). Error bars indicate ± one standard deviation.*

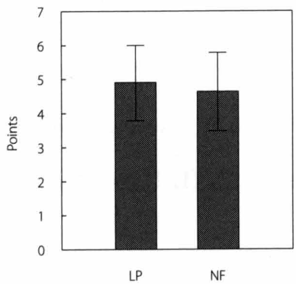


Figure 2.18. *Effect of filter (Sentence 4). Error bars indicate ± one standard deviation.*

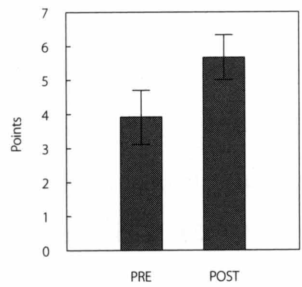


Figure 2.19. *Effect of phase (Sentence 4). Error bars indicate ± one standard deviation.*

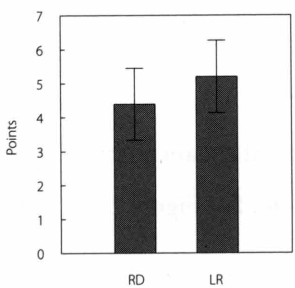


Figure 2.20. *Effect of task (Sentence 4). Error bars indicate ± one standard deviation.*

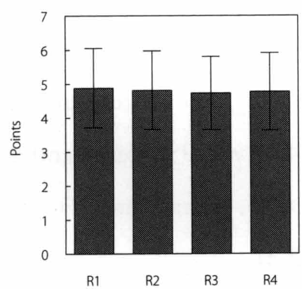


Figure 2.21. *Effect of rater (Sentence 4). Error bars indicate ± one standard deviation.*

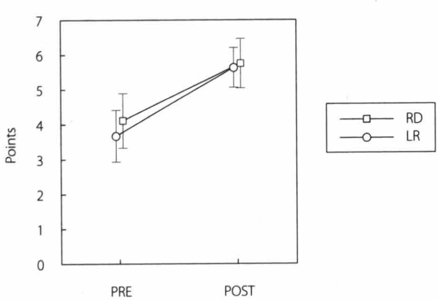


Figure 2.22. *Interaction between phase and task (Sentence 4). Error bars indicate ± one standard deviation.*

2.27. Regarding main effects, phase [$F(1, 11) = 229.10$, $p < .05$], task [$F(1, 11) = 217.21$, $p < .05$], and rater [$F(3, 33) = 2.98$, $p < .05$] were significant whereas the main effect of filter was not significant [$F(1, 11) = 0.65$, *ns*]. See Figure 2.23, 2.24, 2.25, and 2.26.

Sentence 6

The results of analysis of Sentence 6 were slightly different from other sentences. An interaction between filter and task was observed [$F(1, 11) = 12.91$, $p < .05$]. However, the interaction is not extensive as in Figure 2.32. Regarding main effects, phase [$F(1, 11) = 243.99$, $p < .05$], task [$F(1, 11) = 154.47$, $p < .05$], were significant whereas the main effects of filter [$F(1, 11) = 0.35$, *ns*] and rater [$F(3, 33) = 0.41$, *ns*]. See Figure 2.28, 2.29, 2.30, and 2.31.

Sentence 7

The statistical results of Sentence 7 were similar to those of Sentence 5. An interaction between task and rater was observed [$F(3, 33) = 3.25$, $p < .05$]. However, the interaction is not extensive as in Figure 2.37. Regarding main effects, phase [$F(1, 11) = 80.30$, $p < .05$], task [$F(1, 11) = 59.53$, $p < .05$], and rater [$F(3, 33) = 4.31$, $p < .05$] were significant whereas the main effect of filter was not significant [$F(1,11) = 0.30$, *ns*]. See Figure 2.33, 2.34, 2.35, and 2.36.

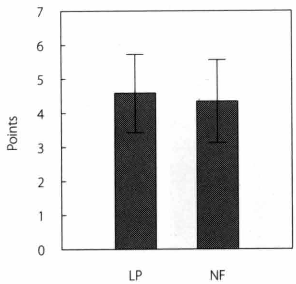


Figure 2.23. *Effect of filter (Sentence 5). Error bars indicate ± one standard deviation.*

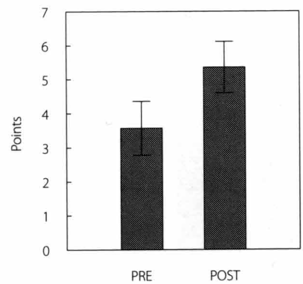


Figure 2.24. *Effect of phase (Sentence 5). Error bars indicate ± one standard deviation.*

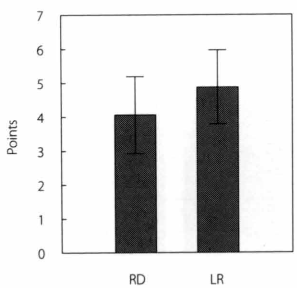


Figure 2.25. *Effect of task (Sentence 5). Error bars indicate ± one standard deviation.*

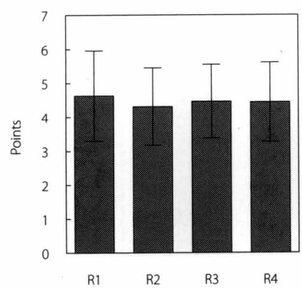


Figure 2.26. *Effect of rater (Sentence 5). Error bars indicate ± one standard deviation.*

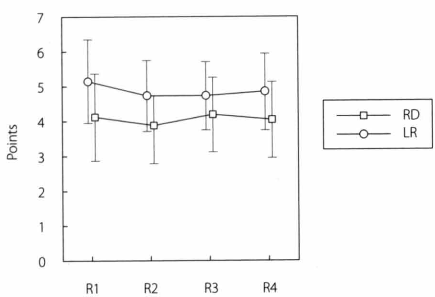


Figure 2.27. *Interaction between task and rater (Sentence 5). Error bars indicate ± one standard deviation.*

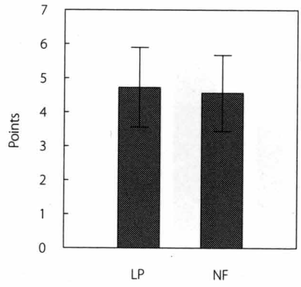


Figure 2.28. *Effect of filter (Sentence 6). Error bars indicate ± one standard deviation.*

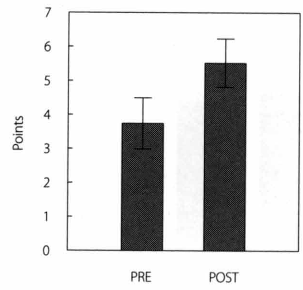


Figure 2.29. *Effect of phase (Sentence 6). Error bars indicate ± one standard deviation.*

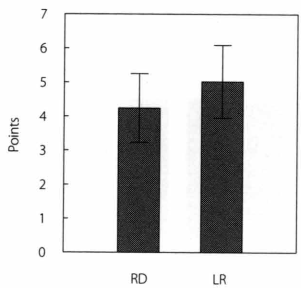


Figure 2.30. *Effect of task (Sentence 6). Error bars indicate ± one standard deviation.*

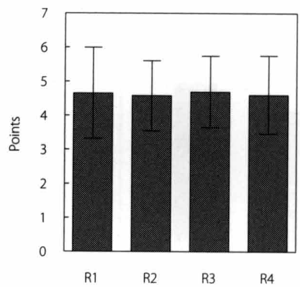


Figure 2.31. *Effect of rater (Sentence 6). Error bars indicate ± one standard deviation.*

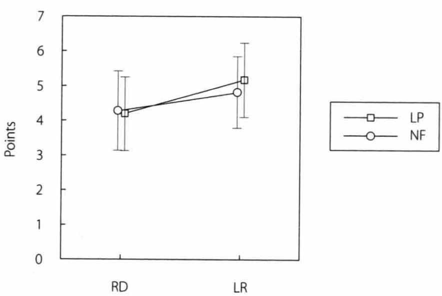


Figure 2.32. *Interaction between filter and task (Sentence 6). Error bars indicate ± one standard deviation.*

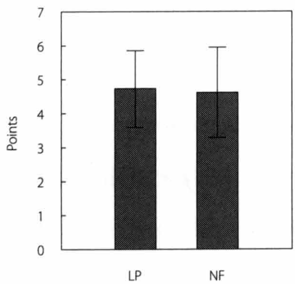


Figure 2.33. *Effect of filter (Sentence 7). Error bars indicate ± one standard deviation.*

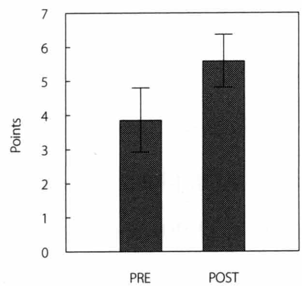


Figure 2.34. *Effect of phase (Sentence 7). Error bars indicate ± one standard deviation.*

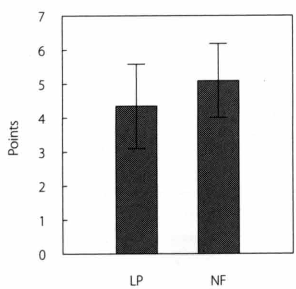


Figure 2.35. *Effect of task (Sentence 7). Error bars indicate ± one standard deviation.*

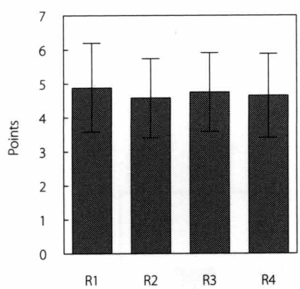


Figure 2.36. *Effect of rater (Sentence 7). Error bars indicate ± one standard deviation.*

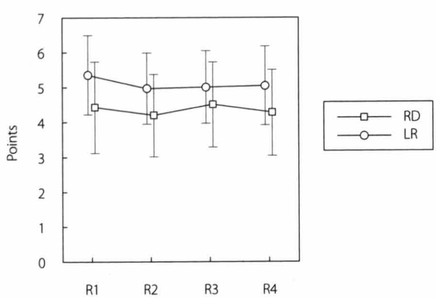


Figure 2.37. *Interaction between task and rater (Sentence 7). Error bars indicate ± one standard deviation.*

Sentence 8

In Sentence 8, four interactions were found between phase and task [$F(1, 11) = 17.24$, $p < .05$], phase and rater [$F(3, 33) = 22.05$, $p < .05$], task and rater [$F(3, 33) = 21.90$, $p < .05$], and phase, task, and rater [$F(3, 33) = 20.90$, $p < .05$]. However, the interactions were not extensive as in Figure 2.42 to Figure 2.44. Regarding main effects, phase [$F(1, 11) = 172{,}04$, $p < .05$] and task [$F(1, 11) = 28.81$, $p < .05$], were significant whereas the main effects of filter [$F(1, 11) = 0.11$, *ns*] and rater [$F(3, 33) = 0.95$, *ns*] were not significant. See Figure 2.38, 2.39, 2.40, and 2.41.

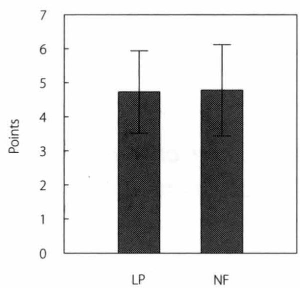


Figure 2.38. *Effect of filter (Sentence 8). Error bars indicate ± one standard deviation.*

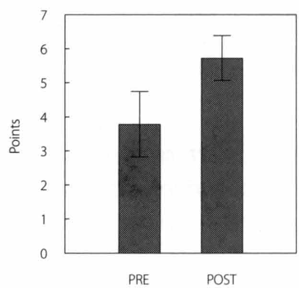


Figure 2.39. *Effect of phase (Sentence 8). Error bars indicate ± one standard deviation.*

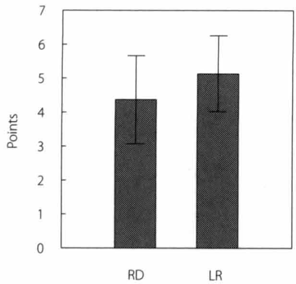


Figure 2.40. *Effect of task (Sentence 8). Error bars indicate ± one standard deviation.*

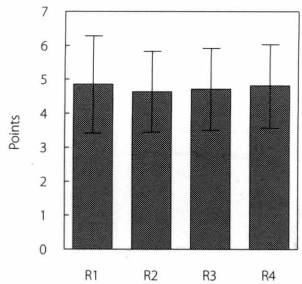


Figure 2.41. *Effect of rater (Sentence 8). Error bars indicate ± one standard deviation.*

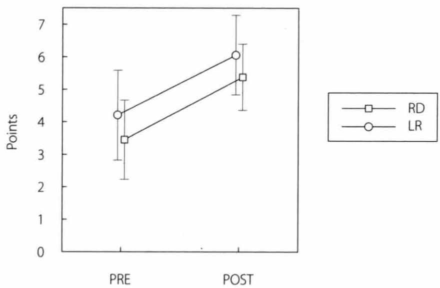


Figure 2.42. *Interaction between phase and task (Sentence 8). Error bars indicate ± one standard deviation.*

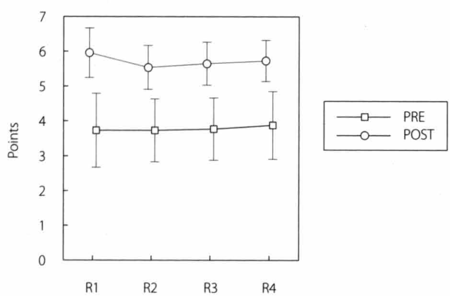


Figure 2.43. *Interaction between phase and rater (Sentence 8). Error bars indicate ± one standard deviation.*

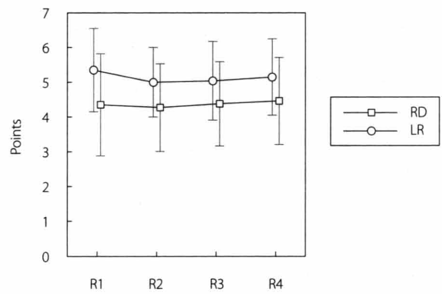


Figure 2.44. *Interaction between task and rater (Sentence 8). Error bars indicate ± one standard deviation.*

Minimal Pair 1 and Minimal Pair 2

Minimal Pair 1 and Minimal Pair 2 showed the same tendency in the results of ANOVA. There were two interactions between phase and rater [Minimal Pair 1: $F(3, 33) = 3.13, p < .05$, Minimal Pair 2: $F(3, 33) = 3.13, p < .05$], and task and rater [Minimal Pair 1: $F(3, 33) = 4.23, p < .05$, Minimal Pair 2: $F(3, 33) = 4.23, p < .05$]. However, the interactions were not extensive as in Figure2.53 to 2.56. Regarding main effects, phase [Minimal Pair 1: $F(1, 11) = 32.86, p < .05$, Minimal Pair 2: $F(1, 11) = 32.86, p < .05$] and task [Minimal Pair 1: $F(1, 11) = 12.07, p < .05$, Minimal Pair 2: $F(1, 11) = 12.07, p < .05$], were significant whereas the main effects of filter [$F(1, 11) = 0.05$, *ns*] and rater [$F(3, 33) = 1.68$, *ns*] were not significant. See Figure 2.45, 2.46, 2.47, 2.48, 2.49, 2.50, 2.51, and 2.52.

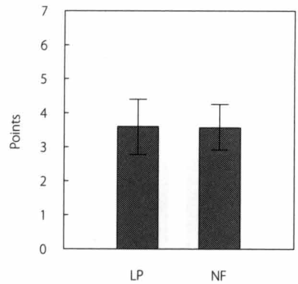


Figure 2.45. *Effect of filter (Minimal Pair 1). Error bars indicate ± one standard deviation.*

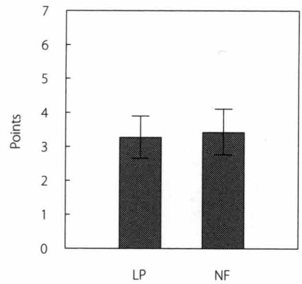


Figure 2.46. *Effect of filter (Minimal Pair 2). Error bars indicate ± one standard deviation.*

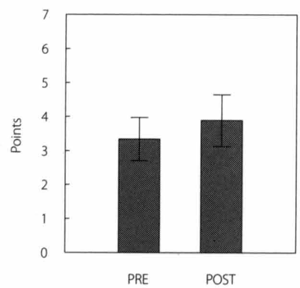


Figure 2.47. *Effect of phase (Minimal Pair 1). Error bars indicate ± one standard deviation.*

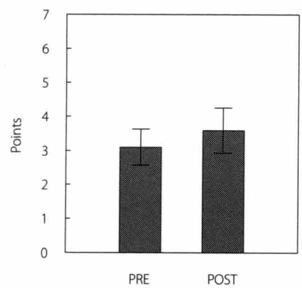


Figure 2.48. *Effect of phase (Minimal Pair 2). Error bars indicate ± one standard deviation.*

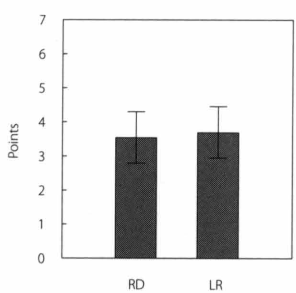


Figure 2.49. *Effect of task (Minimal Pair 1). Error bars indicate ± one standard deviation.*

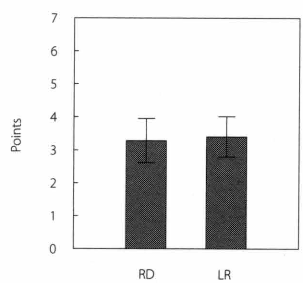


Figure 2.50. *Effect of task (Minimal Pair 2). Error bars indicate ± one standard deviation.*

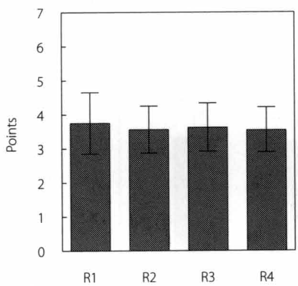


Figure 2.51. *Effect of rator (Minimal Pair 1). Error bars indicate ± one standard deviation.*

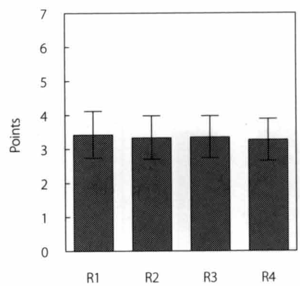


Figure 2.52. *Effect of rator (Minimal Pair 2). Error bars indicate ± one standard deviation.*

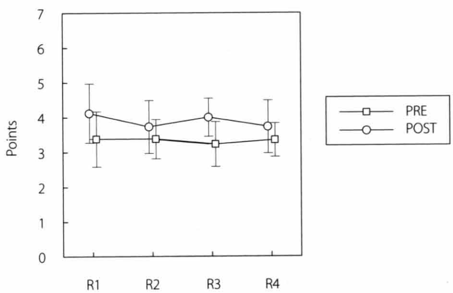


Figure 2.53. *Interaction between phase and rater (Minimal Pair 1). Error bars indicate ± one standard deviation.*

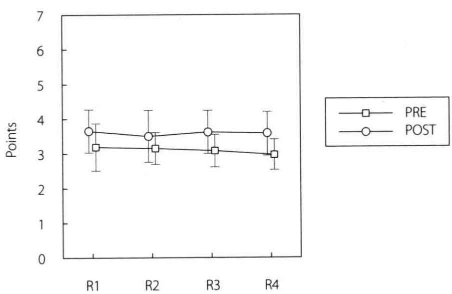


Figure 2.54. *Interaction between phase and rater (Minimal Pair 2). Error bars indicate ± one standard deviation.*

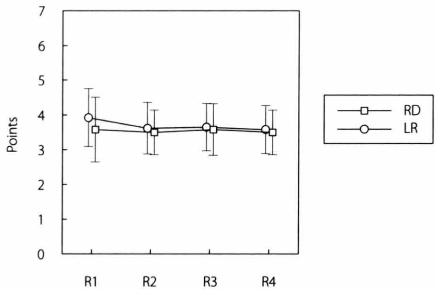


Figure 2.55. *Interaction between task and rater (Minimal Pair 1). Error bars indicate ± one standard deviation.*

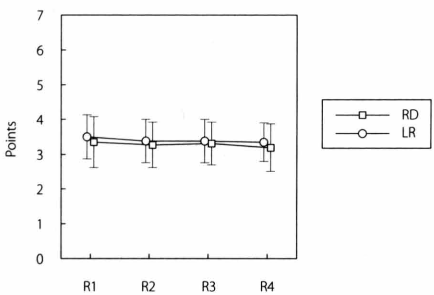


Figure 2.56. *Interaction between task and rater (Minimal Pair 2). Error bars indicate ± one standard deviation.*

2.3.4. Questionnaire responses: participants' feedback

At the conclusion of the training program, all participants were given a questionnaire to complete and return to me anonymously. The purpose of the questionnaire was to assess participants' perceived value of speech technology such as this in foreign language instruction. The responses shown below are listed according to frequency of occurrence on the returned questionnaires. The low-pass filtered training was positively accepted by the participants.

1 How do you feel after finishing the training program?
   - I feel more confident about my pronunciation.
   - I was dismayed to see the difference between my pitch contours and the model's but gradually I came to approximate the model.
   - I am not sure whether my pronunciation of English has improved but I feel as though I have achieved something.

2 What have you noticed about your own English pronunciation?
   - My intonation was rather monotonous.
   - I have noticed how my pronunciation in English differs from that of native speakers.

3 (LP participants only) How did you feel when you were listening to the LP sounds?
   - I found it tiring at first. The LP sounds felt uncomfortable because I couldn't understand what was being said. This was frustrating.

However, as I attended the training sessions, I became more used to them. Now I am not uncomfortable with the sounds at all.

- They sounded as if someone was talking under water.
- I recognized the rhythm and intonation of the speech although I didn't understand what was said.
- After repeatedly listening to the LP sounds followed by the NF sounds, I felt the NF sounds became clear and easy to understand.

2.4. Discussion and concluding remarks

The results of the acoustic analyses of this experiment revealed significant differences between the LP group and the NF group as in 2.3.1. Throughout the low-pass applied training sessions, the LP participants appear to have become more sensitive to prosodic features than the NF participants. However, as for subjective evaluations as in 2.3.3, there were no significant differences between the LP group and NF group. All raters recognized that the production of each participant in posttest sounded more natural as English than in pretest. While more data from both of the groups are needed to fully validate the robustness of the hypothesis, it would be reasonable to conclude that computer-assisted pronunciation training using low-pass filters is more effective to train accurate pronunciation than training without digital filters on the basis of quantitative analyses. Yet, on the level of human perception, the differences are subtle and hard to judge.

The following are my research questions and their answers in this study.

1 *Is there any difference in speech production between the group of participants who attended 10 sessions of training with electric low-pass filtered sounds and the controlled group who did the same training without filtered sounds?*

Yes, there is. As for the duration of the eight sentences in the posttest, the LP group was significantly different from the NP group and they approximated more to the model sounds. By comparing pitch contours of the pre/posttests, the LP group showed closer approximation of the model sounds.

Yet, regarding subjective evaluations, all raters found that both groups of participants improved their English pronunciation.

Did the low-pass group become more accurate in production than the non-filtered group, or vice versa? Were both improved?

It depends on the definition of accuracy, but as for the approximation to the model sounds, the LP group became more accurate in the production of prosodic features than the NP group although both groups improved their production skills.

2 *How did the participants feel after finishing the 10 training sessions?*
See 2.3.4, questionnaire responses.

There are some remaining issues. As the number of participants was limited in this study, as mentioned earlier, a follow-up experiment is needed with a larger pool of participants. As for the difference between the results of acoustic analyses and those of subjective evaluations, more analyses are needed. The training period might not have been long enough. One of the possible reasons of the difference is that the human raters' sense of "Englishness" may differ from that of machines. The precision of computer speech analysis is becoming ever more accurate. In this experiment, as raters were required to focus on how naturally the utterance sounded as English, they didn't concentrate on the duration of sentences.

In conclusion, the English prosody training using low-pass filters contributed to acquiring accuracy of pronunciation. Yet, training without low-pass filters was also effective in improving English prosody according to the subjective evaluations. The results suggest that we could apply low-pass filtering to English prosody training in the interests of accuracy. The efficient collection of more data remains as a challenge for future work.

Notes

1 Without binary variables for sentences (i.e., binary variables Sentence 1 through to Sentence 8), I recognize this result is robust. The parameter is -145.15 with t-statistics 2.81, p-value 0.005, and $n = 416$.

2 Without binary variables for sentences (i.e., binary variables Sentence 1 through to Sentence 8), this result is almost robust. The parameter is -120.18 with t-statistics 1.90 and p-value 0.059.

Chapter 3 (study 3)

The Apparent Priority of Prosodic Features over Individual Sounds in Second Language Speech Learning

3.1. Introduction

3.1.1. Background and outline of the study

PHONETICS, phonology and other aspects of language have long been viewed as being best learnt through a bottom-up approach. Traditional works on phonetic pedagogy, such as Gimson (1981), began with the description of vowels, consonants, words and connected speech. It was natural for more practical textbooks on pronunciation to deal with individual speech sounds first, then morphemes, vocabulary, phrases, and discourse units. According to this approach, learners would reach a particular level of proficiency by accumulating the mastered entities of the target language. Over the years, however, there has been a shift towards a more holistic top-down approach in the field of English language teaching. The current emphasis in pronunciation teaching seems to reside in the prosodic features, or the suprasegmentals of language such as stress, rhythm, intonation, and pitch as opposed to the segmentals such as individual vowel and consonant sounds. Brown (1991) expressed the view that "the suprasegmentals are more basic and contribute more to intelligibility and accent. They should therefore appear first in textbooks and be mastered first by learners." However, investigations to support this claim were not carried out. Moreover, in Japan, there seems to be relatively little interest in pronunciation teaching and the bottom-up approach is still common in the English language classroom. Indeed, concerning both approaches, theory has not been sufficiently supported by empirical research to establish which approach is more effective in language acquisition.

This study aims to examine which approach is more effective in pronouncing English naturally. There were four groups of participants; Group A: learning consonants and vowels first, then prosody in phrases or sentences, Group B: learning prosody in phrases or sentences first then consonants and vowels, Group C: learning prosody and individual sounds together, and Group D: a control group which didn't practice but took the pretest, the midtest, and the posttest. The participants in this study were all Japanese and joined the experiment voluntarily. Most of them were enrolled in English listening courses at a university in Japan. Sound data of the pre/mid/post-tests were collected using original software. Sentence or phrase duration and F0 ranges were measured using Praat software. Data were also judged by four raters as to whether they sounded natural as English. The results were analyzed in Analysis of Variance (ANOVA). Findings show that Group B achieved the highest results in both objective evaluation (the measurement of duration and pitch ranges) and subjective evaluation (raters' judgments). Therefore, the findings seem to indicate that repeating sentences or phrases at the beginning of a series of sessions played an important role in acquiring the prosody of a target language. The research in this area will shed much light on our understanding of the process of speech perception in general.

3.1.2. Purpose of the study

This study investigates the prosodic aspects of second language acquisition. My principal concern is order effects: Which group will acquire the prosodic features of English most effectively?

- Beginning training sessions with individual sounds followed by prosody
- Beginning training sessions with prosody followed by individual

sounds

- Training individual sounds and prosody together

Sentence duration and F0 ranges of 17 stimuli were measured in pre/mid/posttest. The durational ratios and F0 range ratios of those stimuli were also figured and analyzed in ANOVA. Subjective evaluations were also conducted. The raters of the experiment were the same professional English teachers in Study 1 and 2. The only criterion for rating was "How natural does the utterance sound as English?" Participants' recorded productions were evaluated on a seven-point scale. ANOVA was used for analysis of the results and compared with the consequences of the objective evaluation stated above.

3.2. Experiment

3.2.1. Materials

The same stimuli were used in the pretest, the midtest, and the posttest. The total number of stimuli in each test was 17. The stimuli were largely collected and selected by Professor Shinobu Mizuguchi at Kobe University. They were recorded by two native speakers of British English at a studio in Osaka. The recorded productions were inserted in the original software as model sounds. The test software was created by ATR. Table

3.1 shows the contents of the stimuli.

Selection of the stimuli for testing followed these guidelines: 1) familiar vocabulary, 2) structural variety, 3) sustained phonation which may provide a visually obvious display of pitch contour, 4) a variety of relatively short sentences or phrases and longer sentences.

Table 3.1. *The stimuli in the pre/mid/posttests*

No.	Sentence
1	The BBC.
2	You'll have to take the tube.
3	Pardon?
4	Roads are rough in rural areas.
5	What a good idea!
6	What did Mary bring?
7	The wine.
8	Isn't she pretty!
9	Pork or beef?
10	Sorry, I don't eat meat.
11	I like chocolate, but I'm on a diet.
12	Milk, I believe, comes from cows.
13	Would you pass me the water?
14	Hey, are you going to return those books of mine you borrowed?
15	Which books? I can't remember borrowing any.
16	The ones about biology and language.
17	Oh, those books. Er ... could I keep them a few more days?

3.2.2. Speakers

In the pre/mid/posttests, a male speaker and a female speaker of standard British English recorded the test items as models. Between the pretest and midtest, participants attended 10 sessions of pronunciation training, and following the midtest, they participated in the 10 additional sessions. Two pieces of software were used in the 20 sessions. In that used for training prosodic features, the same speakers' voices in the pre/mid/post tests were inserted. In that used for individual sounds, a male speaker of standard American English and a male speaker of standard British English recorded the models.

3.2.3. Participants

The total number of participants is 80. At first, approximately 120 native speakers of Japanese volunteered to participate in this study. All of them were undergraduate students at Konan University in Kobe. They belonged to different faculties of the university. None had spent more than two months in an English speaking country. Their ages ranged from 19 to 23. They reported normal hearing and vision. All of them except the students in the control group were taking English listening courses. As for the control group, they were taking an English translation course.

The participants took a short version of the TOEIC test to assess their English proficiency. The participants for the experiment were divided into three nearly homogeneous groups. Regarding the participants in the control group, they took the TOEIC to begin with, and then the pretest.

As the pre/mid/posttests and the 20 sessions were individually performed using 10 computers in a self-study room at the university, some of the data of the three tests were missing or some of them were not recorded clearly enough to analyze. Furthermore, the period of the training sessions was about two month long and the participants were supposed to attend 20 sessions. Consequently, the number of complete data was reduced. Barely 20 participants' data of each group can be analyzed as complete sets of three tests. Regarding the number of male/female participants, see Table 3.2. The number of female students who volunteered to join this experiment was originally higher than that of the male students. Accordingly, the number of female students who completed the three tests was higher.

3.2.4. Procedure

A pretest-midtest-posttest design was used to measure the effects of two months training (20 sessions of about 40 to 60 minutes each) using computerized visual displays of pitch contours, wave forms and power as feedback (See Figure 3.1). As for the pre/mid/posttests, the same software that was provided by ATR (Advanced Telecommunications Research Institute International) was used. Users can customize it by inputting the stimuli. 17 stimuli for each test were set in the software. See Table 3.1. In

Table 3.2. *The number of the participants*

Group	Male	Female	Total
A	6	14	20
B	6	14	20
C	7	13	20
D	8	12	20
Total	27	53	80

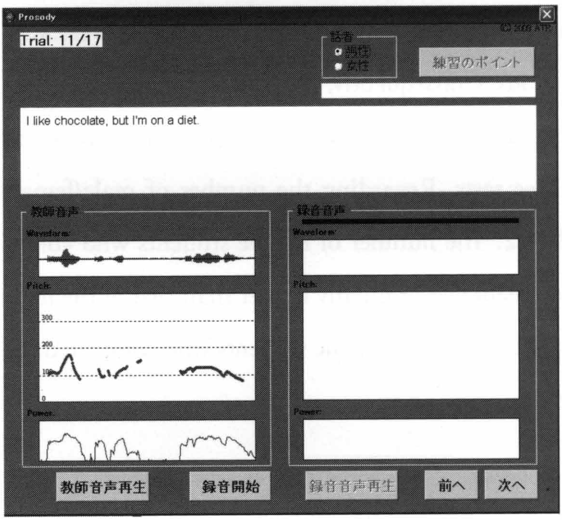


Figure 3.1. *The interface of pre/mid/posttest. The same software was used in this study as with Study 1 and Study 2. The contents were customized by ATR.*

the pre/mid/posttest, the participants recorded their voices which were saved in the computer server as WAV files. They were analyzed by computer software called Praat.

Participants were asked to come to a self-study room at any time during the training period, where the training sessions were performed on 10 computers. For the first time, they were asked to read instructions about using the software and took the pretest by computer. Their voices were automatically recorded and stored in the server. During the training period, they took the training sessions at any time they liked. Some of the participants came daily and finished the training sessions relatively early. Some of them came to the room as regularly as twice a week and others came quite irregularly.

The training session groups

Figure 3.2 shows the three tests and the two-part training sessions. The following is the description of each group.

Group A (Individual-sound-first group)

After taking the pretest, the participants of Group A trained individually with software to practice English pronunciation that I created in 2004 and made available on the web. Details of the software are mentioned in the next section. The participants took 10 training sessions focusing on practicing particular individual sounds such as /r/ and /l/ in one ses-

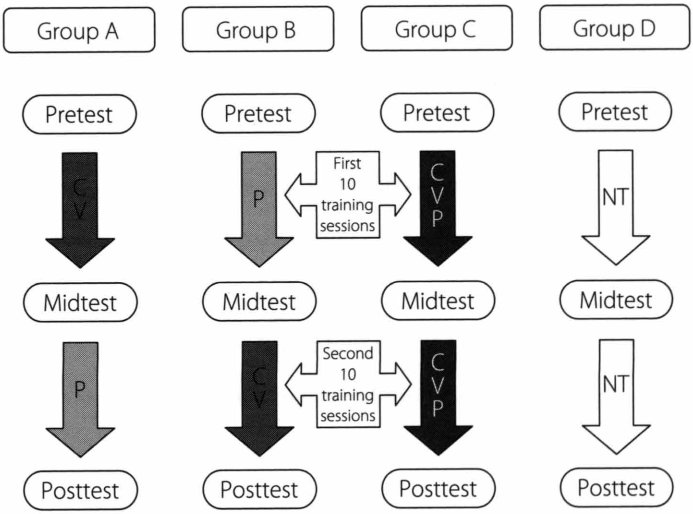


Figure 3.2. *The three tests and training sessions. In the arrows, "CV" means Consonants and Vowels, i.e., individual sounds, "P" means "Prosody," "CVP" means "Consonants, Vowels, and Prosody," and "NT"means" No training."*

sion. After that, they took the midtest then participated in the second 10 training sessions. This time they used different software the details of which are described below. The second 10 training sessions focused on acquiring prosodic features. After they finished the second 10 sessions, they took the posttest.

Group B (Prosody-first group)

The participants of Group B trained in the opposite way to Group A. In the first 10 sessions, they practiced prosodic features and in the second 10 sessions, they practiced individual sounds.

Group C (Mixed training group)

In the first and second 10 training sessions, the participants of Group C trained with both pieces of software together. They practiced prosodic features and individual sounds in one session.

Group D (Control group)

The participants of Group D didn't join the training sessions at all. They took the pretest first, and after three weeks they took the midtest, and finished the posttest three weeks later.

The two pieces of training software

A) "English Pronunciation Practice for Japanese Learners" for practicing individual sounds

As mentioned in the above section, two different pieces of software were used in this experiment. As for training individual sounds, software called "English Pronunciation Practice for Japanese Learners" was used. I created the software with financial assistance from Konan University and it is now available on the web. See Figure 3.3.

Prosodic features can also be practiced with this software, but on this occasion, the software was used to train individual sounds. In the 10 training sessions with this software, the participant practiced items in Table 3.3.

In a single training session, participants read the explanation on the computer display about how to produce a given consonant or vowel, and then performed five to ten exercises. The participant was asked to repeat each exercise at least five times. As the software is not equipped with a recording function, the production of the participants at any session was not recorded

B) "Prosody" for practicing prosodic features

As for the training sessions in prosodic features, different software named "Prosody" supported by Grants-in-Aid for Scientific Research was used. The format is the same as pre/mid/posttest but the inputted contents

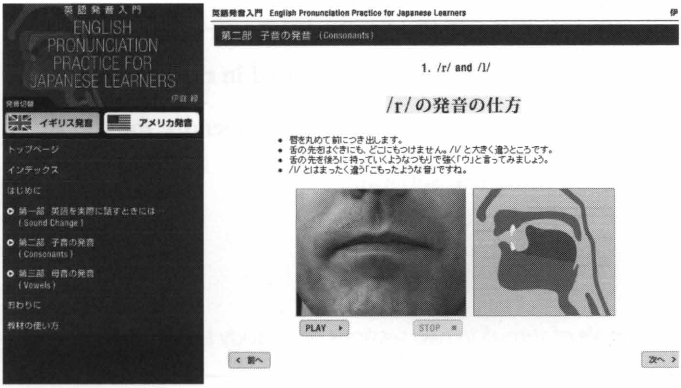


Figure 3.3. *The interface of software, "English Pronunciation Practice for Japanese Learners" (http://kccn.konn-u.ac.jp/ilc/english/)*

Table 3.3. *The contents of 10 training sessions for individual sounds*

No.	Contents
Session 1	/r/ and /l/
Session 2	/f/ and /h/
Session 3	/b/ and /v/
Session 4	/T/ and /s/
Session 5	/Δ/ and /z/
Session 6	/s/ and /Σ/
Session 7	/n/ and /N/
Session 8	The difference between /i/, /I/ and "い"
Session 9	The difference between /e/ and "え"
	The difference between /æ/, /Θ/and "あ"
Session 10	The difference between /ɒ/, /ɑ/ and "お"
	The difference between /u/, /ʊ/ and "う"

were different. The software can be customized according to the needs of users. Table 3.4 shows an example of stimuli used in the first and second sessions of prosody training. See Table 18 in Appendix for the entire stimuli of all sessions using this software.

Table 3.4. *Example of stimuli in two sessions of prosody training*

No.	Sentence	
Session 1 Tones		
1	Great!	
2	Thanks!	
3	Pardon?	
4	Yes.	
5	Bass.	
6	A: I'll be there by five.	B: Great!
7	A: Care for a drink?	B: Thanks!
8	A: You'll have to take the tube.	B: Pardon?
9	A: You were there, weren't you?	B: Yes.
10	A: He sings tenor.	B: Bass.
Session 2 Statements		
1	This is a pen.	
2	I think it's great.	
3	A: When'll they finish?	B: Next Wednesday.
4	I won't eat anything.	
5	I won't eat anything.	
6	Will you eat anything?	
7	Roads are rough in rural areas.	
8	It's not hot, it's cold.	
9	A: Who's that?	B: I know her face.
10	A: She's working in Oxford.	B: Cambridge.

As mentioned before, the interface of this software is the same as the pre/mid/posttest. Participants were asked to repeat the model utterance for which the text was shown on the screen at least 10 times. Along with the text, the participant saw the waveform, pitch and power of the model sound. Then he/she pushed the recording button and read the text aloud. The waveform, pitch and power of the participant were also shown on the same screen. See Figure 3.4. The production during the sessions with this software was saved on the computer automatically but not used for the analysis of this experiment. Only the production of the pre/mid/posttest was analyzed later.

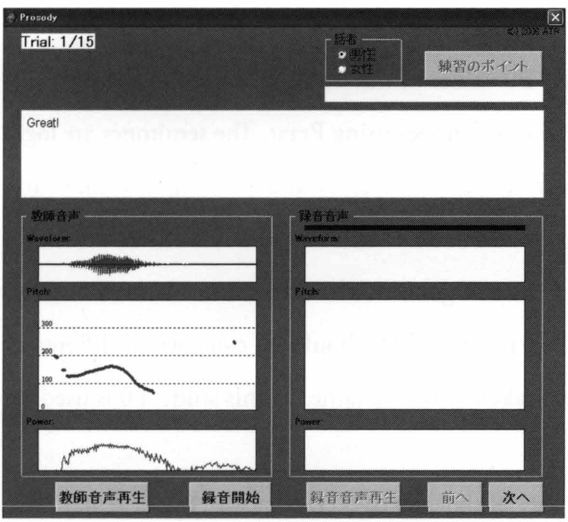


Figure 3.4. *The interface of software "Prosody"*

Acoustic analysis

Sentence duration and F0 ranges were measured using waveform displays and wideband spectrograms of Praat. See Figure 3.5. There were 8,160 data (4,080 for sentence duration, 4,080 for F0 ranges) in total.

Sentence duration

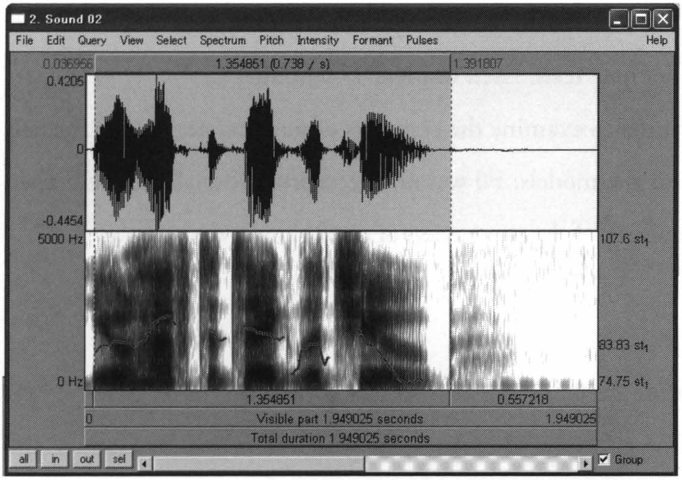
There were four groups in all. Each group had 20 participants and took the three tests. Each test had 17 stimuli.

In order to examine the contrasts in duration among the participants and the models, duration was analyzed proportionally as well.

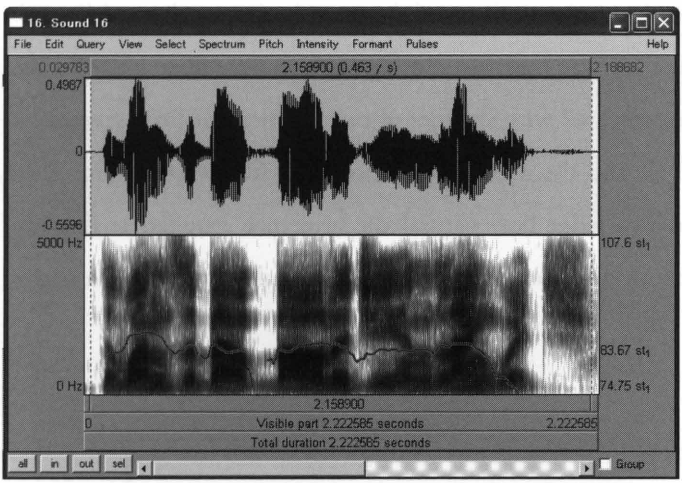
F0 ranges

The highest and lowest F0 values in the whole utterance in the pre/mid/posttest were measured in semitones using Praat. The semitones are logarithmic scales of Hertz. Usually F0 values and F0 ranges of male and female speakers are quite different. However, there is no more difference between the ranges of the two genders when they are converted to semitones. Strictly speaking, pitch and F0 should be categorized differently though they are widely taken to be the same. In this study, F0 is used for considering pitch.

The total number of the data of F0 ranges was 4,080. It was measured using Praat displays in the same way as sentence duration. See



(a) A model pronounced Stimulus 2 in the pre/mid/posttest: "You'll have to take the tube."



(b) A model pronounced Stimulus 16 in the pre/mid/posttest: " The ones about biology and language."

Figure 3.5. *Praat interface*

Figure 3.5. There were four groups in all. Each group had 20 participants and took the three tests. Each test had 17 stimuli.

In order to examine the contrasts of pitch ranges among the participants and the models, F0 was analyzed proportionally as well. Then the mean average of the proportion of pitch ranges was calculated.

Subjective evaluations

The procedure of subjective evaluations is the same as in Study 1 and Study 2. The raters are also the same group of people in Study 1 and Study 2. The participants' recorded productions were evaluated on a seven-point scale (Figure 3.6) by a total of four teachers of English at a university in Kobe, Japan. Two of them are native speakers of English and the other two are Japanese. Rating sessions were done individually. Raters were presented with the files of each participant from sentence 1 to sentence 17, and the order of presenting the files was random. They were required to judge how natural the utterance sounded as English. If a rater felt an utterance was as natural as English spoken by a native

1 – – –2– – –3– – –4– – –5– – –6– – –7

most unnatural as English ↑ (1) most natural as English ↑ (7)

Figure 3.6. *Seven point scale for subjective evaluations*

speaker or near-native, seven points would be added to the utterance. As there were 80 participants and each participant's recorded productions were 51, each rater listened to 4,080 sound files on computer. The total number of the rating results of the four raters is 16,320.

3.3. Results

3.3.1. Durational ratios of the 17 stimuli of the three tests

ANOVA was conducted on the results of the mean average of durational ratios to test differences between means for significance. See Table 19 in Appendix.

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed a significant main effect of phase, [$F(2, 76) = 20.35, p < .001$], a significant group × phase interaction [$F(6, 152) = 2.23, p < .05$].

The interaction between group and phase was further explored. The significant simple main effect was observed only for the factor phase for Group B, [$F(2, 152) = 19.72, p < .001$]. See Figure 3.7. Observing the line graph, the lines other than Group D appear to lean enough to suggest that the difference among pre/mid/post is significant. However, regarding main effects, only the result of Group B was significant. The model sounds were treated as 1 in ratio. *Post hoc* pairwise comparisons

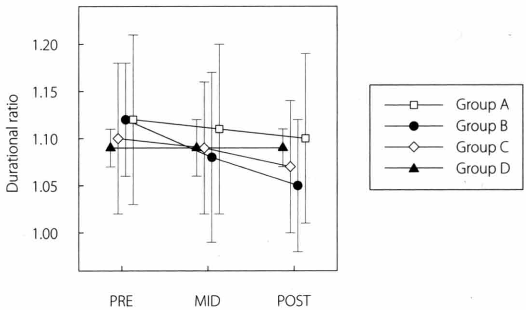


Figure 3.7. *The mean averages of durational ratios of each group. Error bars indicate ± one standard deviation.*

using Ryan's method, where .05 as a significance level, showed that there were significant differences between pretest vs. midtest, midtest vs. posttest, and pretest vs. posttest for Group B.

The results show that only durational ratios of Group B significantly approximate the model ratio from the pretest to the posttest.

3.3.2. F0 range ratios of the 17 stimuli

ANOVA was conducted on the results of the mean average of F0 range ratios to test differences between means for significance (Table 20 in Appendix).

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed a significant main effect of

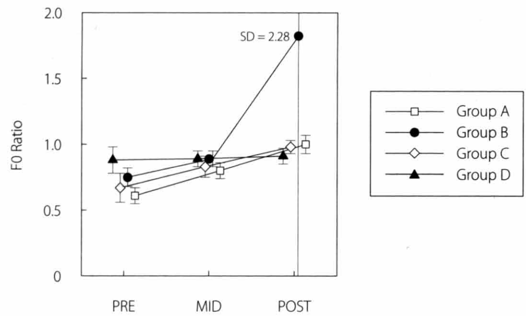


Figure 3.8. *The mean averages of F0 ratios of each group. Error bars indicate ± one standard deviation.*

group [$F(3, 76) = 3.30, p < .05$] and phase [$F(2, 76) = 9.44, p < .001$]. A significant group × phase interaction [$F(6, 152) = 2.23, p < .05$] was also found.

The interaction between group and phase was further explored. The significant simple main effect was observed for the factor phase for Group B, [$F(2, 152) = 14.82, p < .001$]. There was also a significant difference between groups in posttest, [$F(3, 228) = 8.02, p < .001$]. See Figure 3.8. Observing the above line graph, the lines other than Group B do not appear to rise enough to suggest that the difference among pre/mid/post is significant. The model sounds were treated as 1 in ratio. *Post hoc* pairwise comparisons using Ryan's method, where .05 as a significance level, showed that there were significant differences between pretest vs.

posttest and midtest vs. posttest for Group B.

As seen in Figure 3.8, the results show that F0 range ratios of Group B significantly changed throughout the tests especially from the midtest to the posttest. Group D, the control group, did not show any difference throughout the three tests. As for the results of the posttest of the other groups, Group A and Group C approximate to the model ratio.

3.3.3. Subjective evaluations

Regarding the results of all the ratings. Table 21 in Appendix shows mean subject evaluations of the four raters.

A two-factor ANOVA with group (A, B, C, and D) and phase (pre, mid, and posttest) as factors showed significant main effects of phase [$F(2, 76) = 1008.67$, $p < .001$] and group [$F(3, 76) = 62.73$, $p < .001$]. The interaction between phase and group was also significant [$F(6, 152) = 113.87$, $p < .001$].

The interaction between group and phase was further explored. The significant simple main effects were observed in most cases except for the effect of group for pretest (tendency, $p < .10$), and the effect of phase for Group D (*ns*). *Post hoc* pairwise comparisons using Ryan's method, with .05 as a significance level, were conducted. There were significant differences between groups in the midtest and posttest phases except for the Group A vs. Group C in the midtest. In contrast, there were no significant differences between groups in the pretest phase. There were differences between pretest vs. midtest, midtest vs. posttest and pretest vs. posttest in all groups except Group D, which was a control untrained group. Group B achieved the highest evaluation in the posttest.

Observing the inclination of the lines of each group in Figure 3.9, we note that Group B obtained the highest evaluation among the raters in posttest. Group A and Group C also showed steady growth in evaluation. As their lines are almost overlapped, it can be said that they had the same growth tendency.

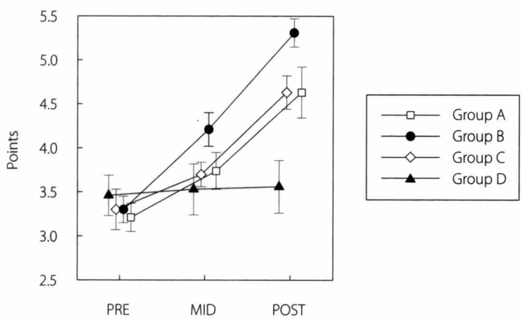


Figure 3.9. *Mean subject evaluations by four raters. Error bars indicate ± one standard deviation.*

3.4. Discussion and concluding remarks

The main goal of this study was to investigate prosodic aspects in second language speech perception and production, especially concerning the order effects of training prosodic features and individual sounds.

This study provides some encouraging new data regarding the acquisition of non-native prosodic features and individual sounds. The findings are as follows.

Effects of training

Regarding the effects of the training, particularly in the subjective evaluations, all the experimental groups showed significant differences from pretest to posttest whereas the control group did not show any changes throughout the tests. Raters detected an improvement from the pretest to the posttest productions of young adult learners. The results support the report by Hakuta, Bialystok, and Wiley (2003) that adult learners who have passed the putative end of a critical period exhibit verall L2 success.

In the training sessions of this study, the participants were exposed to a great many model sounds. That means acoustic templates must have played an important role in monitoring the articulatory output.

As for the results of acoustic analysis of Group A, the individual-sound-first group, and Group C, the mixed training group, the durational ratios did not change significantly but F0 range ratios approximated the model sound's ratios throughout the tests. These two groups obtained better subjective evaluations in the posttest by raters.

The importance of the order of training: the apparent priority of training prosodic features over individual sounds

Group B, the prosody-first group, showed remarkable results in the three analyses of durational ratios, F0 range ratios and subjective evaluations. In the pretest, all groups were homogeneous in English proficiency as confirmed by acoustic analysis and subjective evaluations. However, as the training sessions went by, the participants of this group began to exhibit differences from the other groups. What does this indicate? The present data may not provide conclusive evidence to support the hypothesis that prosody training should be prior to individual sounds training because the tasks in this study were quite different. Two different pieces of software were used. In the prosody training, participants can check their voice recordings acoustically and visually against pitch contours, wave forms, and intensity on the computer display. Thus they had some feedback. On the other hand, in the individual sounds training, participants read instructions on the screen and practice without recording. In the latter task, the quality of feedback that the participants received was different. Further experiment is needed to confirm the priority of training prosody over segmentals by using the same task.

Chapter 4

General Discussion and Implications for Second Language Pronunciation Teaching

4.1. Summary of findings of the three studies

THE summarized findings of Study 1, Study 2, and Study 3 follow below.

Finding 1: The significant effects of the model reading on the participants' production as to prosodic features (from Study 1)

The duration of sentences, consonants, and vowels approximated that of the model reading. As for the unfamiliar non-native consonants, there was no significant change between Before Listening (BL) and After Listening (AL). That means the participants could recognize the prosodic features by listening to the model reading only once or twice and produce the sound in a similar way.

Finding 2: The insignificant effects of the model reading on the participants' production as to non-native consonants (from Study 1)

Regarding non-native consonants, most participants could not recognize the difference after listening to the unfamiliar model sounds once or twice only. In consequence, they couldn't produce them.

Finding 3: The significant differences between the low-pass-filtered training (LP) group and the non-filtered training (NF) group in acoustic analyses (from Study 2)

The results of the acoustic analyses of the experiment revealed significant differences between the LP group and the NF group. As for the duration of the eight sentences in the posttest, the LP group was significantly different from the NP group and they approximated more to the model sounds. By comparing pitch contours of the pre/posttests, the LP group showed closer approximation of the model sounds.

Finding 4: The insignificant differences between the low-pass-filtered training (LP) group and non-filtered training (NF) group in subjective evaluations (from Study 2)

As for subjective evaluations, there were no significant differences between the LP group and NF group. All raters recognized that the production of each participant in posttest sounded more natural as English than in pretest. They found that both groups of participants improved their English pronunciation.

Finding 5: The high acceptability of the low-pass filtered training by participants (from Study 2)

At the conclusion of the training program, all participants were given a questionnaire to complete and return to me anonymously. The low-pass filtered training was positively accepted by all the participants. Initially, some of them found it tiring. The LP sounds felt uncomfortable because they couldn't understand what was being said. However, as they attended the training sessions, they became more used to them.

Finding 6: Importance of the order of training: the apparent priority of training prosodic features over individual sounds (from Study 3)

Group B, the prosody-first group, showed appreciable results in the three analyses of durational ratios, F0 range ratios and subjective evaluations. In the pretest, all groups were homogeneous in English proficiency as confirmed by acoustic analysis and subjective evaluations. However, as the training sessions developed, the participants of this group began to exhibit differences from the other groups.

Finding 7: The consistent effects of training (from Study 3)

Regarding the effects of the training, particularly in the subjective evaluations, all the experimental groups showed significant differences from pretest to posttest whereas the control group did not show any changes throughout the tests. Raters detected an improvement from the pretest to the posttest productions of young adult learners. The mean evaluation score of Group B was the highest among the groups.

4.2. General discussion of results of the three studies

4.2.1. The order of training

The major finding of this book is Finding 6 from Study 3 described in the previous section (4.1), which demonstrated the importance of the order of training. In that study, two groups of participants received two kinds of trainings in different order, segmental-training first in Group A and prosody-training first in Group B. Surprisingly, even though both groups received the same amount of training, the training effect as measured by the improvement from pretest to post-test differed significantly between the groups. The degree of improvement in Group B, which received prosody-training first, was significantly greater than the other groups. So far as I know, this is the first study to show the significant effects of the order of training tasks in speech production training.

Note that two factors, processing level factor (prosody vs. segment) and interactiveness factor (high vs. low), covary in the current study. The group which showed the greater improvement received "prosody training" first and "segmental training" was introduced subsequently. However, the "prosody training" in the current study used a highly interactive task, in which participants reproduced words and sentences after listening to the speech model, and the acoustical characteristics of the model's and the participant's utterances were given as instant feedback. In contrast, the "segmental training" used a rather low-interactive task, in which participants listened to the speech models and were instructed how to reproduce them. There was no feedback for their utterances. Thus, the above results demonstrate that training which introduced "prosody training" and/or "high-interactive production training" first is more effective than training which introduced "segmental training" and/or "low-interactive production training" first. Because of this covariance, it is somewhat difficult to conclude which factor contributed more to the effects of training.

Although further experiments are expected, it may be said that priority of suprasegmentals in the early stages of pronunciation training did not hinder the learning process of the participants, and I believe that this training order will contribute to modify the structure of the learner's phonetic system. Previous studies have investigated the relationship between perception and production of non-native contrasts such as /r/ and /l/. For example, in perceptual training for non-native contrasts, Yamada et al (1994) used the high-variability training technique and found significant improvements in the Japanese trainees' productions of /r/ and /l/ as a result. Thus, the effectiveness of training individual sounds is clearly apparent. If the order-effect hypothesis that prosody training should be prior to individual sounds (i.e. non-native contrasts) training is proven correct, then this study will have contributed to develop a more effective training program on computer or one that can be used in the classroom.

4.2.2. Low-pass experiment and quantity of training

A second important finding concerns the low-pass experiment (Finding 3 and Finding 4). The results of acoustic analyses showed a significant difference between the pretest and posttest, but as to the results of the subjective evaluations, there was no significant difference between low-pass training participants and non-filtered participants. To account for the difference in results, more analyses are needed. Yet, we could hypothesize that the training period might not have been long enough. For example the results of Study 3 were more conspicuous than those of Study 2. The training period of Study 3 was approximately twice as long as that of Study 2. Thus, the quantity of training was different. If the "high-variability" approach that presents the trainee with many exemplars of the target categories over a lengthy period is adopted as found in

Bradlow et al (1997), the results of the low-pass experiment would differ from those of Study 2. The quantity of training must be considered when discussing the results of any experiments.

4.2.3. Low-pass experiment and accuracy

Finding 3 provides some indication that prosody training using low-pass filters contributed to acquiring accuracy of pronunciation. The results suggest that we could apply low-pass filtering to English prosody training in the interests of accuracy. The efficient collection of more data remains as a challenge for future work.

4.2.4. Acoustic analyses and subjective evaluations

Both acoustic analyses and subjective evaluations are commonly adopted in analyzing acoustic experiments. Sometimes the results of these two differ, as seen in Study 2 and Study 3 in this book. The precision of computer speech analysis is becoming ever more accurate. In this experiment, raters were required to focus on how natural the utterance sounded as English, and didn't concentrate on the duration of sentences. The criteria, in this case, were substantially different. These questions have to be taken into consideration in each case rather than automatically adopting acoustic analyses or subjective evaluations.

4.2.5. The importance of suprasegmentals and segmentals in L2 production

One of the pedagogical implications from Finding 1 and Finding 2 is that both suprasegmentals and segmentals are indispensable in second language production. Many segmentals are vital for the preservation of phonological intelligibility. On the other hand, segmental errors have a rather less serious effect on intelligibility than do suprasegmental error as Brown (1991) states.

4.3. Conclusion: Implications for second language pronunciation teaching

As mentioned in the Introduction, as a teacher of English my goal in this research was to develop new techniques for modifying the structure of the learner's phonetic system and to investigate the nature of prosody in speech perception and production.

In applying my findings to the classroom, or in developing a perceptual and productive training program, I would like to make the following suggestions based on the three studies of the present book.

4.3.1. Communicative pronunciation teaching for the classroom

In the early stages of L2 speech training, I suggest that pronunciation tasks be combined and integrated with other communicative grammar, vocabulary, listening, and situational/functional exercises. Integrated exercises are more likely to be communicatively meaningful, and are necessary due to constraints on our classroom time. We should encourage our students to shift their attention from the segmental to the suprasegmental aspects of pronunciation. We focus our attention on the meanings carried by the different areas of English pronunciation within the context of communication, and attempt to integrate this focus as much as possible with our other communicative learning activities. Rounds of minimal-pair practice cannot be recommended at these stages. I would like to suggest that our students be made aware of the communicative value of intonation and placement of emphatic and contrastive stress within sentences. Every dialogue that is taught is an opportunity to do this, and each should be introduced with both demonstration and practice relating to how its intonation contour contributes to meaning and communicative function within the context. Once variation between stressed and unstressed syllables has been introduced, and in particular, once unstressed vowels have begun to be reduced or even deleted, the obtrusive insertion of vowels within consonant clusters and at the end of words will subside.

Following this, individual sounds should be introduced. We should make our students aware of those sounds and contrasts that are missing in Japanese, through the careful and contextualized contrast with minimal pairs

4.3.2. Developing communicative pronunciation software

Many language classrooms are severely constrained by lack of class time and large class sizes. Using pronunciation software in or out of class offers a solution for time and class size management. I have used computer software throughout the three studies. I also created pronunciation training software as in Study 3. However, I do not claim that computer-assisted training is superior to other approaches. A question might arise as to whether non computer-based training approaches such as traditional teacher-led instruction would be as or more effective. One might attempt to compare these approaches experimentally. Such a comparison is irrelevant because there are numerous elements that make up an approach.

Introducing suprasegmentals communicatively in the early stages of L2 speech training is recommended when developing software. Meaningful dialogues, rather than a collection of isolated sentences or phrases, should provide the contents. With these contents, students will learn prosodic features of English naturally. Following this, individual sounds are introduced. In Study 3, the approximate time of practicing suprasegmentals was about 500 minutes (10 training sessions: 50 minutes for each session). This figure is not a conclusive one but may be used as an example.

The goal of teaching pronunciation is not to make learners sound like native speakers of English. There are indeed some groups of people whose oral communication needs mandate a high level of intelligibility such as diplomats, international business people, and teachers of English as a foreign language. However, the realistic goal in general is to enable learners to communicate with intelligible pronunciation. I believe that pronunciation teaching should support the autonomy of students through self-study. Computers are the ideal medium to teach pronunciation because users can learn pronunciation on their own without fear, hesitation, or embarrassment. It is my hope that the findings of this study can contribute to develop better practices, and shed light on our understanding of the process of speech perception and production in general.

Chapter 5

Teaching English Pronunciation for Global Communication

5.1. Introduction

ENGLISH as an international language now plays an important role in the lives of millions of speakers in the world who are not very proficient in the language. They communicate with other non-proficient users in a wide range of different fields such as business, science, and official negotiations. Some sociolinguists call this trend "linguistic imperialism," and many of them try to address the problem that a privileged few enjoy the benefits of globalization. On the other hand, there are defenders of "Standard English" who insist on protecting their sole property, Anglo-American language and culture. In this complex situation, what should we, as ELT practitioners, do? The teaching of English pronunciation is especially problematic because verbal communication is the main topic of the above argument.

In this article, I would like to reappraise the model for the teaching of English pronunciation from the perspective of a non-native instructor of the English language. Firstly, we will see the degree to which the English language has spread all over the world. Secondly, the pros and cons of choosing a model for the pronunciation of English will be examined, and the question of whether a model is still necessary or useful will be asked. Then finally, I would like to state my own proposal regarding this issue.

5.2. The spread of English: Has English swept away all else before it?

According to Crystal (1997), about 2,090,000,000 people (well over a third of the world's population) are, as he puts it "routinely exposed to English." As he points out, what is impressive about this staggering figure is "not so much the grand total but the speed with which expansion has taken place since the 1950s (Crystal 1997, p. 61)." In 1950, making the case for English being a world language would have hardly been possible. Within a little more than a generation, the case is entirely plausible. So what happened to the language? Kachru has dealt with this issue in many of his writings[1]. He has classified and visualized the spread of English around the world as three concentric circles (Figure 5.1). They represent different ways in which the language has been acquired and is currently used.

The inner circle covers the historical bases of English in the area where it is the first language: the USA, UK, Ireland, Canada, Australia, and New Zealand.

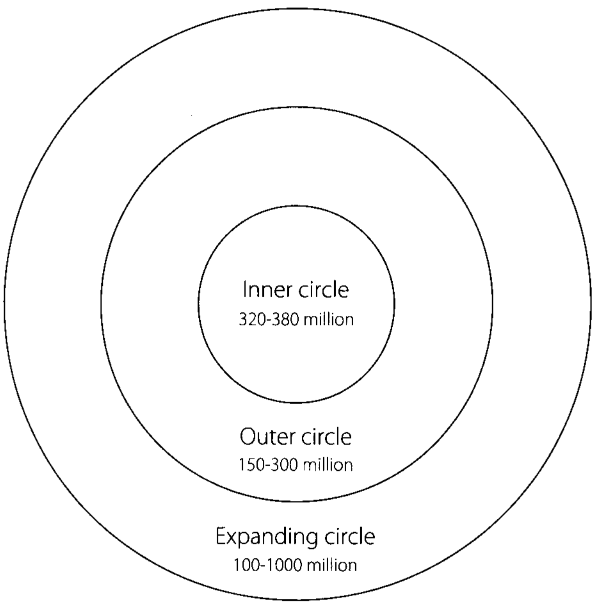


Figure 5.1. *Kachru's model*

The outer circle comprises regions, colonized mainly by Britain, where English is used as a second language in non-native settings and has become part of the country's chief institutions: India, Singapore, and numerous other territories mainly in Africa, South-east Asia and the South Pacific.

The expanding circle refers to the nations where English is recognized as an international language but which have never been colonized[2] by the inner circle countries, and in which English does not have special status in their language policies. English is taught as a foreign language in these areas: China, Russia, Japan, Greece, Poland and other countries.

Estimates for the total number of each category vary enormously, especially in the expanding circle: as low as 100 million and as high as 1 billion. This gap exists because the criterion for defining English use differs. If we demand that people in the expanding circle have "near-native fluency," the lowest total would be chosen. The same can be applied to the case of the outer circle. A reasonable command of English would expand the grand total, though the grand total is, in fact, less accurate because estimates are not available for many countries in this category.

Not all the countries would fit in Kachru's model, but it is widely recognized as a useful classification. The expansion of English indicated above in the concentric circles is inarguably the result of the movement of English around the world. It started with the pioneer voyages to America and Asia, and continued with the colonization of Africa and the South

Pacific in the nineteenth century. There were wars or movements of independence in those colonized countries in the mid-twentieth century where English naturally achieved the status of being a second language. In this section, I will not focus on the historical background of the spread of the language, but examine instead how much it has expanded all over the world. As Crystal (1997) says, the present-day world status of English is primarily the result of two factors: the expansion of British colonial power and the emergence of the United States as the leading economic power of the twentieth century. If we consider the diffusion of the language as a form of a genealogical-tree-like list, it can be indicated as Table 5.1. Countries listed are from the inner circle or the outer circle. The names of the countries are from an atlas published in Japan (Teikoku-shoin editorial department, 2001).

Table 5.1 refers only to English as L1 and L2, and we should notice the dramatic increase of speakers of English in the twentieth century has been recognized as coming from the expanding circle. A geographical and historical survey will help us to understand the spread of English in the past, but questions more important to us are related to the expanding circle. We can predict various fields in which English is a significant medium, for instance, international organizations, the worldwide computing network (i.e. the Internet), business transactions, scientific periodicals, motion pictures, popular music and so forth. Here I will focus on two major fields: international organizations and the Internet,

Table 5.1. *British and American English as L1 and L2*

English	English	English
British	British	American
(Europe)	(Asia)	(North America)
UK	(South Asia)	USA
Ireland	India	Canada
Malta	Pakistan	(Central America)
(Africa)	Sri Lanka	Puerto Rico
(West Africa)	Bangladesh	(Asia)
Cameroon	Grenada	Philippines
Nigeria	(Southeast Asia)	(Africa)
Ghana	Singapore	Liberia
Sierra Leone	Malaysia	(Oceania)
(East Africa)	Brunei	American Samoa
Kenya	(East Asia)	
Uganda	Hong Kong	
Tanzania	(Oceania)	
Rwanda	Australia	
(South Africa)	New Zealand	
Namibia	Papua New Guinea	
Botswana	(Pacific Islands)	
Zimbabwe	Solomon Islands	
Zambia	Tonga	
Malawi	Micronesia	
Seychelles	Marshall Islands	
	Fiji	

because they both play an important role in modern society.

5.2.1. International Organizations

The United Nations has six official languages: English, French, Spanish, Chinese, Russian and Arabic. This organization is overwhelmingly larger than other international gatherings. It consists of over fifty organizations, programs, agencies and as many as other regional commissions. Official languages are, of course, used in their proceedings or documents, but not all of them are regularly adopted. If you access the website of the UN, you will find that quite a large number of the related sites are available only in English.

As for other organizations, wherever they may be based, English is the main or chief auxiliary language (UIA, 2000). For example, in the Association of South-East Asian Nations, English is the working language, though there is no member whose first language is English. This trend is reflected even in Europe, where other languages such as French or German are more likely to be predicted as being more dominant. Yet English is the official language of the European Central Bank, even though the bank is in Frankfurt, and neither Britain nor any other English-speaking country is a member of the European Monetary Fund[3].

5.2.2. The Internet

The USSR launched Sputnik, the first artificial earth satellite, in 1957. In response, the United States formed the Advanced Research Projects Agency (ARPA) within the Department of Defense to establish the US lead in science and technology applicable to the military. In 1968, the agency devised a military research network called the ARPANET, the aim of which was to link important academic and government institutions so they would survive local damage in the event of a nuclear strike. Its language was accordingly English, and when it was opened to people and organizations overseas in the 1980s, the language was chiefly used in forming links to the network. In the meantime, English continues to be the dominant language of the Internet.

It is impossible to confirm the number of English pages on the web. However, according to Deutscher Akademischer Austaush Dienst (DAAD, German Academic Exchange Service), the proportion of English pages in all was 56.4 % in 2003[4]. The proportion is going to become much smaller today, as more people in the world are going on line and using their own languages. For example, there were approximately 10,000,000 web pages written in Japanese in January 2000 (Yamamoto, Kitajima, Takagi, and Cho, 2001). Yet this figure is still minor compared to the total number of the web pages; even in 1998, there were around 320,000,000 pages on the net[5]. We Japanese know that our language is a minor presence on line and it is quite common for us to attach an English version when we make a web page, but the reverse hardly ever happens.

In this section, we have seen how much English has spread all over the world, and we may conclude that the language has reached a status that no other language could compete with. Yet I am rather doubtful that this status will last forever. There are over 7,000 languages in the world now[6], so how can we be so confident about the eternal dominance of one language? No one can deny the possibility that a non-English-speaking country will overwhelm the globe in a few centuries. Then the situation will change and English would not be able to maintain its status as a world language.

So we, as instructors of English, should maintain our perspective in regard to the language. Adding to that, if we adopt the phrase "lingua franca" to English, we should notice that English is no longer the sole property of people of the inner circle. Conceptual and practical changes are needed for the teaching of English language. This will be related to the topic of the next section.

5.3. Pros and cons of choosing a model for English pronunciation

5.3.1. The conservatives

The British people are well known for being sensitive about how they and others speak English. Accent differences seem to receive more attention in that country than anywhere in the world. Thus, if we define English as a genuinely global language in which usage and accents are not restricted, it is natural that such people should feel uncomfortable. It is also understandable that people feel their language is abused when other people use it at their own discretion, and in the end they announce to them that it is no longer their own language. We may agree that we have a preference for the usage or accent of our own language, yet it is going too far to accept the comment by a radical defender of "Standard English" such as John Honey (1997), who says that if only people would speak "Standard English," then rich vistas of intellectual development would open up before them. For Honey, some sociolinguists are viewed as "enemies" of Standard English and he sees a conspiracy among them to keep non-standard English speaking people in their deprived state by persuading them that there is no need to learn Standard English. Of course there are many other more sensible linguists like Peter Trudgill who explains his own ideas about Standard English as follows: "the social and educational role of Standard English in modern society should be dealt with, and the benefits of mastery of Standard English stressed[7]."

So which variety of English will be chosen among the moderate conservatives as the model of English? If learners of the language are in the USA, General American (GA) will be widely accepted, and if in Britain, Standard British English (SBE) will be preferred. Received Pronunciation (RP) was once held up as the most desirable model of English for teaching purposes because it was based on the educated pronunciation of London and the Home Countries. Nowadays, however, there is general agreement that this form of English was an ideological and increasingly redundant model of language and that it is not a realistic model to pursue. The emphasis was then placed on SBE, a less artificial but still regional and class-based (Southern, middle class) English. There is a view that this model is also becoming outdated and that British English is changing again towards what is known as Estuary English[8]. In reality, however, most schools or institutions of higher education at present seem to stay with SBE for teaching purposes to avoid a wide-ranging sociolinguistic debate.

5.3.2. The radicals or reformers

When *Linguistic Imperialism* by Robert Phillipson was released in 1992, it elicited a great variety of opinions. While some felt inspired by the book, believing that it was a necessary blow to those who had been teaching English but were unaware of their political role in the spread of the language and the cultures it represents, others reacted against the claims that those in the teaching profession were emissaries of the hegemonic power of the United States and Great Britain over the developing countries. Phillipson's view was rather pessimistic and deterministic. He adopted a suspicious attitude toward the subject of global English, repeating that "Global English is a myth." His purpose was not to discourage the teaching of English as a foreign language, but to help instructors achieve an awareness of the potentially harmful effects of their undertakings. He believes that educators need to understand that English is not culturally neutral but carries with it assumptions that serve to promote the interests of the rich and powerful. The language is too closely connected to the workings of late capitalism and the needs of multinational corporations, as well as to the voices of authority from the United Nations.

Societies that have adopted the idea that English is a necessary skill have automatically experienced inequalities and emotional distress, Phillipson says. At the lowest level, students throughout Asia are familiar with the trauma of failure. English speaking academics and critics dominate the world of academic publishing. Phillipson notes that "International scholars are disadvantaged when forced to write in English." He also says that the postulation of native speaker superiority, sometimes known as linguicism' is grounded in the belief systems left behind by colonial imperialists.

Pennycook (1994), an acute linguist like Phillipson, however, criticizes Phillipson for not adequately considering how English can be used in diverse contexts.

Pennycook discusses the spread of English in terms of the concept of "center" and "periphery." He points out how English media from developed countries have penetrated the media of developing nations.

This one-way flow of information erodes the national sovereignty, cultural identity, and political independence of developing nations. Institutions on the periphery tend to become distributors of knowledge from the center, but Pennycook emphasizes that the actual situation is often more complex. Many institutions in the third world are more than passive information receivers. He also disputes the assumptions that the way English spreading is natural, neutral or beneficial. A theme of his book is the nature of education. Instead of viewing schools as neutral sites where a curricular body of information is passed on to students, he urges readers to regard them as cultural and political arenas in which different values are in struggle.

His work is thought-provoking, but he doesn't offer any prescription for teaching methodology. He just expresses the negative vision of his personal philosophy toward the language. Other people, like Samuel Ahulu, are more realistic. Ahulu (1997) suggests that the concept of Standard English should be redefined. According to his view, English is now a global language and it has developed and continues to develop forms or features divergent from British or American English. He claims that it is unrealistic to treat any divergence as errors. Standard English should accommodate to the developments of new Englishes. The validity of those new Englishes should be seen as styles of speech or expression that make up a part of the speakers' repertoire. They should not be thought of errors. English lacks standard codification that would reflect its international character adequately. Thus, one of the major problems with new Englishes appears to be the issue of codification (Ahulu 1997, pp. 17–19).

Jenkins (2000) claims that Non-Bilingual English Speakers (NBES) now outnumber native speakers of English and that their needs must be considered when the nature of English as an International Language (EIL) is defined. It is no longer acceptable to establish the structure of EIL from the perspective of the regions or nations of traditional native speaker populations. Jenkins suggests that we need to evaluate which parts of English pronunciation are essential for interaction between NBESs and which can be regarded as peripheral. She proposes that we need to develop a Lingua Franca Core (LFC) for English, choosing the phonological features that are necessary for communication, and that in the future English language teaching should be based on the LFC. She even tries to exclude the non-core features from the model pronunciation. The problem here is that the selection of features that constitute the LFC is highly controversial. For example, dental fricatives /θ/ and /ð/ will be eliminated from LFC because these are rather rare in languages around the world and hard to pronounce for non-native speakers. However, if the dental fricatives are to be replaced in the pronunciation core, which sounds should be the alternatives? I have been teaching English pronunciation to Japanese students for years and found that most of them use /s/ for /θ/ and /z/ for /ð/, so if we encourage the replacement of the dental fricatives with alveolar fricatives, what will happen in communication with other NBES? Some other NBES who use alveolar plosives /t/ and /d/ for /θ/ and /ð/, will actually be less understandable when they communicate with Japanese learners who are taught alveolar fricatives for dental fricatives. If it is too difficult or almost impossible to pronounce dental fricatives, the alternatives should be accepted. However, with a little correction, Japanese learners' pronunciation of dental fricatives is easily corrected. Actually, the difficulty of /θ/ and /ð/ lies not so much in their articulation, which most learners can perform in isolation, as in their combination with other fricatives such as /s/ and /z/. So if they practice with drills containing such combinations involving rapid tongue glides, for example, /s/±/θ/ as in "this thing," or /s/+/ð/ as in "pass the salt," they can gradually pronounce /θ/ and /ð/ accurately without much difficulty.

What is more controversial about Jenkins's claim is that word stress and intonation, apart from the placement of nuclear stress, do not cause problems of misunderstanding between NBESs and so should be taken off from the LFC. It might be wrong in the interaction between NBESs whose first languages are different. I myself sometimes have difficulty understanding what French people are saying when they speak in English because their stress of words is affected by the French language. Therefore, the reverse must also be true.

Jenkins' attitude towards NBES is very generous and she also thinks highly of non-native teachers of English because they have undergone the painful process and stress of learning English, and, in consequence, they are more attuned to the needs of their learners than native speakers. She also suggests that native speakers of English are now largely irrelevant, and non-native speakers should actually provide the new model for EIL. She also proposes that native speakers need to undergo retraining so that they use the modified pronunciation of the LEC. Her suggestions must be highly unpopular among teachers of English all over the world, yet her books are worth reading because of the fact that her arguments are based on substantial empirical findings.

5.4. My proposal

If you like to play the piano, you may start lessons with a traditional acoustic piano or a digital keyboard. Whichever keyboard you may choose, the arrangement of keys is invariable all over the world. First of all, musical scales are universal. Thus, a pianist from China and a pianist from South Africa can play the notes of a piece of music identically, though the interpretation of the emotional language of the piece may differ. If you turn out to be a very talented player, you may have a chance to join a world-famous competition such as Chopin's International Piano Contest in Warsaw without any interpreter on the stage. Music is universal because the scales are the same all over the world.

It may not be appropriate to compare language learning to piano lesson because language learning is linked to more serious fields: politics, economics, or the history of human beings. Languages can be more related to the origins of conflict than music. When I think of a model of English pronunciation, however, I always imagine a simple piano lesson. Some linguists are overly conceptual or emotional when it comes to arguments about the English language. The spread of English is no longer in the hands of the educators or ideologues. It is already out there. Information technology is introducing a new way to approach the language outside the classroom. In my opinion, if English as an international language is simply a communicative tool, we should teach basic English pronunciation just as scales in music so that learners make themselves intelligible in English; at the same time, we should allow them to retain their phonological characteristics to the greatest degree possible. A model for pronunciation is all the more needed for better intelligibility in a global situation, but it is not a goal for learners.

As for choosing a model pronunciation of English, I would rather stay in a neutral position. I cannot be a strong supporter of Standard English because Daniel Jones' good old days are gone and the definition of "standard" is hardly tenable today, as we have seen in former sections, though I personally prefer a British accent. I can hardly accept the notion of "core-based English" suggested by Jenkins because it will surely cause additional communication problems, even among non-native speakers of English. It is not a good idea to restrict usage, vocabulary, and pronunciation of a language intentionally. A language changes in its own way. If we face up to reality, we will find that numerous people want to learn a "prestigious" variety of English which some sociolinguists hate. There are many language schools in Britain which provide Standard British English as a model of English, and those schools are popular among people from Europe, Arab countries, Africa and Asia. I wonder why a number of sociolinguists ignore the majority opinion. I cannot agree with them on this point, but it is true, as they say, that it is impossible to learn a foreign language without being influenced ideologically, politically and culturally, so the spread of English can potentially wreak havoc on a number of languages and cultures.

Let us turn now to the implications for the classroom. In my case, I teach English pronunciation basically according to one of the major varieties of English, General American, because American English is dominant and popular in Japan and there are two major language proficiency tests (TOEFL and TOEIC) which are made by the English Testing Service, a semi-governmental company in the USA. However, I have never forced my students to pronounce the sounds "correctly." I just show them how to pronounce the target sounds, make them practice, and advise them individually. Some of the students make great progress in a short period, others not. For the students with a thick accent, I reassure them of the legitimacy of their varieties, since not doing so will perpetuate the idea that to "correctly" use English they must turn their English into whatever concept they have of American English. Such a belief frustrates students because in most instances their performance never equals their mental model of what an ideal English should sound like. Since pronunciation correction is an optional task in a listening class, I don't grade their pronunciation. I sometimes show them videos to expose them to other major varieties of British or Australian English, or Asian varieties such as Indian or Singaporean English. I think it is of paramount importance to make them familiar with a wide range of varieties of spoken English.

We now find ourselves faced with serious challenges. Some people may criticize of my proposal toward teaching English pronunciation for being inconsistent and yielding, and, indeed, it is a compromise plan. In this fast changing world, however, isn't it rather realistic? Above all, we should think of the learners' benefit first. We should support them in their efforts to participate on the global stage and, at the same time, preserve their unique identities. To a great extent, success or failure depends on us, ELT practitioners, though as I mentioned before, English has spread outside the classroom. In view of the great varieties of pronunciation that teachers and learners will be expected to deal with, we will have to be even more phonetically aware than we are now. I am not suggesting that we should insist on accuracy or check our students' production against some imaginary model of perfection, but we should know more about varieties of English.

Superficiality is hardly avoidable when attempting to cover such a topic within approximately 20 pages, and we all know there is no perfect remedy to the issue of a model of teaching English pronunciation. However, I hope this article will be thought-provoking enough to raise the reader's awareness of this important topic.

Notes

1 Braj B. Kachru is currently a Professor of Linguistics at University of Illinois. He has written a large number of publications about English as an International language, starting from the early 1980s (See Kachru (1986, 1992)).

2 In the strict sense of the word, Japan was under US-led occupation after World War II (1945–52).

3 Cited from the website written by Barbara Wallraff (http://www.theatlantic.com). The website was accessible in August 2001.

4 DAAD Tokyo (http://tokyo.daad.de/wp/grunde_zum_deutschlernen/).

5 World Wide Web User Statistics (http://www.why-not.com).The website was accessible in August 2001.

6 Michael Ashby referred to this issue in his lecture at University College London during a summer course on English Linguistics and Phonetics held in August 2001.

7 http://www.phon.ucl.ac.uk/home/estuary/honeyrev.htm

8 This variety of English is spoken in the East of London on both sides of the Thames River.

References

Ahulu, S. (1997). General English. A consideration of the nature of English as an international medium. *English Today, 13*(1), 17–24.

Akahane-Yamada, R., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic presentation and HMM-based speech recognition scores. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, 1998 [CD-Rom] Paper 0429.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 345–360.

Bohn, O. S., & Munro, M. J. (2007). *Language Experience in Second*

Language Speech Learning: In honor of James Emil Flege. Philadelphia: John Benjamins Publishing Company.

Bourdieu, P. (1991). *Language and symbolic power.* (Translated by J. B. Thompson), Cambridge, MA: Harvard University Press.

Bourdieu, P. (1995). *The logic of practice.* Palo Alto: Stanford University Press.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tokuhara, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America, 101,* 2299–2310.

Brown, A. (1991). *Pronunciation Models.* Singapore: Singapore University Press.

Celce-Murcia, M., Briton, D. M., & Goodwin, J. M. (1996). *Teaching Pronunciation.* New York: Cambridge University Press.

Crystal, D. (1997). *English as a Global Language.* Cambridge: Cambridge University Press.

Dauer, R. (1983). Stress-timed and syllable-timing reanalyzed. *Journal of Phonetics, 11,* 51–62.

Derrida, J. (1982). Différance. In Derrida, J. (Ed.), *Margins of philosophy* (pp. 1–27). Chicago: The University of Chicago Press.

Flege, J. (1988). The production and perception of foreign language speech sounds. In H. Wintz (Ed.), *Human communication and its disorders.* Norwood, NJ: Ablex.

Flege, J. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). Timonium, MD: York Press.

Freadman, A. (1994). *Models of genre for language teaching.* Paper delivered as The Sonia Marks Memorial Lecture. Sydney: University of Sydney.

Gimson, A. C. (1981). *An introduction to the pronunciation of English.* 3rd ed. London: Edward Arnold.

Guberina, P. (1972). *Restricted bands of frequencies in auditory rehabilitation of deaf.* Zagreb: Institute of Phonetics Faculty of Arts.

Guberina, P. (1976). Structuration et dépassement des structures perceptives et psycholinguistiques dans la méthodologie SGAV. In *Actes du 3e Colloque international SGAV pour l'enseignement des langues* (pp. 41–58). Paris: Didier.

Hakuta, K., Bialystok, E., & Wiley E., (2003). Critical evidence: A test of critical period hypothesis for second-language acquisition. *Psychological Science, 14*, 31–38.

Hardison, D. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics, 24,* 495–522.

Honey, J. (1997). *Language is Power: The Story of Standard English and its Enemies.* London: Faber & Faber.

Hunter, M. D., Phang, S., Lee, K., & Woodruff, P. W. R. (2005). Gender-specific sensitivity to low frequencies in male speech. *Neuroscience Letters, 375*(3), 148–150.

Iba, M. (2002). Teaching English pronunciation for global communication. *The Journal of the Institute for Language and Culture, 6*, 21–32.

Iba, M. (2007). The influence of model sounds on the speech production of Japanese learners of English. *The Journal of the Institute for Language and Culture, 12*, 45–66.

Jenkins, J. (2000). *The phonology of English as an International language.* Oxford: Oxford University Press.

Joudry, R. (2004). *Why Aren't I Learning?* Perth: Sound Therapy International.

Kachru, B. (1986). *The Alchemy of English: The Spread, Functions, and Models of Non-Native English.* London: Pergamon Press.

Kachru, B. (1992). *The Other Tongue: English Across Cultures.* Urbana: University of Illinois Press.

Keith, R. W. (2000). *SCAN-C: Test for Auditory Processing Disorders in Children-Revised.* San Antonio: Psychological Corporation.

Keith, R. W., & Farrer, S. (1981). Filtered word testing in the assessment of children with central auditory disorders. *Ear Hear, 12*, 267–269.

Ladefoged, P. (1982). *A Course in phonetics.* 2nd ed. New York: Harcourt

Brace Jovanovich.

Ladefoged, P. (2003). *Phonetic data analysis.* Oxford: Brackwell Publishing.

Lian, A. B. (2003). *Beyond illusions and facts: Toward a methodology of dialogue and dialogue-enhancing environments.* Paper presented at The International Conference on Computers and Philosophy, ANU, Australia and Rice University, Texas, USA. Retrieved from http://www.anialian.com/On_Time.html

McKay, S. L. (2002). *Teaching English as an International Language.* New York: Oxford University Press.

Moyer, A. (1999). Ultimate attainment in L2 phonology. The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition, 21*, 81–108.

Munro, M. J. & Tracey M. Derwing (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning, 45*, 73–97.

Pennycook, A. (1994). *The Cultural Politics of English as an International Language.* Harlow: Longman Group Limited.

Phillipson, R. (1992). *Linguistic Imperialism.* Oxford: Oxford University Press.

Roach, P. (1991). *English phonetics and phonology.* 2nd ed. Cambridge: Cambridge University Press.

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379–410). Timonium, MD: York Press.

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech, Language, and Hearing Research, 37,* 347–357.

Spolsky, B. (1999). *Concise encyclopedia of educational linguistics.* Oxford: Elservier Science Ltd.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36,* 131–145.

Teikoku-shoin editorial department. (2001). 旅に出たくなる地図　世界 [Maps to Make You Want To Travel]. Tokyo: Teikoku-shoin.

Ueyama, M. (1996). Phrase-final lengthening and stress-timed shortening in the speech of native speakers and Japanese learners of English. *Proceeding of ICSLP,* vol. 2, 610–613.

UIA. (2000). *The Union of International Organizasions' Yearbook 2000–2001.* Munich: K. G. Saur Verlag.

Yamada, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *Journal of the Acoustical Society of America, 93,* 2391(A).

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke III, W. D. (1994). The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference on Spoken Language Processing*, Yokohama, Japan, 1994, 2023–2026.

Yamada, R. A., McDermott, E., Adachi, T., Kawahara, H., & Pruitt, J. S. (1998). Computer-based second language production training by using spectrographic presentation and MM-based speech recognition scores. *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, 1998 [CD-Rom] Paper 0429.

Yamamoto, T., Kitajima, M., Takagi, H., & Cho, Y. (2001). Markov 連鎖を用いたウェッブナビゲーション過程の評価 [An evaluation of the process of web navigation using the Markov chains]. Retrieved from http://staff.aist.go.jp/kitajima.muneo/Japanese/PAPERS(J)/Qsymposium2001.html

Appendix

Table 1. *Test sentences (study 1)*

(a) Before Listening to the model reading

No.	Sentence
1	Thank you very much for everything.
2	I bought them drinks.
3	What's she saying?
4	They are all afraid.
5	What would you like to do?
6	Give it to him.
7	That is better than that.
8	I know it's true.
9	think/sink
10	right/light
11	He went over the path.
12	They thought about it.
13	We've fired them.
14	Repeat the word.
15	I knew it was wrong.
16	He didn't collect the papers.
17	He didn't correct the papers.
18	Put all these things in the bag.
19	food/hood
20	cars/cards

(b) After Listening to the model reading

No.	Sentence
21	He went over the path.
22	They thought about it.
23	We've fired them.
24	Repeat the word.
25	I knew it was wrong.
26	He didn't collect the papers.
27	He didn't correct the papers.
28	Put all these things in the bag.
29	food/hood
30	cars/cards
31	Look at the train.
32	Don't disturb them while they are praying.
33	Humpty Dumpty sat on a wall.
34	Humpty Dumpty had a great fall.
35	All the king's horses and all the king's men.
36	Couldn't put Humpty together again.
37	What a wonderful life he lived!
38	How beautiful you are!
39	ban/van
40	deaf/death

Note. There are two main reasons why No. 1–10 and No. 31–40 were chosen and recorded. One is because the participants would not concentrate on the forcal ten materials in Table 1.1, and speak naturally. The other is because No. 1–19 and No. 31–40 can be used for further experiments.

Table 2. *The duration of Sentence 1, 2, 3 and 4 (ms)*

	Sentence 1	Sentence 2	Sentence 3	Sentence 4
Model	1685	1158	1068	1101

	Sentence 1		Sentence 2		Sentence 3		Sentence 4	
Participant	BL	AL	BL	AL	BL	AL	BL	AL
P1	1297	1763	1206	1117	1112	1217	1100	1133
P2	1587	1562	1320	1108	1983	1412	886	1007
P3	1519	1668	1239	1369	937	1026	885	1198
P4	1778	1565	1580	1185	1292	1203	998	1099
P5	1424	1479	1252	1059	949	950	949	1090
P6	1605	1570	1326	1162	1407	1110	964	1152
P7	1475	1379	1170	1200	1209	1057	865	916
P8	1600	1869	1467	1100	1270	1225	1026	1039
P9	1920	1361	1261	1171	1214	1229	938	972
P10	1492	1716	1280	1195	1218	1236	902	1077
P11	1541	1667	1132	1131	1126	1145	976	1135
P12	1798	1535	1414	1066	1045	1101	984	942
P13	2584	1604	1613	1311	2117	1415	2565	1030
P14	1468	1323	1681	1160	1399	1275	1006	1095
P15	1404	1333	1217	1307	1141	1057	1170	966
P16	1617	1491	1616	1459	1506	1442	1104	1228
P17	1805	1546	1637	1375	1497	1201	999	978
P18	1404	1461	1292	1263	1123	1142	928	944
P19	1561	1464	1103	1062	1174	937	957	900
P20	1437	1467	1663	1238	1046	1159	898	1065
M	1615.8	1541.15	1373.45	1201.9	1288.25	1176.95	1055	1048.3

Table 3. *The duration of Sentence 5, 6, 7 and 8 (ms)*

	Sentence 5	Sentence 6	Sentence 7	Sentence 8
Model	1118	1608	1761	1683

	Sentence 5		Sentence 6		Sentence 7		Sentence 8	
Participant	BL	AL	BL	AL	BL	AL	BL	AL
P1	1236	1189	2117	1940	1791	1912	1697	1694
P2	1419	1385	1861	1941	1993	1948	2065	2104
P3	1256	1181	2325	2014	2226	1968	1976	1830
P4	1822	1464	2265	2193	2630	1970	1977	1785
P5	1135	1313	1752	1599	1528	1889	2082	1971
P6	1379	1130	2187	2138	2476	2232	2700	2142
P7	1477	1419	2375	1831	2782	1889	3126	1930
P8	1179	1228	1783	1666	1990	1842	2029	1790
P9	1200	1176	1796	1909	1908	1914	2126	1883
P10	1288	1215	1784	1901	1877	1910	2364	1923
P11	1440	1120	1912	1999	2136	1927	2383	2088
P12	1824	1445	1587	1625	1531	1965	3524	1849
P13	1732	1225	2667	1864	2367	2003	3554	2583
P14	1308	1440	1948	2271	2162	1806	1918	1627
P15	1302	1450	2113	1810	2216	1753	2223	1939
P16	1373	1402	2089	1994	2399	2113	2262	2151
P17	1514	1391	1993	2133	1888	2094	2159	2104
P18	1177	1182	1899	1865	2030	1724	2128	1691
P19	1388	1285	1739	1635	1720	1569	1872	1640
P20	1250	1360	2054	2180	2101	2260	2744	1951
M	1384.95	1300	2012.3	1925.4	2087.55	1934.4	2345.45	1933.75

Table 4. *Duration of consonants prior to stressed vowels (ms)*

	Sen. 2 /θ/ in thought	Sen. 3 /f/ in fired	Sen. 8 /ð/in these
Model	141	176	72

	Sen. 2 /θ/ in thought		Sen. 3 /f/ in fired		Sen. 8 /ð/in these	
Participant	BL	AL	BL	AL	BL	AL
P1	122	126	199	177	44	96
P2	95	120	217	187	97	70
P3	118	164	74	163	35	59
P4	186	163	125	188	57	64
P5	121	124	132	186	98	99
P6	140	170	162	142	41	70
P7	105	116	140	147	101	76
P8	211	147	226	197	62	40
P9	120	179	150	134	87	70
P10	117	191	148	166	50	94
P11	105	131	159	174	42	80
P12	152	117	135	125	157	158
P13	170	127	165	170	59	87
P14	141	109	108	120	26	93
P15	92	137	131	152	55	45
P16	42	133	154	198	64	63
P17	158	148	230	164	32	88
P18	104	115	132	125	20	54
P19	63	83	93	105	79	71
P20	73	141	111	189	36	53
M	121.75	137.05	149.55	160.45	62.1	76.5

Table 5. *Duration of consonants prior to unstressed vowels (ms)*

	Sen.1 /ð/ in the	Sen. 4 /ð/ in the	Sen. 8 /ð/ in the
Model	64	38	20

	Sen.1 /ð/ in the		Sen. 4 /ð/ in the		Sen. 8 /ð/ in the	
Participant	BL	AL	BL	AL	BL	AL
P1	62	42	57	40	28	21
P2	27	63	88	32	57	31
P3	47	63	27	35	58	16
P4	21	29	68	55	33	16
P5	58	83	57	46	30	30
P6	68	78	98	39	49	38
P7	95	65	33	59	73	21
P8	52	62	63	65	54	21
P9	49	49	47	63	50	37
P10	39	83	41	45	87	50
P11	54	59	51	38	46	28
P12	39	30	173	59	50	25
P13	131	96	72	59	79	72
P14	94	83	55	137	25	38
P15	49	41	39	30	48	20
P16	80	53	51	38	25	21
P17	46	72	51	40	25	26
P18	63	57	110	37	26	22
P19	52	72	55	68	35	29
P20	46	43	68	43	53	30
M	58.6	61.15	65.2	51.4	46.55	29.6

Table 6. *Duration of phrase-final vowels (ms)*

	Sen.1 /ɚ/ in path	Sen. 4 /ɔ/ in word	Sen. 5 /ɔ̞/ in wrong	Sen. 8 /æ̞/ in bag
Model	280	305	254	324

	Sen.1 /ɚ/ in path		Sen. 4 /ɔ/ in word		Sen. 5 /ɔ̞/ in wrong		Sen. 8 /æ̞/ in bag	
Participant	BL	AL	BL	AL	BL	AL	BL	AL
P1	138	251	249	291	204	196	141	195
P2	140	209	240	327	221	**384**	155	203
P3	236	251	277	372	242	248	210	328
P4	198	230	228	322	249	203	191	268
P5	171	176	236	275	105	206	139	267
P6	176	207	243	238	277	197	208	231
P7	130	155	248	203	212	222	124	223
P8	162	231	282	243	214	258	237	260
P9	159	194	232	303	211	175	128	322
P10	141	156	229	270	215	204	174	290
P11	158	227	239	277	**292**	264	158	283
P12	140	224	182	208	151	233	137	205
P13	134	186	242	230	313	200	140	178
P14	147	83	265	315	186	205	148	196
P15	180	170	246	243	169	205	210	243
P16	185	232	263	267	174	305	167	155
P17	136	227	252	396	245	264	142	208
P18	154	194	186	298	122	248	120	144
P19	165	234	260	216	166	232	186	167
P20	156	187	230	314	146	225	110	174
M	147	219	241.45	280.4	205.7	233.7	161.25	227

Table 7. *Auditory impressions by raters*

	Sentence 1–8			Minimal pair 9, 10		
	BL	AL	Same	BL	AL	Same
Native 1	7	128	25	3	14	23
Native 2	3	139	18	0	9	31
Japanese 1	2	137	21	0	10	30
Japanese 2	2	130	18	0	6	36
M	3.5	134	20.5	0.75	9.75	30

Table 8. *Results of rating of Sentence 1*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	3	2	5	5	6	5	4	5	5	4	6	7	7	6
P2	4	3	4	3	6	4	6	5	5	4	5	5	7	7	6	6
P3	4	3	4	3	4	4	5	4	5	6	6	5	7	7	6	6
P4	4	4	3	3	5	4	4	4	6	5	5	6	7	6	6	6
P5	4	4	4	4	5	5	4	5	5	6	5	5	6	6	6	6
P6	3	3	4	3	5	4	4	4	6	6	5	5	7	6	6	7
P7	4	4	3	4	5	4	4	5	6	6	5	6	7	6	6	6
NF																
P8	3	3	3	4	5	4	4	4	6	5	6	6	6	6	6	6
P9	5	4	5	4	5	5	5	5	6	6	6	5	6	6	6	6
P10	5	5	5	4	6	5	5	5	7	6	6	6	7	6	6	6
P11	2	3	2	3	3	3	3	4	5	4	4	4	5	5	4	5
P12	3	3	3	3	4	5	4	4	6	6	5	5	7	6	6	7
P13	3	2	3	2	3	3	3	3	5	5	4	5	6	5	5	6

Note. P for "participant", R for "rater."

Table 9. *Results of rating of Sentence 2*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	3	3	4	4	4	3	4	5	5	5	5	5	6	5
P2	3	2	2	3	5	3	3	4	4	4	5	5	6	5	6	6
P3	4	3	3	3	4	4	4	4	5	5	6	5	7	6	7	5
P4	4	3	4	3	5	4	4	4	6	5	5	5	6	6	6	6
P5	4	4	4	3	5	4	4	4	5	5	5	5	5	6	6	5
P6	4	4	3	3	5	5	4	4	6	6	6	5	7	6	6	6
P7	4	4	3	4	5	5	4	4	6	6	5	5	7	6	6	6
NF																
P8	3	4	3	3	4	4	4	4	5	6	6	5	6	6	6	6
P9	5	4	4	4	5	5	5	5	6	6	6	6	7	7	6	6
P10	5	4	5	5	6	5	5	5	6	6	6	6	6	6	6	6
P11	3	3	2	3	4	3	3	3	5	5	5	4	6	5	6	5
P12	3	3	4	3	4	4	4	4	6	5	5	5	7	6	5	6
P13	3	3	3	3	3	4	4	4	4	4	5	5	4	4	5	5

Note. P for "participant", R for "rater."

Table 10. *Results of rating of Sentence 3*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	3	3	4	4	3	4	5	6	5	5	6	6	6	6
P2	2	2	2	2	4	3	3	3	5	5	5	5	7	6	5	6
P3	4	4	3	3	4	5	4	4	6	6	5	6	6	6	6	6
P4	3	3	3	3	4	4	4	4	4	5	6	5	5	5	6	6
P5	2	3	3	2	4	3	3	3	4	5	5	4	5	5	5	5
P6	3	3	3	3	4	4	4	4	6	5	5	5	6	6	5	6
P7	4	4	4	4	5	5	5	4	6	6	5	6	6	7	6	6
NF																
P8	3	3	3	3	5	4	4	4	6	5	6	5	6	6	6	6
P9	4	4	3	4	5	5	4	5	6	6	6	6	6	6	6	7
P10	5	4	5	5	5	5	5	5	6	6	6	6	6	6	6	6
P11	2	3	2	2	3	3	3	3	5	4	4	5	5	5	5	5
P12	3	3	3	3	3	4	3	3	5	5	4	5	6	6	5	5
P13	3	3	2	2	3	4	3	3	4	5	4	4	5	5	5	5

Note. P for "participant", R for "rater."

Table 11. *Results of rating of Sentence 4*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	4	4	3	4	5	5	4	5	4	5	5	5	6	6	6	6
P2	4	4	4	4	6	5	5	6	6	6	6	6	7	7	6	7
P3	4	3	3	4	5	4	4	4	6	5	5	6	7	6	6	7
P4	4	4	4	3	6	5	5	4	6	6	6	5	7	7	6	6
P5	4	4	4	4	5	5	4	5	5	6	5	5	6	6	6	6
P6	3	3	3	3	4	4	3	4	5	6	4	5	6	6	5	5
P7	4	3	4	3	4	4	5	4	5	5	6	5	6	6	6	6
NF																
P8	3	3	3	3	4	4	5	4	5	5	5	6	5	6	6	6
P9	4	3	4	4	5	4	5	4	6	5	6	5	6	6	7	6
P10	4	4	3	3	5	5	4	4	6	6	5	5	6	6	5	5
P11	3	3	4	4	3	4	4	5	5	5	5	6	6	6	6	6
P12	2	3	3	2	3	3	3	3	4	5	5	5	6	6	6	6
P13	4	3	4	4	4	4	4	4	5	5	5	5	6	6	6	6

Note. P for "participant", R for "rater."

Table 12. *Results of rating of Sentence 5*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	4	3	5	4	4	4	5	5	5	5	7	6	6	6
P2	3	3	3	3	5	4	4	4	4	4	5	4	5	5	6	5
P3	3	3	3	3	5	4	4	4	6	5	6	5	7	6	6	7
P4	4	3	3	3	5	4	4	4	5	5	5	5	6	6	5	6
P5	2	2	4	3	4	3	4	4	4	4	5	4	5	5	5	5
P6	4	3	4	3	5	4	5	4	6	5	6	6	7	6	6	7
P7	4	3	4	4	4	4	4	4	5	5	5	5	6	6	6	6
NF																
P8	3	3	3	3	3	3	3	3	6	5	6	5	6	6	6	6
P9	3	3	3	3	3	3	3	3	5	5	5	5	5	5	6	5
P10	5	4	4	4	5	4	4	4	6	6	5	6	6	6	5	6
P11	2	2	2	2	2	2	2	2	4	4	4	5	4	5	4	5
P12	2	3	3	3	2	3	3	3	5	5	4	5	7	6	5	5
P13	3	3	3	3	3	3	3	3	5	5	5	5	6	5	6	6

Note. P for "participant", R for "rater."

Table 13. *Results of rating of Sentence 6*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	4	4	3	6	5	5	4	5	6	6	5	7	6	7	6
P2	3	4	3	3	5	4	4	4	6	5	5	5	7	6	6	6
P3	3	3	3	4	4	4	4	5	5	5	5	5	7	6	6	6
P4	3	3	4	3	5	4	5	4	5	5	5	6	7	6	6	7
P5	3	3	5	3	4	4	4	4	4	5	4	5	5	6	5	6
P6	3	3	3	4	4	4	4	4	6	5	5	5	6	6	6	6
P7	3	3	3	3	4	4	4	4	4	5	6	5	4	6	6	6
NF																
P8	4	4	4	4	5	4	5	4	6	5	6	5	6	5	6	6
P9	4	4	4	4	4	4	4	4	5	5	5	5	5	6	5	6
P10	4	4	4	4	5	5	5	5	6	5	6	6	6	5	6	6
P11	3	3	3	3	4	4	4	3	6	5	5	5	6	6	5	6
P12	2	3	3	3	3	4	4	4	6	6	5	5	6	6	6	6
P13	2	3	3	2	3	3	3	3	4	4	5	4	6	5	5	5

Note. P for "participant", R for "rater."

Table 14. *Results of rating of Sentence 7*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	4	4	4	4	6	5	5	5	5	5	6	6	6	6	6	6
P2	3	3	3	3	4	4	4	4	5	5	6	5	7	6	6	6
P3	3	3	3	3	4	4	4	4	5	6	5	5	7	6	6	7
P4	5	4	4	5	5	5	5	5	6	6	6	5	6	6	6	6
P5	4	4	5	3	5	5	4	4	5	6	5	5	6	6	6	6
P6	3	3	3	3	4	4	4	4	6	5	6	6	7	6	6	7
P7	3	3	4	3	4	4	4	4	4	4	5	4	4	5	5	5
NF																
P8	3	3	3	3	4	3	4	4	5	4	5	5	6	5	6	5
P9	5	4	5	5	5	5	5	6	6	5	6	6	6	6	6	6
P10	5	5	5	5	5	5	5	5	7	6	6	6	7	6	6	6
P11	2	2	2	2	4	3	3	3	5	5	4	5	6	6	5	6
P12	3	3	3	3	4	4	4	4	6	5	6	5	7	6	7	6
P13	2	2	3	2	4	3	3	3	5	4	4	4	6	5	5	4

Note. P for "participant", R for "rater."

Table 15. *Results of rating of Sentence 8*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	3	3	4	4	4	5	5	5	5	5	6	6	6	6
P2	2	2	3	2	4	3	3	3	5	4	4	5	6	5	5	6
P3	3	3	3	3	5	4	5	4	5	5	5	5	6	6	6	6
P4	3	4	4	3	5	5	5	5	6	6	6	6	6	6	7	6
P5	3	3	4	3	4	4	4	4	5	5	6	6	7	6	6	6
P6	4	3	3	4	5	4	4	5	6	5	5	6	7	6	6	6
P7	4	4	4	4	5	5	4	5	6	6	5	6	6	6	6	6
NF																
P8	3	3	4	4	4	4	4	4	5	6	5	6	6	6	6	6
P9	5	5	5	5	5	5	5	6	6	6	6	6	7	6	6	7
P10	5	5	5	5	5	5	5	5	7	6	6	6	7	6	6	6
P11	2	3	2	3	3	4	3	3	6	5	5	4	6	6	6	5
P12	2	3	3	3	4	4	4	4	5	5	6	5	7	6	6	6
P13	2	2	2	3	3	3	3	3	5	4	5	5	6	5	6	6

Note. P for "participant", R for "rater."

Table 16. *Results of rating of Minimal Pair 1*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	2	3	2	3	3	3	2	3	3	3	3	3	3	3	3	3
P2	2	3	3	3	3	3	3	3	3	3	4	3	4	3	4	3
P3	3	3	3	3	4	4	4	4	5	4	4	4	5	4	4	4
P4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5
P5	3	3	2	3	3	3	3	3	3	4	4	3	3	4	4	3
P6	4	4	4	4	4	4	4	4	4	4	5	4	5	4	5	4
P7	3	3	3	3	4	3	3	3	5	4	4	4	6	5	4	5
NF																
P8	3	3	3	3	3	3	3	3	4	3	4	3	4	3	4	3
P9	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
P10	5	4	4	4	5	5	4	4	5	5	4	5	5	5	4	5
P11	3	3	3	3	3	3	3	3	4	3	4	3	4	3	4	3
P12	3	3	3	3	4	3	3	3	4	3	4	4	4	3	4	4
P13	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Note. P for "participant", R for "rater."

Table 17. *Results of rating of Minimal Pair 2*

	PRE								POST							
	RD				LR				RD				LR			
	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4	R1	R2	R3	R4
LP																
P1	3	3	3	3	3	3	3	3	3	3	4	3	4	3	4	4
P2	2	2	2	2	3	3	2	3	3	3	2	3	3	3	2	3
P3	3	3	3	3	4	4	3	3	3	4	4	4	3	4	4	4
P4	3	3	3	2	3	3	3	3	4	3	4	3	4	4	4	3
P5	4	4	3	3	4	4	4	3	4	5	4	4	4	5	4	4
P6	3	3	3	3	3	3	4	3	3	3	4	3	3	3	4	3
P7	3	3	3	3	3	3	3	3	4	3	4	3	4	3	4	3
NF																
P8	3	3	3	3	3	3	3	3	4	3	3	4	4	3	3	4
P9	3	3	3	3	3	3	3	3	3	4	3	3	3	4	3	3
P10	5	4	4	4	5	4	4	4	5	5	4	5	5	5	4	5
P11	3	3	3	3	3	3	3	3	4	3	4	4	4	3	4	4
P12	3	3	3	3	3	3	3	3	4	3	4	4	4	3	4	4
P13	2	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3

Note. P for "participant", R for "rater."

Table 18. *The stimulus used in the 10 sessions of prosody training (study 3) 1*

(a) Session 1 Tones

No.	Sentence	
1	Great!	
2	Thanks!	
3	Pardon?	
4	Yes.	
5	Bass.	
6	A: I'll be there by five.	B: Great!
7	A: Care for a drink?	B: Thanks!
8	A: You'll have to take the tube.	B: Pardon?
9	A: You were there, weren't you?	B: Yes.
10	A: He sings tenor.	B: Bass.

(b) Session 2 Statements

No.	Sentence	
1	This is a pen.	
2	I think it's great.	
3	A: When'll they finish?	B: Next Wednesday.
4	I won't eat anything.	
5	I won't eat anything.	
6	Will you eat anything?	
7	Roads are rough in rural areas.	
8	It's not hot, it's cold.	
9	A: Who's that?	B: I know her face.
10	A: She's working in Oxford.	B: Cambridge.

Note. No. 4 and 5 are difference in tones.

Table 18. *The stimulus used in the 10 sessions of prosody training (study 3) 2*

(c) Session 3 Questions

No.	Sentence	
1	What's your name?	
2	What's your name?	
3	A: Come and see me tomorrow.	B: When?
4	Would you pass me the water?	
5	A: Would you like some tea?	
6		B: Sorry, what was that?
7	A: Would you like some tea?	
8	It's snowing, is it?	
9	Answer the phone, will you?	
10	A: It's a sort of green.	B: Blue, isn't it?

Note. No. 1 and 2 are difference in tones.

(d) Session 4 Other sentence types

No.	Sentence	
1	What a good idea!	
2	A: These are my two sons.	B: Haven't they grown!
3	Do be careful.	
4	Do be careful.	
5	What a pity you will miss the party!	
6	What do you think, okay?	
7	Good morning!	
8	Good morning!	
9	A: Here are your tickets.	B: Thank you.
10	A: Here are your tickets.	B: Thank you.

Note. No. 3 and 4, 7and 8, 9 and 10 are difference in tones.

(e) Session 5 Sequence of tones

No.	Sentence	
1	We're going to Spain in August.	
2	After lunch we could call on Mary.	
3	After lunch we could call on Mary.	
4	Please shut the window.	
5	Please shut the window.	
6	A: I've got some chocolate here.	
7		B: Oh dear. I like chocolate, but I'm on a diet.
8		B: Oh good. I like chocolate. Pass it over.
9	This is my colleague, Charles.	
10	This is my colleague, Charles.	

Note. No. 2 and 3, 4 and 5 are difference in tones.

(f) Session 6 Review

No.	Sentence	
1	A: I'd like some chicken, please.	B: Fried?
2	A: Are you sure?	B: I think so.
3	Guns don't kill people.	
4	People kill people.	
5	You'll be coming to dinner?	
6	How do you spell friend?	
7	Isn't she pretty?	
8	Never mind.	
9	If I were you, I'd reject it.	
10	I'd reject it if I were you.	

Table 18. *The stimulus used in the 10 sessions of prosody training (study 3) 3*

(g) Session 7 Tonicity

No.	Sentence	
1	the BBC	
2	She's done it.	
3	Tell me about it.	
4	I think you ought to tell me about it.	
5	newspaper	
6	peace keepers	
7	Don't look at the keyboard.	
8	I've lost my credit cards.	
9	A: What's the problem?	B: Let's look at the operating system.
10	A: How can I write it?	B: Use a word processor.

(h) Session 8 The old/new focus

No.	Sentence	
1	A: Yes madam?	B: I'd like a gin and tonic.
2	A: How about a gin and tonic?	B: Oh, I'd prefer a vodka and tonic.
3	A: I'd like to speak to the manager.	B: She's much too busy.
4	A: I'd like to speak to the manager.	B: She's much too busy.
5	We enjoy swimming, hiking, and tennis.	
6	A: What happened next?	B: Everyone burst out laughing.
7	A: Who brought the wine?	B: Mary.
8	A: What did Mary bring?	B: The wine.
9	A: What do you like about me?	B: Well I love your hair.
10	A: But darling, don't you like my hair?	B: I love your hair.

Note. No. 3 and 4 are difference in tones.

(i) Session 9 Accenting old material/what is known

No.	Sentence	
1	A: I can't stand whisky.	B: You can't stand whisky?
2	A: You say your name's Smith?	B: Yes, Smith.
3	Surprise, surprise!	
4	The noise got louder and louder.	
5	I can't accept it, I just can't accept it.	
6	The next station is Oval.	
7	This station is Oval.	
8	I like the cooler climate in the fall .	
9	North Korea faces severe food shortages.	
10	A: Pork or beef?	B: Sorry, I don't eat meat.

(j) Session 10 Tonality

No.	Sentence
1	Milk comes from cows, and wool comes from sheep.
2	Milk, I believe, comes from cows.
3	I'll talk to the students in the garden.
4	I'll talk to the students in the garden.
5	He was looking up the street.
6	He was looking up the street.
7	He was looking up the street.
8	Linda, could I have a word?
9	You'll remember that, won't you?
10	You, Jimmy, have made a mistake.

Note. No. 5, 6, and 7 are difference in tones.

Table 19. *The mean average of durational ratios*

(a) group A

Participant No.	PRE	MID	POST
P1	1.10	0.98	0.97
P2	1.05	1.02	1.03
P3	1.10	1.20	1.13
P4	1.04	1.07	1.01
P5	1.11	1.10	1.09
P6	1.25	1.23	1.22
P7	1.01	0.96	0.99
P8	1.18	1.16	1.26
P9	1.15	1.23	1.18
P10	0.98	1.00	0.97
P11	1.18	1.13	1.11
P12	1.15	1.14	1.06
P13	1.35	1.21	1.19
P14	1.04	1.01	1.01
P15	1.03	1.02	1.05
P16	1.10	1.13	1.04
P17	1.15	1.19	1.16
P18	1.20	1.20	1.22
P19	1.14	0.98	1.13
P20	1.08	1.17	1.19

(b) group B

Participant No.	PRE	MID	POST
P21	1.11	1.11	1.07
P22	1.12	1.05	1.05
P23	1.09	0.99	0.99
P24	1.19	1.17	1.14
P25	1.06	1.02	1.01
P26	1.14	1.02	1.07
P27	1.06	1.02	0.97
P28	1.14	1.11	1.06
P29	1.23	1.28	1.18
P30	1.08	1.03	0.99
P31	1.09	1.02	1.00
P32	1.19	1.13	1.05
P33	1.18	1.13	1.08
P34	1.10	1.07	1.07
P35	1.10	1.09	1.04
P36	1.05	1.01	1.00
P37	1.07	1.04	1.03
P38	1.25	1.29	1.26
P39	1.02	0.98	1.00
P40	1.13	1.06	1.01

(c) group C

Participant No.	PRE	MID	POST
P41	1.03	1.03	0.99
P42	1.06	1.03	1.01
P43	1.09	1.06	1.11
P44	1.12	1.11	1.13
P45	1.10	1.05	1.03
P46	1.19	1.17	1.15
P47	1.11	1.10	1.11
P48	1.21	1.11	1.14
P49	1.06	1.13	1.12
P50	1.05	1.06	1.07
P51	1.05	1.17	1.10
P52	1.06	1.17	1.04
P53	1.28	1.13	1.10
P54	1.14	1.13	1.18
P55	1.00	1.09	1.04
P56	1.13	1.13	1.08
P86	1.20	1.17	1.17
P58	1.08	1.05	1.04
P59	0.94	0.87	0.91
P60	1.02	0.98	0.98

(d) group D

Participant No.	PRE	MID	POST
P61	1.11	1.10	1.09
P62	1.08	1.05	1.04
P63	1.06	1.04	1.04
P64	1.09	1.08	1.07
P65	1.09	1.09	1.04
P66	1.06	1.05	1.07
P67	1.09	1.10	1.08
P68	1.12	1.09	1.08
P69	1.08	1.08	1.08
P70	1.09	1.08	1.09
P71	1.09	1.08	1.08
P72	1.12	1.09	1.09
P73	1.12	1.12	1.11
P74	1.10	1.08	1.04
P75	1.09	1.11	1.08
P76	1.09	1.07	1.05
P77	1.08	1.10	1.08
P78	1.07	1.02	1.02
P79	1.08	1.04	1.04
P80	1.11	1.07	1.04

Table 20. *The mean average of F0 range ratios*

(a) group A

Participant No.	PRE	MID	POST
P1	0.71	0.84	1.06
P2	0.68	0.94	1.20
P3	0.58	0.79	0.96
P4	0.63	0.81	0.99
P5	0.53	0.80	0.99
P6	0.69	0.83	1.00
P7	0.55	0.72	0.90
P8	0.66	0.73	1.02
P9	0.50	0.73	0.91
P10	0.62	0.73	1.01
P11	0.73	0.82	1.10
P12	0.56	0.84	0.99
P13	0.65	0.78	0.93
P14	0.58	0.70	1.02
P15	0.61	0.75	0.91
P16	0.62	0.73	0.96
P17	0.56	0.87	0.94
P18	0.57	0.84	1.05
P19	0.56	0.82	0.99
P20	0.64	0.82	1.06

(b) group B

Participant No.	PRE	MID	POST
P21	0.53	0.77	0.97
P22	0.64	0.75	0.98
P23	0.60	0.73	0.90
P24	0.54	0.72	0.96
P25	0.64	0.77	0.93
P26	0.62	0.77	0.98
P27	0.56	0.85	0.94
P28	0.64	0.78	1.06
P29	0.76	0.79	0.96
P30	0.57	0.74	0.98
P31	0.83	0.94	1.00
P32	0.86	0.99	1.09
P33	0.88	0.93	1.00
P34	0.84	0.96	1.05
P35	0.66	0.80	0.97
P36	0.75	0.89	1.01
P37	0.66	0.88	1.04
P38	0.66	0.88	0.94
P39	0.58	0.80	0.89
P40	0.65	0.85	0.96

(c) group C

Participant No.	PRE	MID	POST
P41	0.73	0.87	1.07
P42	0.73	0.88	1.05
P43	0.82	0.89	1.17
P44	0.91	1.04	1.17
P45	0.73	0.84	1.18
P46	0.64	0.82	0.97
P47	0.72	0.85	1.05
P48	0.73	0.87	1.00
P49	0.80	0.83	1.14
P50	0.82	0.90	1.09
P51	0.68	0.91	1.02
P52	0.63	0.83	7.07
P53	0.80	0.91	1.05
P54	0.69	0.80	1.12
P55	0.76	0.97	1.06
P56	0.74	1.00	1.05
P86	0.74	0.87	0.97
P58	0.82	0.89	9.99
P59	0.70	0.90	1.16
P60	0.74	0.86	1.04

(d) group D

Participant No.	PRE	MID	POST
P61	0.82	0.99	0.97
P62	0.81	0.84	0.93
P63	1.02	0.95	0.89
P64	0.89	0.95	0.92
P65	1.26	0.86	0.89
P66	0.88	0.88	0.87
P67	0.84	0.89	0.90
P68	0.86	0.76	0.92
P69	0.87	0.91	0.84
P70	0.75	0.98	0.84
P71	0.90	0.88	1.01
P72	0.84	0.89	0.86
P73	0.88	0.88	0.86
P74	0.88	0.87	0.86
P75	0.82	0.88	0.94
P76	0.81	0.81	0.98
P77	0.85	0.87	0.83
P78	0.83	0.80	0.90
P79	0.91	1.02	1.01
P80	0.82	0.87	1.01

Table 21. *Mean subject evaluations of four raters*

(a) group A

Participant No.	PRE	MID	POST
P1	3.44	3.99	4.79
P2	3.37	3.84	4.71
P3	3.53	3.88	4.57
P4	3.35	4.18	5.19
P5	3.28	3.91	4.62
P6	3.16	3.69	4.66
P7	3.31	3.87	4.50
P8	3.15	3.59	4.40
P9	3.40	3.62	5.06
P10	3.31	3.90	4.96
P11	3.19	4.00	4.71
P12	3.40	3.93	4.69
P13	3.16	3.78	4.53
P14	3.41	3.72	4.56
P15	3.56	3.49	4.38
P16	3.37	3.69	4.35
P17	3.21	3.37	4.12
P18	3.25	3.47	4.13
P19	2.87	3.94	5.12
P20	3.10	3.43	4.49

(b) group B

Participant No.	PRE	MID	POST
P21	3.22	4.25	5.37
P22	3.29	4.13	5.09
P23	3.53	4.34	5.63
P24	3.79	4.21	5.51
P25	3.13	4.12	5.47
P26	2.94	4.13	5.32
P27	3.49	4.19	5.49
P28	3.75	4.41	5.49
P29	3.44	4.13	5.31
P30	3.29	4.63	5.32
P31	3.29	4.13	5.24
P32	3.31	4.12	5.09
P33	3.25	4.07	5.29
P34	3.21	3.88	5.47
P35	3.09	4.13	5.26
P36	3.49	4.07	5.06
P37	3.13	4.72	5.12
P38	3.41	4.25	5.34
P39	2.93	4.24	5.13
P40	3.00	4.03	5.18

(c) group C

Participant No.	PRE	MID	POST
P41	3.29	3.87	4.59
P42	3.54	3.66	4.35
P43	3.35	3.65	4.51
P44	3.29	3.68	4.69
P45	3.60	3.62	4.47
P46	3.12	3.72	4.96
P47	3.29	3.74	4.57
P48	2.99	3.57	4.35
P49	3.41	3.62	4.54
P50	3.25	3.56	4.26
P51	3.41	3.56	4.88
P52	3.38	3.94	4.79
P53	3.22	3.63	4.75
P54	3.24	3.91	4.75
P55	3.32	3.82	4.54
P56	3.40	3.79	4.78
P86	3.15	3.38	4.74
P58	3.44	3.62	4.44
P59	3.10	3.69	4.93
P60	3.32	3.88	4.66

(d) group D

Participant No.	PRE	MID	POST
P61	3.50	3.50	3.69
P62	4.06	4.35	4.41
P63	3.50	3.56	3.68
P64	3.47	3.56	3.47
P65	3.53	3.68	3.72
P66	3.46	3.53	3.40
P67	3.43	3.47	3.54
P68	3.49	3.56	3.63
P69	3.54	3.63	3.72
P70	3.66	3.51	3.13
P71	3.82	3.76	4.06
P72	3.53	3.68	3.40
P73	3.34	3.84	3.65
P74	3.37	3.69	3.82
P75	3.41	3.29	3.35
P76	3.44	3.12	3.25
P77	3.16	3.18	3.18
P78	3.16	3.13	3.32
P79	3.24	3.32	3.46
P80	3.00	3.13	3.25

伊庭 緑（いば みどり）

津田塾大学学芸学部英文学科卒業。神戸大学大学院教育学研究科修士課程修了。神戸大学大学院国際文化学研究科博士後期課程修了。学術博士（神戸大学）。2002 年、ロンドン大学客員研究員。現在甲南大学国際言語文化センター准教授。

Midori Iba is an Associate Professor at the Institute for Language and Culture, Konan University in Kobe, Japan. She was an Affiliate Academic at the Department of Phonetics and Linguistics, UCL in 2002. She completed her Ph.D. at Kobe University in 2010. She loves cats.

Experimental Studies on Prosodic Features in Second Language Acquisition:

Training Japanese learners to produce natural English

2010年8月26日　初版発行　　定価（本体4,600円＋税）

著　者	伊庭緑
発行者	三浦衛
発行所	春風社
	横浜市西区紅葉ヶ丘53　横浜市教育会館3階
	電話　045-261-3168
	FAX　045-261-3169
	http://www.shumpu.com
	info@shumpu.com
	振替　00200-1-37524
装　丁	後藤葉子
印刷・製本	シナノ書籍印刷株式会社

© Midori Iba 2010. ISBN 978-4-86110-235-6 C3082 ¥ 4600E