

Studies of Improvement in Pronunciation of Japanese EFL
Learners by Using Multimedia Applications

.....

A Dissertation Submitted to
The Graduate School of Foreign Language Education and Research,
Kansai University

.....

In Partial Fulfillment of the Requirements for the Degree
Doctor of Philosophy in Foreign Language Education and Research

.....

by

SONOBE, Hideyuki

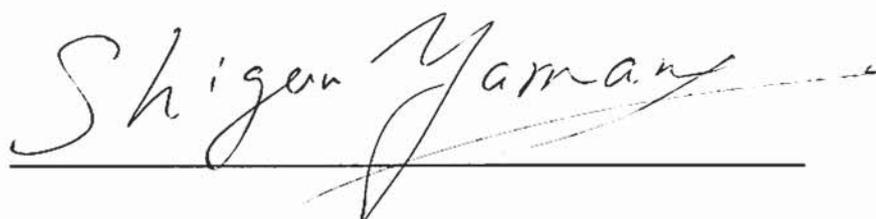
March 31, 2010

© Copyright by SONOBE, Hideyuki, 2010

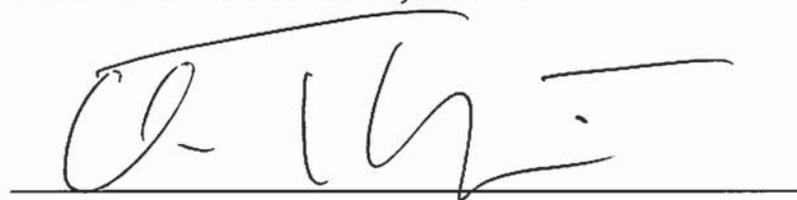
This dissertation of SONOBE, Hideyuki is approved.

Doctoral Committee:

Professor S. Yamane, Ph. D. (Chair)

A handwritten signature in cursive script, reading "Shigen Yamane", written above a horizontal line.

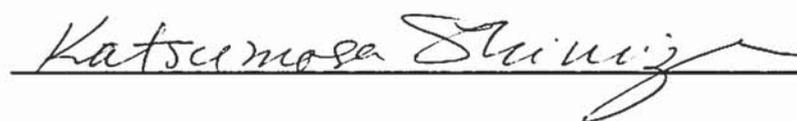
Professor O. Takeuchi, Ph. D.

A handwritten signature in cursive script, reading "O. Takeuchi", written above a horizontal line.

Professor S. Yoshida, Ph. D.

A handwritten signature in cursive script, reading "S. Yoshida", written above a horizontal line.

Professor K. Shimizu, Ph. D. (External)

A handwritten signature in cursive script, reading "Katsunasa Shimizu", written above a horizontal line.

The Graduate School of Foreign Language Education and Research,
Kansai University, Osaka, Japan
2010

論文要旨（概要）

本論文は日本人 EFL 学習者の口語英語の発音について、英語教育の立場から教材を作成、提案し、その音韻面での特徴から発音の進歩を検証した4つの実証的研究の結果をまとめたものである。

第一章では本研究の位置づけ及び本研究にいたった経緯と本論文の目的について概説する。本研究の位置づけ及び本研究にいたった経緯として日本人 EFL 学習者の口語英語の低い発音評価の要因となる発音指導上の取り組み方、英語教師による指導の現状、発音指導教材の問題点を紹介する。そして本研究の目的として英語の分節素と同様に口語英語の発音上重要であると言われるプロソディーの各特徴を改善することが発音評価者による評価や数量的評価を向上させることに繋がる事を明確にしようとした。そしてそのために超分節素の各特徴を改善するための教材を作成、検証し、今後の発音指導に寄与することを目的とした。

第二章では先行研究から日本人 EFL 学習者の口語英語の韻律面での主要素である Rhythm, Duration, Pitch, Intonation について英語母語話者と比較しながらその特徴について述べる。同時に音声学習を促進する働きのある Visual Information のメカニズムについて説明する。またこの章では口語英語を評価する上で基準となる Intelligibility (明瞭性), Comprehensibility (理解しやすさ), Naturalness as Spoken English (英語の自然らしさ) の定義を試みた。本研究では英語の韻律面での評価を行うことを目的とするので英語の自然らしさを全体的音声評価の基準とした。

第三章では第二章で述べた理論に基づき日本人 EFL 学習者 21 名に対する実験 1 を実施した。実験 1 では録音された口語英語の音声を英語話者と比較した上で、その発話時間長と発音の自然さ(Naturalness)との相関を検証した。発話

時間長は(1) 全文、(2) 各語彙、(3) 強勢グループ、(4) 弱形の発音語彙を測定し発音の「英語としての自然さ」との相関を検証した。

第四章では第三章の実験 1 にて検証された発話時間長についてさらに検証し、実験 2 にてモーラ言語である日本語を母語とする英語学習者に発話時間長の制御を改善すると期待される文字の縦横の動きを取り入れ、視覚に訴える文字アニメ教材を作成し実験を実施した。実験はアニメ教材を使用した実験群、静止画を使用した統制群に分け、CALL 教室にて実施した。文字アニメ教材は英語の自然な強勢型リズムである発話時間長制御を可能とする教材として独自に作成した。口語英語の評価は 2 名の英語母語話者及び 1 名の日本人英語教師による発音評価と音声分析ソフトの *WaveSurfer* による数量的評価で実施した。発話時間長によるクラスター分析結果では、英語母語話者に近い各句の時間長を持つグループが一番高い評価結果となった。また、学習者の認知的な働きを検証するための発音面及び動機づけに関するアンケートでは共に高い好意的な感想を得た。

第五章では第四章の実験 2 の結果から次の点を改良および機能を追加した教材を作成した。(1) 文字の色、(2) 文字の伸縮、(3) 練習回数表示、(4) 強勢音への効果音、(5) 英語話者の顔の Video の導入、(6) 音声認識装置、そして独自の 3 段階学習システム (3-Step Learning System) による学習法にて実施した。改良した文字アニメ教材はウェブ上に実験群用として設定し、統制群用には音声+静止文字の条件で実験 3 を実施した。実験 3 の数量的検証では音声分析ソフト *Praat* を使用し、練習文及び練習文以外の文への発音力向上の転移も同時に検証した。その結果、3 名の評価者よる 10 段階評価では両方の英文において有意差が見られ、ピッチ幅も広がりを示し、韻律面での向上が見受けられた。特にピッチ、イントネーションの変化を視覚的に伝える情報が英語学習者の音声情報を促進させ、認知的に「気づき」が起こり韻律面での向上につながったと考えられる。実験後のアンケートではリズムやイントネーションの習

得などに関するスキル面、音声学習の楽しさや練習の動機づけなどに関する情意面ともに好意的な評価が得られた。

第六章ではこれまで実施してきた実験 1 から 3 の集大成として日本人 EFL 学習者に更に効果的と思われる教材を開発し実験 4 に臨んだ。教材は口語コミュニケーションにおいて最も重要な役割を果たすと言われるイントネーションの向上に焦点を置いたもので、学習者がモデルとなる英語母語話者のピッチ曲線に近づけることを目標とした発話練習用教材である。実験 3 では文字アニメーション教材を使用することで強音節のピッチ幅を拓げることに成功し、それが有意な発音評価に結びついたと考えられるが、文全体の流れであるイントネーションを視覚情報として学習者に提示することはより効果的に韻律面での向上を促進するであろうと考え、オリジナルソフトを作成した。実験はオフラインでも使用できる環境で中学生、大学生を参加者として検証した。発音評価は英語母語話者 3 名、日本人英語教師 3 名の計 6 名で実施し、*t* 検定による分析結果では中学生、大学生ともに統計上有意差があった。また練習後はセミトーンによるイントネーションやピッチ幅においてもより英語母語話者のイントネーションに近づく結果となった。この結果からイントネーションを改善することは韻律面での発音向上に大変重要な役割を果たすと言える。そして練習後のアンケートではオリジナルソフトの使い易さ、学習意欲、イントネーション学習の成果においていずれも高い好意的な感想を得た。

第七章ではコンピュータ技術の発達にともない変容する学習環境の中で日本人 EFL 学習者のための英語音声学習について第二章から第六章までの研究に共通した問題点や限界点をふまえ、以上 4 つの実験から得られた結果から次の 4 点の研究課題について見直し、今後の研究の方向性を示した。(1) 日本人英語学習者のプロソディーの特徴について、(2) 本オリジナルソフトは発音学習に役立ったか、(3) もし役立ったとすると特にどの点が向上したか、(4) どの点での韻律上での改善が評価に繋がったのか。そして本章では音声情報の理解

促進のため、韻律情報を視覚に訴えるためのオリジナルマルチメディア教材を作成し、メタ認知による気づきを促し、自律学習者へと導くための理論的背景を述べる。具体的には発音学習の効果を上げるため、モデル音声と比較したピッチ曲線を視覚情報として学習者に提供することにより学習者に気づきや自己訂正を起こさせる活動をすることが発音の向上を促し、それが自動化に繋がるメタ認知の働きについて教育的示唆を行った。

最後にこれまでの実験結果から発音において学習効果が認められたオリジナルマルチメディア教材を **Blended Learning** に活用することを提案した。**Blended Learning** とはこれまでの伝統的な対面式の授業と時間や場所に制約されないウェブ教材による e ラーニングを組み合わせた学習法であり、これまでの実験から発音学習において有効であることが証明された。**Blended Learning** は(1)多様化する学習者要因に対応でき、(2)音声や視覚情報を有効に活用するマルチメディアの特性を生かすことが可能であり、(3)学習動機づけを高めるインターラクティブな学習法であると考えられる。

本研究にて紹介したオリジナルのマルチメディア教材が英語発音学習において多くの先生方に活用されることを切に希望する。

Acknowledgements

I have been fortunate to have had excellent supervisors who were always there to offer valuable advice while working on this dissertation.

First, I wish to express my deepest gratitude to Professor Shigeru Yamane, who has been my supervisor for the second, third and fourth years of my graduate course at Kansai University. Dr. Yamane has patiently guided me so that I was able to become a self-supporting instructor and researcher. Without his constant support and proper guidance, I would never have finished writing this dissertation. He helped nurture my ability to think logically while organizing this dissertation. One of my fondest memories is when I presented a paper at an international conference together with Professor Yamane, who provided me with a valuable insight into the effective ways of presenting a paper in English. This experience served as a great stimulus and encouraged me to explore my research further. I would also like to express my sincere gratitude to Professor Osamu Takeuchi, who was my tutor for the first year and taught me that “Teaching English is a science and everything related to teaching English should be proved.”

I am also very grateful to Professor Makoto Ueda of Aichi Sangyo University for his technical assistance and support with creating the original multimedia software. Professor Fumimasa Ishigami and Instructor Okamoto of Aichi Sangyo Mikawa junior high school are also subjects of my gratitude, as it was they who willingly accepted my request to have their students participate in the experiments. I would like thank Professor Katsumasa Shimizu, a renowned scholar in phonetics and phonology, who gave me warm encouragement. Furthermore, I would like to give special thanks to Lecturer Douglas McDowall, who always cooperated with my requests to make sound files. I feel privileged to have participated in Dr. Yamane’s seminar, whose members

were all dedicated to their research and inspired me a great deal. I feel grateful for the precious comments and suggestions given by Professor Kathleen Yamane of St. Thomas University at the early stages of this dissertation. Mr. Tom Mayes of the British Council was kind enough to check my English and demonstrated his excellent ability as a proofreader.

Finally, I dedicate this achievement to my family, especially to my wife, who has supported me and taken care of my parents. To them I express my gratitude and love from the bottom of my heart.

Table of Contents

Abstract	i
Acknowledgments	v
Table of Contents	vii
1. Introduction	1
1.1 What Motivated the Author to Write This Dissertation?	1
1.1.1 World Englishes	1
1.1.2 Japanese English Teachers	2
1.1.3 Amount of Lessons Allocated to Pronunciation Instruction	3
1.1.4 Problems in Teaching Materials	4
1.2 Research Questions	6
1.3 The Purpose of This Dissertation	7
2. Literature Review	11
2.1 Prosodic Features of English Produced by Japanese EFL Learners	11
2.2.1 Rhythm	12
2.2.2 Duration	15
2.2.3 Pitch	18
2.2.4 Intonation	20
2.3 Visual Information	22
2.4 Intelligibility, Comprehensibility and Perceived Naturalness as Spoken English	24

3. The Relationship Between Duration and Perceived Naturalness	27
3.1 Purpose	27
3.2 Participants	27
3.3 Experiment 1	28
3.4 Results and Discussion	30
3.4.1 Evaluation by Raters.....	30
3.4.2 Objective Measurements.....	31
3.5 Conclusion	36
Notes	38
4. The Effects of Pronunciation Practice with Animated Materials Focusing on Duration	39
4.1 Introduction	39
4.2 Importance of Speech Duration in English Prosody	40
4.3 Creating Multimedia Applications Focusing on Duration	41
4.4 Experiment 2	43
4.4.1 Participants	43
4.4.2 Procedure	44
4.5 Results and Discussion	46
4.5.1 Evaluation by Raters	46
4.5.2 Acoustic Measurements	47
4.5.2.1 Comparison in the Segment Duration Between NSs and NNSs ...	49
4.5.2.2 A Cluster Analysis	51
4.6 Questionnaires	52
4.7 Conclusion and Implication	53

Note	55
5. The Effects of Pronunciation Practice with Animated Materials Focusing	
on English Prosody	56
5.1 Introduction	56
5.2 The Present Study	58
5.2.1 Purposes	58
5.2.2 Creating Two Different Types of Web Materials	58
5.2.3 Three-Step Learning System	60
5.3 Experiment 3	62
5.3.1 Basic Data of the Two Groups	62
5.3.2 Procedures	63
5.4 Results and Discussion	64
5.4.1 Evaluation by Raters	64
5.4.2 Objective Measurements	66
5.4.3 Correlations Between Rater's Evaluation and Objective Measurements	70
5.5 Questionnaires	71
5.6 Conclusion and Implication	72
6. A Study of Intonation Practice Using Original Software: Improvement in	
Pronunciation	75
6.1 Introduction	75
6.2 Difficulties that Japanese EFL Learners may Face with Intonation Practice	77
6.2.1 Perception of Sound	77
6.2.2 Tonetic Notations	77

6.2.3 Lack of Proper Teaching Materials	79
6.3 Purpose	79
6.3.1 Previous Studies	79
6.4 Procedure	80
6.4.1 Creating Original Computer Software Application Focusing on English Intonation	80
6.4.2 Some Features of <i>IntoRec</i>	81
6.4.3 Participants	81
6.4.4 Practice Procedures	82
6.5 Experiment 4	84
6.6 Results and Discussion	87
6.6.1 Evaluation by Raters	87
6.6.2 Acoustic Analyses	89
6.6.3 Correlations Between Raters' Evaluations and Other Parameters	94
6.7 Questionnaires	96
6.8 Conclusion and Implications	98
7. Conclusion and Implications	101
7.1 Research Questions	101
7.2 Pedagogical Implications	106
7.2.1 Importance of Visual Information for Pronunciation Training	106
7.2.2 Metacognition	107
7.2.3 Blended Learning	108
Note	112

References	113
Appendices	119
Appendix A: Questionnaire: Pronunciation Skills (Chapter 4)	119
Appendix B: Questionnaire: On Affective Factors (Chapter 4)	120
Appendix C: Sentences for Rhythm Practice (Chapter 4)	120
Appendix D: Questionnaire: On Communicative Skills (Chapter 5)	121
Appendix E: Questionnaire: On Affective Factors (Chapter 5)	122
Appendix F: Sentences for Rhythmic Practice (Chapter 5)	123
Appendix G: Practice Sentences (Chapter 6)	123
Appendix H: Questionnaires (Chapter 6)	124
Appendix I: List of the Original Software Developed for Each Experiment ·	125

1. Introduction

1.1 What Motivated the Author to Write This Dissertation?

First of all, the author would like to mention the reasons why he came to write this dissertation. Jenkins (2007), investigating the intelligibility of English spoken by non-native speakers, reported that when the natives of seven different non-English speaking countries, including Japan, were asked to rate the English of EFL speakers in other countries, the English of the Japanese EFL learners was ranked the lowest. It was also ranked the second from bottom, followed only by Chinese English, among the same seven countries when assessed by native speakers of English. These results imply that English spoken by the Japanese is not comfortably perceived or understood by native speakers of English, as well as by non-native speakers. In order to promote smooth communication with English speakers of the world, it is necessary for Japanese EFL learners to improve English pronunciation to the extent whereby it does not hinder comfortable communication. If we take the natural process of language acquisition into consideration, it is obvious that spoken English should be taught prior to written English, since oral communication plays a vital role in human communication in general.

Discussed in the following section are four conceivable problems which Japanese EFL learners as well as teachers often encounter in learning English pronunciation.

1.1.1 World Englishes

English is currently regarded as the world's main international language and used as a means for mutual communication not only between native speakers (NSs) of English and non-native speakers (NNs), but also among NNs. Recently, the idea of "World Englishes" (WE), as supported by Kachru (1986) and Jenkins (2002), has come

to be widely approved. Proponents of WE claim that as long as the meaning of the spoken message is clear enough to be understood by the addressee, a variety of English pronunciation phonetically modified with different regional accents should be accepted. If this is the case, which variety of English should be chosen as the model of pronunciation Japanese EFL learners should aim at? The author considers that those varieties which are readily accessible for Japanese EFL learners should be selected. Received Pronunciation (RP), the standard British accent and GA (General American), to name a few, are considered to be possible models of English pronunciation.

It will make it easier for learners to practice their English communication if a pronunciation model is provided for them since they can be sure of the learning target, just like beginner golfers need to watch closely the swing forms of pro-golfers. Mastering a RP or a GA accent should not necessarily be the learning target, but to be able to speak intelligible, comprehensible, and natural English should be their goal. The author will define the adjectives “intelligible,” “comprehensible,” and “natural” English in Chapter 2.

1.1.2 Japanese Teachers of English

English teachers are not necessarily keen on teaching pronunciation to students, and the students, on the other hand, are often unwilling to engage in pronunciation practice. One of the reasons for this lack of eagerness on both sides seems to stem from the fact that improvement in pronunciation is not readily visible and it is difficult to quantify the degree of improvement with measurable scores. English teachers are not usually equipped with enough knowledge to teach pronunciation to Japanese EFL learners and they are not sure about how much of it they are expected to teach.

Some English teachers, who have not received any pronunciation training in university and do not know basic theory in phonetics, are sometimes not sure what to

teach and how to teach pronunciation (Arimoto & Kochiyama, 2006). Though many of them are aware of the importance of pronunciation teaching in Japan for efficient oral communication, it is not generally emphasized enough in actual teaching scenes. Both senior and junior high school teachers, in general, are enthusiastic about teaching English grammar, but they are not necessarily ardent instructors of pronunciation.

The author developed multi-media applications for the purpose of assisting both those instructors who are not knowledgeable enough to teach pronunciation with confidence and the learners who need to improve their English pronunciation. These teaching materials, available on the Internet, will spare teachers a lot of the time that is normally necessary for pronunciation instruction. If a CALL system is available, these multimedia applications can be used in a classroom. If not, students can practice at home at their own pace with their PCs logged in to the Internet. These teaching materials are designed to appeal to the learners' interest in pronunciation, thus enhancing their motivation for pronunciation learning. It is the author's sincere hope that many students access this site and enjoy learning English pronunciation online.

1.1.3 Amount of Lessons Allocated to Pronunciation Instruction

The author feels that not enough time is given to pronunciation learning in junior and senior high schools in Japan. Since the students' main concern in learning English is to pass entrance examinations in many cases, teachers inevitably devote much of their lesson time to these preparations, leaving pronunciation instruction as a second or third priority. Though the Ministry of Education, Culture, Sports, Science and Technology initiated listening tests in 2006 in its National Center Test for University Admissions, a large proportion of the entrance examination was still occupied with questions on grammar, reading comprehension and English composition. "Questions related to pronunciation, accent and intonation make up only 8 percent in English

examination administered by the National Center for University Entrance Examination in 2008” (Sundai & Benesse, 2008).

What we should watch closely is which language skills are focused upon in university entrance examinations. The balance of instruction for preparation towards entrance examinations is usually decided depending on what sort of questions will be asked in the exams. In order to raise the students’ scores in exams, teachers tend to adapt the content of lessons to the types of questions asked in the exam. The author, who felt a lack of pronunciation instruction in Japan, created multimedia teaching materials geared toward pronunciation instruction and put them online. These materials are designed to free both instructors and learners from place and time limitations when learning and to encourage the learners to take more responsibility for their own learning. To put it differently, the materials aim at fostering autonomous learners of pronunciation. No matter how hard it is to find time for instructors to teach pronunciation in class, these teaching materials on the Internet will enable learners to practice pronunciation outside of the classroom whenever they wish.

1.1.4 Problems in Teaching Materials

In Japan, textbooks on English pronunciation dealing with segmental sound such as vowels and consonants are widely available these days. However, there are only a few good textbooks focusing on suprasegmental aspects of pronunciation, i.e., English rhythm and intonation which are considered to play a crucial role in communication. It is needless to say that there is an inevitable limitation to paper-based teaching materials as far as pronunciation learning is concerned. Practical and reliable devices using multimedia applications which provide learners with audio and visual information are necessary to enhance their motivation for pronunciation improvement.

It is close to impossible to teach pronunciation to a large number of students at

the same time, even for instructors with the necessary knowledge on phonetics and experiences of teaching pronunciation, since “pronunciation instruction is implemented in the most effective and efficient manner as a face-to-face activity” (Arimoto, 2002). We know from experience that it is difficult to teach pronunciation even to 20 students at a time. “It would be nearly impossible for an instructor to check the outward shape of the students' mouths during lessons and let the students imitate to produce the similar sound as the instructor” (Richards, 2002, p.27).

Pronunciation instruction systems, taking advantage of recent speech recognition technology, have already been developed by major companies which specialize in e-learning system design. To the best of the author's knowledge, however, few systems incorporated as a part of CALL console are capable of providing reliable feedback on pronunciation to users. Without prompt and proper feedback, students' motivation will be diminished. “Interactiveness” can be one of the most important keywords in this type of practice. Computer applications, used as stand-alone programs on PCs, have also been introduced by brand-name software companies in Japan. They were designed to evaluate learners' pronunciation. However, experimental data as to the effects on learners' pronunciation has not been made public. These applications available on a commercial basis focus too much on the evaluation of the segmental errors committed by Japanese learners, for instance, on the confusion between /l/ and /r/, /s/ or /ʃ/, and so on. Though prosodic features such as pitch and durations are displayed on the monitor, evaluation scores made on a prosodic aspect are not given to learners as feedback.

The multimedia teaching materials the author created are designed so that learners can establish their learning target easily and on an individual basis. While using these materials, learners are able to literally *see* their gradual improvement and feel their confidence in pronunciation. Their motivation will be enhanced by

effectively taking advantage of audio and visual information provided from the software. Setting a feasible and visible target is quite important to make learners continue the practice.

It is easy to speculate that pronunciation practice is likely to be a monotonous task and requires a lot of patience if learners practice by themselves. These teaching materials can also be used as a part of regular classroom activities with a teacher's computer at the beginning of each lesson, for instance. Practicing every day, even for a short period of time, will be more effective for learners to acquire better pronunciation than doing hour-long practice sessions. Thus, "easy access" is another keyword for learning online.

1.2 Research Questions

Evaluating Japanese learners' pronunciation is not an easy task for Japanese English teachers. Learners' skills in grammar, listening, and reading could be gauged by giving them proficiency tests, while the level of English pronunciation is very hard to be assessed since numerous variables could affect the judgment of spoken English. For instance, the rating will be influenced by the articulateness of segmental pronunciation, speech rate, and appropriate use of suprasegmentals such as rhythm and intonation. When the learner's problem lies in their segmentals, their speech will be judged fairly objectively as problems in manners of articulation. For instance, if an 'r' sound is not correctly pronounced, it is quite noticeable even to an inexperienced rater. In teaching pronunciation, more attention tends to be directed to segments (vowels and consonants) than to suprasegmentals (stress, rhythm, intonation, and speech duration). "Errors in intonation may be more serious since they can produce misunderstandings at the pragmatic level when the specific context may not help to disambiguate the intended meaning" (Cenoz & Lecumberri, 1999, p.4).

Results in earlier research suggest that in conversation among non-native English speakers, the correct pronunciation of segmentals becomes more important than the appropriateness of prosody, since clear articulation of segmentals facilitates understanding among NNSs (Arimoto, 2002; Jenkins, 2000; Yamane, 1999). However, even though learners may have problems with their prosody, communicative factors will make fair and objective assessments of their speech very difficult. In other words, raters sometimes place greater emphasis on success in communication rather than on appropriate use of prosody. Even when learners had some problems in prosodies, their intended messages can be conveyed. This is why Japanese English teachers find it difficult to teach prosody, i.e., the teachability of prosodies becomes in effect low, while learners on their part may find setting learning goals difficult.

Reflecting upon these problems which Japanese EFL learners face in learning English pronunciation, the author would like to put forward the following four main research questions: (1) what are the typical prosodic features displayed by Japanese EFL learners?; (2) will the original multi-media application help improve learners' pronunciation?; (3) should it be improved, and in what aspects of pronunciation in particular has it been improved?; and (4) Improvement in which prosodic aspects influence the evaluation of Japanese learners' speech as assessed by raters?

1.3 The Purpose of This Dissertation

The purpose of this dissertation is to investigate the improvement in pronunciation of Japanese EFL learners' by having them use the multimedia applications. Let the author describe the purpose of each chapter in detail.

The basic aim of Chapter 2 is to introduce prosodic features of English spoken by Japanese EFL learners and to attempt to examine those features: rhythm, duration, pitch, and intonation typically displayed by Japanese EFL learners when they speak

English. Each prosodic feature and its function will be described referring to previous studies. It is known that these features play an important role when learners' speech is rated by native speakers of English and have a lot to do with improvement in pronunciation. This chapter also describes the role of visual information which works effectively in directing learners' attention to audio information. Learners' spoken English is often assessed in terms of its "intelligibility," "comprehensibility," and "naturalness." However, the definitions for these terms vary depending on researchers. This chapter discusses these technical terms and how they are used.

Chapter 3 aims to focus on the relationship between speech duration and the perceived assessment of learners' English. Duration is one of the crucial prosodic features but the Japanese learners often have problems in producing English words with correct duration. It is noted that Japanese is a mora-timed language in which production of each mora takes about the same length of time in speech. Japanese EFL learners who are influenced by this phonological structure in their mother tongue often experience difficulties with durational control in producing function words and content words. The author assumes that if Japanese EFL learners learn durational control, their spoken English will be assessed much higher by listeners. In order to prove this hypothesis, in Experiment 1, the speech duration of a whole sentence and that of each segment in the sentence are measured, and also the locations of stressed and unstressed words are identified with the aid of a computer-assisted acoustic analyzer.

Chapter 4 further examines speech duration as analyzed in the previous chapter and aims to reveal the effects of pronunciation practice with animated materials focusing on duration. Reflecting upon the results of Experiment 1, the author sensed the necessity of creating animated materials which would visually appeal to learners and make them aware of sentence and phrase duration. Multi-media teaching materials in which stressed words are stretched out vertically while unstressed ones shrank

horizontally were created online. Experiment 2 was conducted to analyze whether Japanese EFL learners become able to acquire proper English durational control.

Chapter 5 demonstrates how the revised version of the original animated Web materials focusing on English rhythm enables Japanese EFL learners to improve English pronunciation. With this purpose in mind, Experiment 3 was conducted and the results were analyzed in terms of pitch, duration, and the correlation between raters' scores and related acoustic variables. A positive transfer of good pronunciation habits was confirmed.

The purpose of Experiment 4 conducted in Chapter 6 is: (1) to determine to what extent this original software focusing on intonation helps students improve their English pronunciation by examining the raters' evaluation and objective measurement using an acoustic analyzer; (2) to explore which prosodic features of English as spoken by Japanese EFL learners contribute to a higher evaluation; (3) to show how the visual presentation of English intonation contours on the screen contributes to heightening the learners' awareness of correct intonation.

Finally, Chapter 7 examines if the four major research questions raised in Chapter 1 were all sufficiently answered by summarizing the findings made in each chapter. The importance of “metacognition” in pronunciation learning will be suggested when reasons for the learner's improvement are interpreted. The author considers that the notion of metacognition in language learning will provide the theoretical background of this dissertation. In final chapter, the author discusses “blended learning”, combining traditional face-to-face learning with e-learning using multi-media applications whose positive effects on learners' pronunciation were verified in this study. This dissertation concludes with teaching implications, a list of limitations, and suggestions for possible future development in this field of research.

The next chapter reviews previous research which has examined the prosodic features of English spoken by Japanese EFL learners. This is followed by an explanation of the judgment scales used to assess learners' English i.e., "intelligibility" "comprehensibility," and "naturalness."

2. Literature Review

The following three sets of findings taken from previous literature on this subject area helped the author set up his research questions as mentioned in Chapter 1. Chronologically speaking, the first of these findings was that prosody encompassing rhythm, pitch, intonation, and duration contributes greatly to the production of natural-sounding English for Japanese EFL learners and furthermore, that these suprasegmental elements play an important role in successful communication (Dauer, 1993; Date, 1998; Lightbown & Spada, 2006; Suzuki, 1990; Sugito, 1996; Takefuta & John, 1967; Watanabe, 1994; Yamane, 1999). The second of these findings revealed that English spoken with a wider pitch range is judged to be more natural as English than that with monotonous intonation (Dauer, 1993; Shirai, 2001; Sugito, 1996; Suzuki, 1992; Takefuta, 1990; Watanabe, 1994; Yamane, 1999). The third finding was that providing Japanese EFL learners with audio and visual information works effectively in helping them utilize a broader pitch range and eventually enables them to use more native-like pitch patterns (Anderson-Hsieh, 1996; Arimoto, 1997; De Bot, 1983; Lambacher, 1997; Shimizu, 1987; Shirai, 2001; Sonobe, Ueda & Yamane, 2009; Sugito, 1996; Yabuuchi & Satoi, 2001; Yamane, 1999). Shirai (2001) also showed that demonstrating the pitch contour patterns to learners on a PC screen was more effective in improving their pronunciation than just having them listen to English. The above three sets of intriguing findings suggest that providing Japanese EFL learners with visual information on English intonation might work effectively in improving learners' often monotonous intonation.

The basic objectives of Chapter 2 are; (1) to introduce some of the typical prosodic features of English displayed by Japanese EFL learners referring to the findings made in earlier research; and (2) to summarize them in order to lay the foundation for the theoretical background of this study.

2.1 Prosodic Features of English Produced by Japanese EFL Learners

Suzuki (1992) showed that English with correct stressing and intonation but with incorrect segmental pronunciation will be understood better by native speakers of English than English with

correct segmentals but with incorrect stressing and intonation. The ability to pronounce segmentals such as vowels and consonants accurately is important, however, English prosody contributes greatly in making English sound natural and it plays an important role for successful communication (Shimizu, 1995; Suzuki, 1992; Sugito, 1996; Takefuta & Black, 1967; Yamane, 2006). These studies suggest that prosody plays a more important role in making non-native speech sound natural than the accurate pronunciation of segmental sounds.

The following sections in this chapter focus on four important aspects in English prosody; rhythm, duration, pitch and intonation, while reviewing previous literature on prosody. Also discussed below is the important role that visual information is expected to play in assisting learners in both broadening their pitch range and giving them durational control, when provided simultaneously with audio information.

2.2.1 Rhythm

It is known that English is a stress-timed language in which stressed syllables tend to be lengthened and other unstressed syllables shortened. Thus, the rhythm of English is created by constant alternation of strong / stressed syllables and weak / unstressed syllables. The stress syllables are generally longer, clearer, and sometimes higher in pitch than unstressed syllables, while the unstressed syllables tend to be shortened and their vowels reduced (Dauer, 1993; Ladefoged, 2006). In stress-timed languages such as in English, it is asserted that lexically-stressed syllables occur at regular time intervals; by contrast, in more-timed languages such as in Japanese, syllables themselves occur at equal time intervals. This phenomenon, known as isochrony in speech, means that there are approximately equal intervals of time between stressed syllables in connected speech. Unstressed syllables tend to be rapidly articulated, while stressed syllable are lengthened with extra duration. Lehiste (1973) found from acoustic analysis that stressed syllables do not occur exactly at regular intervals of time, though she mentions that there is a tendency for English to be perceptually isochronous. The isochrony in speech is thus considered to be a subjective phenomenon only

observed in perception. In other words, though we sense isochony in English rhythm when we listen to it, exact equal duration between stressed syllables is not physically confirmed.

Japanese which is categorized as a more-timed language has a syllable structure that consists of a sequence of a consonant and a vowel in this order, with the vowel as a mandatory constituent and the consonant as an optional constituent. For instance, the Japanese word “ri-zu-mu” consists of three syllables, or more precisely three “moras,” while the English word “rhythm” comprises one syllable. It is known that each syllable in Japanese has approximately the same duration and it is produced with nearly the same loudness (Abercrombie, 1967). These differences in phonological structures between the two languages make the pronunciation of English by Japanese EFL learners very challenging.

Good English rhythm will be created when learners become able to produce both stressed syllables and unstressed syllables properly. Also important is their ability to group syllables together into larger units (Dauer, 1993; Homma, 1981). This suggests that in linking words together to form a larger unit such as a phrase, no pauses should be placed in ungrammatical junctures. For example, a pause should not be inserted between unstressed function words and stressed content words. If a pause is inserted between “the” and “park” in the sentence, “They went for a walk in the park”, it would destroy the natural English rhythm. Pauses placed in an ungrammatical break like this tend to damage the isochrony in spoken English. Learners equipped with English grammar would know where to pause in a sentence.

As explained above, when English is spoken with correct phrasing and pausing, good rhythm tends to be realized. In order to acquire a good rhythm in English, it is necessary for Japanese EFL learners to practice pronunciation with connected speech. One of the sound changes observed in English connected speech is a “linking.” A linking which is also called a “liaison” occurs when a word ends with a consonant followed by a word beginning with a vowel. It is known that the regular intervals between beat placed on stressed syllables is kept constant when words are linked together. ‘R’ linking, where a word that ends with an “r” is linked to a word beginning with a consonant is a

well-known case of a linking. ‘R’ linking is seen in such common phrases as “for instance,” “clear it away,” and “take care of it.” ‘N’ linking is another phenomenon often observed in connected speech. However, Japanese EFL learners tend to pronounce these two words separately which often makes the maintenance of an English rhythm difficult (Pennington, 1987).

Dauer (1993) claims that in order to achieve a good rhythm in English, learners need to slow down, stretch out, and very clearly pronounce one-syllable content words and the stressed syllables of longer words. She also adds that vowels in unstressed function words as well as those in unstressed syllables in content words must be reduced to what is termed “schwa.” Schwa is also called a neutral vowel or a central vowel due to where it is articulated. Gimson (2008) reported that schwa (symbolized /ə/) is the most frequently used vowel out of all the vowels in RP and it accounts for approximately 27 percent of the all vowels used in spoken English. According to Date (2001), the most frequently used ten function words are “the”, “of”, “and” , “to”, “a”, “in”, “that”, “it”, “is”, “ ’ll (will) ” and these 10 words occupy about 25 percent of commonly spoken and written English words in our daily life. These function words usually contain a weak form of a vowel, schwa which is used more frequently than /i/ or /u/, since those words are in general unstressed.

As mentioned earlier, each syllable in Japanese has approximately the same duration and is produced with about the same loudness, which makes the pronunciation of each Japanese vowel relatively clear. As a result of this feature of their mother tongue, Japanese EFL learners tend to pronounce weak forms of function words rather strong. Sudo and Kiritani (1999) reported that among weak forms, articles and pronouns were the most frequently-misperceived variants. This implies that Japanese learners are not good at perceiving or producing a schwa sound.

Let the author summarize the afore-mentioned arguments concerning rhythm by pointing out five common phenomena observed in English spoken by Japanese EFL speakers: (1) it is difficult for learners to maintain isochrony in English; (2) they tend to place unnecessary pauses at grammatical boundaries, thereby failing to link the words together; (3) they tend to pronounce content words very fast by rushing them or dropping the final consonants; (4) they are not good at making enough vowel

reduction in function words and unstressed syllables; and (5) they are not good at perceiving or producing a schwa sound.

Considerable attention should be directed to prosodic features typically found in English for Japanese EFL learners to overcome these difficulties. It was this realization that encouraged the author to create multimedia applications for the purpose of directing the attention of learners to English rhythm in order to facilitate pronunciation learning.

2.2.2 Duration

Duration in this dissertation is defined as the length of time sound continues in grammatical units such as a word, a phrase, a clause, and a sentence. Duration is categorized as one of the most important prosodic features which can be measured by means of a computer speech analyzer. It is known that the frequency of pauses and their location in utterance greatly affect the listening comprehension of learners. Makino (2005), after conducting the acoustic analysis using 2-syllable words, suggested that both pitch and duration in English play a more important role than stress in its speech rhythm, in spite of the fact that English is labeled as a stress-accent language, concluding that “pitch- and length-accent” characterizes its rhythm.

Japanese EFL learners often have problems in producing English words and phrases of the correct duration, which is partly due to the fact that Japanese is a mora-timed language as mentioned earlier. Japanese words are made up of moras, i.e., a sound segment containing a single consonant followed by a vowel, each pronounced with approximately the same duration (Lambacher, 1997; Shimada, 2005). Ladefoged (2006) indicated that one of the most interesting languages in its use of length is Japanese. Japanese words such as [sukiyaki] “beef stew”, consist of four morae of this type. He noted that in the latter word the high vowel /u/ is voiceless because it occurs between two voiceless consonants, but it still takes about the same length of time to say as do the vowels in the other syllables. Japanese EFL learners tend to make the duration of the weak form in function words the same as that in content words. Sudo and Kaneko (2005), on the other hand, found in their

study that the duration of word-initial stops pronounced by Japanese learners was less than half of that produced by English native speakers. These findings suggest that Japanese learners often fail to make function words which should be short in duration, whereas their word-initial consonants are not long enough. Sudo and Kiritani (1999) who have conducted experiments on the perception of spoken English indicated that pronouns of a longer duration tended to be perceived more correctly by Japanese learners than those of a shorter duration. They also argued that, as a whole, the duration of weak forms influences the accuracy of perception as measured by Japanese learners, but it does not influence the perception of accuracy by American listeners. Their results can be interpreted as suggesting that Japanese learners would have difficulties with stress-related durational control since they are not required to exercise this kind of control when speaking their first language.

The author compared the duration of English sentences spoken by Japanese EFL learners with that produced by native speakers of English (Sonobe, 2006). Figures 2-1 and 2-2 show the results of comparison in speech duration between NSs and NNSs. Two sentences were read aloud by both NSs and NNSs and their speech samples were recorded, followed by careful analysis with a computer speech analyzer, *WaveSurfer*. Figure 2-1 shows the mean durational ratio of each word in a whole sentence, "I look for roses in the garden." This figure shows that when read aloud by four NSs, the mean durational ratio of four function words, i.e., "I", "for", "in", "the" accounts for only 28% of the total sentence duration, while the same mean ratio of NNSs (N=25) amounted to as much as 44%. Put another way, in speech read by NSs, unstressed syllables tend to be rapidly articulated, whereas stressed syllables are given extra length. The durational ratio of the content words, i.e., "look", "roses", and "garden" when read aloud by NSs reached 72% in the whole sentence duration, while the same mean ratio in speech by NNSs was 56%. These results imply that NNSs are more liable to prolong function words than NSs, which often results in arrhythmic speech.

Yamane (2001) suggested that Japanese EFL learners are likely to pronounce unaccented function words with strong forms. For instance, the preposition "for" in "The woman waited for her son at the window." shown in Figure 2-2 tends to be pronounced as /fɔ:/ instead of /fɹ/, which

explains the larger percentage for the duration of function words in a whole sentence when spoken by NNSs.

The most outstanding difference shown in Figure 2-1 between speech made by NSs and NNSs is found in the mean ratio of sentences that begin with the word "I." Note that the initial word, "I", as read by NNSs, occupied 15% of the whole sentence, while the same pronoun accounts for only 5% of reading time by NSs.

Another prominent difference in speech duration between NSs and NNSs is observed in the duration of the content word, "roses". As you can see from Figure 2-1, the durational ratio of this content word in a whole sentence when read by NSs occupies 27%, compared with 16% in non-native speech. This durational difference is considered to be caused by “consonant deletion” by Japanese learners. The NSs pronounced this word as /rouz□z/ , while many of the NNSs said /r□□zi/, deleting the final consonant /z/. The above finding suggests that NNSs not only substituted a diphthong /ou/ with /□□/ in “roses” but deleted the final consonant /z/. Suenobu, Kanzaki and Yamane (1992), examining the intelligibility of English spoken by Japanese learners found that “consonant deletion”, i.e., where a necessary consonant is deleted, is the biggest hindrance to intelligibility when being judged by NSs. They also found that a prosody error like “wrong accentuation”, i.e., when a word “COMmerce” is pronounced as “comMERCE” with a strong stress placed on the second syllable, also causes serious communication breakdown.

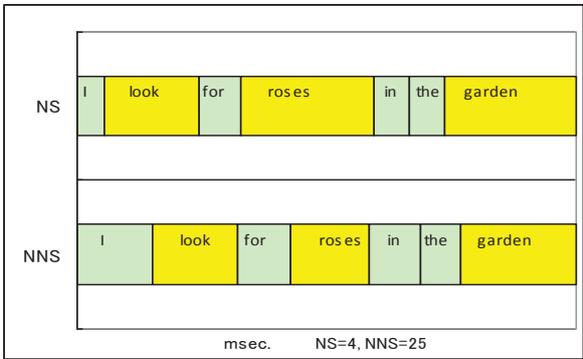


Figure 2-1. Durational Comparison Between NSs and NNSs: “I look for roses in the garden.”
 Note. N = 4 (NSs), 24 (NNSs)
 Cited from Sonobe (2006, p.8)

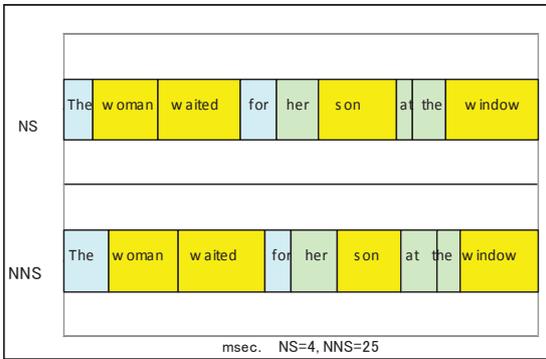


Figure 2-2. Durational Comparison Between NSs and NNSs: “The woman waited for her son at the window.”
 Note. N = 4 (NSs), 24 (NNSs)
 Cited from Sonobe (2006, p.8)

Figure 2-2 also compares speech duration in the sentence "The woman waited for her son at the window" as measured between readings by NSs and NNSs. Figure 2-2 manifested similar durational features to those indicated in Figure 2-1. It became obvious that the NNSs failed to reduce the duration of unstressed words with weak forms. It is likely that Japanese learners would have difficulties with stress-related durational control as they are not accustomed to exercising this control in speaking their first language. One of the possible ways for solving this problem for learners is to learn to pronounce weak syllables by using schwa. It has been emphasized in this section that Japanese EFL learners often have difficulties in durational control. However, Japanese people are supposed to be quite sensitive to word duration, i.e., we can easily distinguish "ojisan" (an uncle) from "ojiisan" (an elderly man) and also "tokei" (watch) from "tohkei" (statistics), whereas native speakers of English often have difficulties in distinguishing those two words (Kawagoe, 1999). The author believes that if the Japanese EFL learners are presented visual clues focusing on duration, it might facilitate both their production and perception of English. It is plausible that the use of multimedia applications focused on prosody as an aid of pronunciation training would be beneficial to Japanese learners of English.

2.2.3 Pitch

It is known that pitch, which is an auditory property, is determined by the following factors. They are: (1) the frequency of vibration of the vocal folds, two small flaps in the larynx. The higher the frequency of vocal fold vibration, the higher the pitch is perceived by listeners. ; (2) the tension of the vocal cords. They can be stretched to become longer and thinner. As a result, they vibrate more quickly, producing a higher pitch. ; (3) the air pressure from the lungs. An increase in the flow of air out of the lungs also causes an increase in pitch, so that stressed sounds usually have a higher pitch. ; and (4) the volume or the size of the vocal cords. Adults or males who possess larger vocal folds generally have lower pitch in their voice than children or females do (Collins & Mees, 2008; Ladefoged, 2006; Roach, 2009).

Accented syllables in English are constructed of three main features: pitch, loudness, and duration. It is generally agreed that English has a “stress accent” in which accented syllables are pronounced higher in pitch, louder, and longer than the other syllables. Thus, accented syllables become more prominent than other unaccented ones. The Japanese language, on the other hand, is labeled as a “pitch-accent” language in which a higher pitch on accented syllables is the sole prosodic feature that distinguishes them from unaccented ones in slow and formal speech. To put it another way, Japanese does not use a combination of intensity and duration to signal accent as in English. No syllable in Japanese is reduced especially in deliberate and careful speech, whereas unstressed vowels are likely to be reduced to a schwa sound in English, as has already been discussed in the earlier section. Researchers (Arimoto, 1997; Celce-Murcai, Brinton & Goodwill, 1996; Sugito, 1996) agree that this significant difference in accent patterns between the two languages makes acquisition of the English “stress accent” extremely challenging for Japanese EFL learners.

It has already been made clear by Yamane (2004) that the pitch range of English spoken by Japanese learners tends to be small compared with that of native English speakers, which makes the learner’s English monotonous. Yabuuchi and Sato (2001), examining prosodic features which differentiate “good readers” from “poor readers”, found that the evaluated scores of learners’ English reading significantly correlated with both the maximum F0 values and the pitch range. Their findings imply that the higher and the wider learners read English sentences, the higher they are assessed as “good” readers.” The F0 value represented in Hertz is often used in acoustic analysis. F0 or the fundamental frequency is determined by the rate of vibration of the vocal folds and it is intimately connected with the pitch we perceive” (Ashby & Maidment, 2005, p.154). As Ladefoged (2006) noted, “fundamental frequency is expressed with F0 and pitch can be expressed and measured with F0. When a speech sound goes up in frequency, it also goes up in pitch” (p. 23). Thus, the F0 value is a commonly used instrument in acoustic research to denote the pitch level in speech.

It is plausible that the inability of Japanese learners to use broader pitch in speaking English

lowers the intelligibility of their English, hindering smoother communication with NS (Celce-Murcai et al., 1996; Nema & Suzuki, 2000; Nishihara, 2003).

Shimizu (1987), analyzing the utterance produced by Japanese university students, revealed “incorrect placement of F0 peak to the unnecessary syllable component” as one of the typical features observed in learner prosody (p.13). The F0 peak is usually placed on accented syllables or nuclei in a sentence to express the speaker’s intension.

Another noteworthy finding concerning pitch was made by Mori (2005) and Watanabe (1994). They found that Japanese learners tend to start an English sentence with high pitch even if the first word of the sentence is an unstressed function word. They claim that this sentence-initial pitch pattern derives from the pitch pattern in Japanese. The experiment conducted by Mori revealed that about half or more native Japanese participants (college students majoring in English) produced unfocused subject pronouns *I*, *they*, *you*, and *it* in the sentence-initial position at a higher pitch than they did lexical verbs that followed. For example, in the utterance *I think*, the subject pronoun *I* is on a higher pitch than the following lexical verb *think*. She continued that “this intonational feature of Japanese speakers may lead to misunderstanding in English communication since a high pitch on the subject pronoun may be interpreted as the placement of focus or emphasis, such as when intending to contrast *I* with some other potential subject such as *you* or *he*”(p. 25).

We have seen in this section that there exists a significant difference in accent patterns between English and Japanese. In order to overcome this difference, the author considers that intentional training focused on prosody is essential to improve the pronunciation of Japanese learners.

2.2.4 Intonation

Dauer (1993) described intonation as follows (p. 219);

Intonation is the melody of speech, the changes in the pitch of the voice over time.

Intonation is fundamentally different from the other aspects of speech [...]. Consonants,

vowels, and stress have no meaning apart from the words they belong to. Intonation, on the other hand, can convey meaning directly. Besides being closely connected to grammar and words, it can express a speaker's emotions (anger, surprise), relationship to the listener, and attitude toward what he or she is saying (serious, joking). Most people use and respond to the intonation patterns of their native language without being aware of what they are doing.

As Dauer (1993) mentioned above, intonation plays an indispensable role in conveying speakers' intended message to listeners. "Within any given context, an utterance can be given a variety of different meanings according to the intonation patterns chosen by the speaker" (Underhill, 1994). However, for Japanese EFL learners, it is often very difficult to choose appropriate intonation for a given situation.

Wells (2006) cited a dialog as an example in which Japanese EFL learners usually make a mistake. In giving an answer to the question "What job would you like to have?" a possible reply would be "I haven't thought about it." NSs would put a nucleus on the word "thought" along with intonation peak on this word, while Japanese EFL learners tend to put it on "haven't." It is because in the Japanese language, the nucleus falls on a denying word in unmarked meaning. Watanabe (1994) illustrated the difficulties in intonation by citing different examples: (1) I didn't `TELEphone ||because I was `ANGry. ; (2) I didn't telephone because I was `ANGry. These two sentences are exactly the same when they are written, however, when they are read out in the ways shown by the tonetic notations, they begin to carry different meanings, i.e., the former sentence means "the reason why I didn't call was that I was actually angry, while the latter connotes another reason for not calling". As Underhill (1994) maintained that "while within certain limits of intelligibility mistakes or inappropriacies of pronunciation, grammar and even vocabulary can be accommodated by the native listener, inappropriate intonation can at times give rise not just to obscuration of the message, but to reception of a quite different message" (p.75). Nema and Suzuki (2000) insists that "in

Japanese sentences with falling intonation there are two stages in pitch range, while in English sentences with the same falling tone there exist three stages of pitch range” (p.141). This can be interpreted as meaning that the intonation change in English is more dynamic than that in Japanese. Shimizu (1995) suggests that “five major intonation patterns in English make things more complex for Japanese EFL learners” (p.94). Wells (2006) also mentions that “English makes more elaborate use of intonation to signal meaning than do most other languages” (p.11). The aforementioned findings all indicate that English has a more complex intonation system than Japanese, which makes it very difficult for Japanese EFL learners to acquire proper command of English intonation. It is indispensable for them to try to learn its use in order to maintain comfortable communication with both native and non-native English speakers.

Teaching materials in written media such as textbooks are able to explain how to read English sentences by illustrating the mechanism of pronunciation. Nevertheless, it is very difficult for non-native speakers to learn the actual use of appropriate rhythm and intonation from written media, since these features are basically mediated through voice. So far, several different ways to represent intonation have been created by phoneticians as will be explained in Chapter 6. However, as Hasebe (1987) suggested, there is an inevitable limitation to paper-based teaching materials and it has been said that English intonation is especially difficult for Japanese learners of English. The above circumstances motivated the author to introduce ICT (information and computer technology) teaching materials of pronunciation for the purpose of benefiting learners in acquiring the natural use of English prosody.

2.3 Visual Information

As we have discussed in the earlier section of this chapter, Japanese EFL learners often have difficulties in producing English words with appropriate duration, which is partly due to the fact that Japanese is a mora-timed language. Japanese words are made up of morae, each of which is pronounced with approximately the same duration. Are there any effective methods that help learners

acquire durational control in speaking English? Lambacher (1997) suggested that a visual image shown on a PC screen as a wave form assists students in grasping the duration of vowels and consonants and helps them learn the durational patterns. The dual-modal presentation, i.e., audio as well as visual, works effectively for learners to process the incoming information. This “modality effect” of multimedia content has been demonstrated in many previous experiments. According to previous studies (De Bot, 1983; Higgins, 1997; Lambacher, 1997; Shimizu, 1987; Shirai, 2001; Sonobe, Ueda & Yamane, 2009; Yamane, 1997), for example, providing learners with audio and visual information simultaneously is more effective for the learning of English prosody than when providing them with audio information only. Yamada (2004), on the other hand, claims that “auditory information plays a dominant role when audio-visual information is provided to learners at the same time” (p.43). However, “showing a speaker's face to learners can also serve as effective visual information in letting them recognize and produce sound” (McGurk & MacDonald, p.746). Although taking advantage of visual information such as pictures or moving images on a computer screen is rapidly becoming more popular in many ESL and foreign language programs around the world, very little research to date has been conducted on the effectiveness of visual feedback in teaching L2 pronunciation to second language learners.

Earlier studies place much emphasis on the effectiveness of visual information on learners' motivation. “Many Japanese students find it very helpful to sit in the phonetics laboratory looking at a computer display showing a picture of their pitch movements” (Roach, 1991, p.143). De Bot (1983) and Yamane (2006) also reported that students who are presented learning materials with both audio and visual information tended to repeat sentences more often and made more effort to correct their mistakes than the students who were given only audio. Another study reported that “after a few attempts using visual information they [learners] were able to correct their mistakes significantly” (Higgins, 1997). As Reid (1987) maintains “Japanese ESL students have strong visual and auditory preferences in their learning style” (p.87). These studies all imply that dual-modal presentation using multimedia applications has high potential to enhance learners' motivation.

It is easy to speculate from the above discussion that without proper phonetic instructions, producing English sentences with appropriate properties of prosodic features is difficult for Japanese EFL learners. Japanese learners are often unable to recognize where the problem with their pronunciation lies and what they should do in order to improve their spoken English. In many cases, they are not aware of their own pronunciation problems. Visual information is expected to play a very important role in having learners “notice” English prosodic features which are quite different from those in their L1, thus facilitating learners’ understanding of English sound and its prosodic features.

As previous findings imply, presenting visual information is expected to help: (1) motivate learners to practice; (2) recognize prosodic features, e.g., a pitch contour ; (3) notice their mistakes in prosody; and (4) set a feasible goal, e.g., the target number of repetition. When the number of times they have practiced is shown on a PC screen, it will work as a clear reminder of their practice. Learners will be more successful in improving spoken English by monitoring their speech and correcting themselves with the aid of visual information. It is easy to “see and compare” prosodic features such as intonation and speech duration, whereas it is relatively difficult to visualize segmentals on a screen. Takeuchi (2000) pointed out that teaching materials utilizing visual information should be created considering the function of the metacognitive processing mechanism (p. 138). A visualized image of prosody is expected to improve the learning process of pronunciation with the assistance of the metacognition of learners, such as knowing about learning or planning, monitoring, and self-evaluated learning.

2.4 Intelligibility, Comprehensibility and Perceived Naturalness as Spoken English

For more than fifty years, it has been considered that native-like English is the most desirable goal of pronunciation for Japanese EFL learners and they have spent a lot of time imitating British English (RP) and American English (GA). However, as Jenkins (2000) insists, it is now recognized that English is used for international communication by more non-native speakers than native

speakers. The author mentioned in Chapter 1 that as long as the meaning of the spoken message is clear enough to be understood by the addressee, a variety of English pronunciation phonetically modified with different regional accents should be accepted. However, in teaching NNSs, the author considers that a pronunciation model is necessary. Without it, learners are unable to set a definite goal which they can endeavor to approach. It is the author's belief that the goal of teaching pronunciation should not be to make learners sound like NSs but to make them speak intelligible English so that their pronunciation will not cause communication breakdown.

In the previous sections of this chapter, various prosodic features of English spoken by Japanese learners are discussed. It is easy to speculate that unnatural prosody, or even worse, prosodic errors in NNS speech would detract from the intelligibility of learners' English. NNS speech need not be exactly native-like, but it should be "intelligible", "comprehensible", and as "natural" as possible so as to secure comfortable communication.

Definitions of such terms as "intelligibility" and "comprehensibility" are diverse depending on researchers, and the methodologies of measuring these factors also vary. Munro and Derwing (1995) defined intelligibility as "the extent to which a speaker's message is actually understood by a listener" (p. 76). In Munro and Derwing (1995), intelligibility is measured by the ability of judges to transcribe the actual words of an utterance. When intelligibility is measured as exact word matches between learners' speech and the listeners' transcriptions as was done in their research, it became plausible that the intelligibility score would be affected more by segmental errors than prosody errors. In other words, unless segmentals are adequately pronounced, NS judges are not able to transcribe NNS speech correctly. The author doubts that accuracy in prosody would be properly measured with their method. Munro and Derwing also established an additional element to assess NNS speech, i.e., "comprehensibility," which was measured by an overall rating of how easy it was to understand a given speaker. It became clear from their researches that comprehensibility, which was assessed based upon the rater's total impression, was more strongly affected by the appropriateness of intonation than segmental correctness. This result implies that in measuring learners' prosody, rating

their speech impressionistically on a Likert scale would work effectively.

In many of other earlier researches, NNS speech has been assessed impressionistically on a Likert scale (Anderson-Hsieh, Johnson & Koehler, 1992; Kenworthy, 1987; Koster & Koet, 1993). In judging non-native speech holistically, “naturalness” of English would be an effective index in eliciting degrees of prosody correctness. Kashiwagi, Snyder and Craig (2006) contends, after summarizing findings in earlier researches on intelligibility of NNS speech, that “a listener’s impression of NNS pronunciation, whether it be comprehensibility or accentedness, tends to show high correlations with his/her suprasegmental scores.” Thus, the author decided to use “perceived naturalness” as a scale when NNS speech is assessed by NS and NNS raters in the 4 experiments to be conducted in the present research.

In the following chapter, the relationship between speech duration and the perceived assessment of learners' English will be discussed. As it was already explained in this chapter, speech duration is one of the crucial prosodic features which affect perceived naturalness of English. Experiment 1, described in Chapter 3, will illustrate that Japanese learners often have problems in producing English words with correct duration.

3. The Relationship Between Duration and Perceived Naturalness

3.1 Purpose

The purpose of this chapter in the current study is to compare English spoken by native speakers of American English to that spoken by Japanese EFL learners, focusing on the correlation between duration and perceived naturalness of learners' English when assessed by raters. In the previous chapter, the author attempted to elucidate reasons why English spoken by Japanese learners tends to sound unnatural, focusing on the temporal aspect of non-native speech. It has already been pointed out by Sudo and Kaneko (2005) that durational control affects the listener's understanding. In other words, they found that well-controlled speech duration makes the learner's English sound natural, thus facilitating the listener's comprehension. Though temporal control is considered to be one of the most important elements in influencing perceived naturalness, simply shortening the time to read a sentence, in other words, increasing speech rate does not always bring about higher perceived naturalness, as will be explained in Chapter 6 in detail. When English is spoken, a group of words are often linked together so that it constitutes a meaningful unit called a "chunk," which usually corresponds to a phrase or a clause. As we have already seen in Chapter 2, Japanese learners are not good at grouping words by "linking" them together. The failure in linking makes sentence duration longer, as is often the case with non-native speech. In order to create good rhythm in spoken English, a temporal control in stressed and unstressed syllables become indispensable, together with the mastery of chunking and linking.

3.2 Participants

Participants of this experiment were 21 junior college students (3 males and 18 females) enrolled in a correspondence course in English as their major. All of them were born in Japan and had learned English for at least six years in Japanese junior high and high school. Two of the participants had spent about three years in the United States accompanying their husbands on business. They were included as participants since they had stayed in the English speaking country

after the “critical period.”¹” Among the participants there were seven students who had experience of staying in English speaking countries either for travel or for taking short-term English lessons of around three weeks in duration. Their ages ranged from 58 to 20 years with a mean age of 30.04, however, if two elder participants are excluded from the mean, it comes down to 27.07.

Before conducting the experiment, a listening test was given to the participants to measure their English abilities. The test was selected from the 2nd grade of the STEP (often referred to as EIKEN) test which is administered by The Society for Testing English Proficiency. A listening test was implemented as a proficiency test, since it had been revealed from earlier studies that learners’ listening ability is highly correlated with the three other language skills , i.e., speaking, reading, and writing (DeMauro, 1992; Thompson, 1996). Table 3.1 below explains the English proficiency level of the participants. The mean score of the listening test turned out to be 62.0 out of the full mark of 100, which meant that English proficiency of most of the participants was equal or close to EIKEN level 2. According to the EIKEN official home page, the pass score of the first-stage test at 2nd to 5th grade levels is 60 % of the full mark (EIKEN, 2010).

Table 3-1
Basic Data of the Participants

Participants	<i>N</i>	<i>M</i>	<i>SD</i>
junior college students	21 (3 males, 18 females)	62.0	12.57

Note. The full mark is 100.

3.3 Experiment 1

An experiment was conducted during the three-day intensive lessons which the author offered. Prior to the experiment, participants were given a 90-minute lecture concerning English phonetics with special emphasis on the pronunciation of the weak forms in functional words. In the lesson, they learned how to pronounce unstressed and stressed words. In the following class, the experiment was conducted as explained below.

For the first step of the experiment, all participants were asked to practice reading aloud the four sentences shown in Table 3-2 for 30 minutes. The practice time of thirty minutes was considered to be sufficient to familiarize them with the pronunciation of the four sentences. The four practice sentences presented to the participants were simple sentences and arranged in order of the number of words and stressed syllables contained in the sentence, so that learners could practice shorter sentences earlier in the practice session. No demonstration reading by a NS was offered to the participants. In other words, they practiced on their own. One of the practice sentences was used as a test sentence for detailed acoustic analysis. Upon completion of the assigned practices, as a second step, they were asked to read a test sentence (*I should have sent her a message.*) aloud and their speech was recorded in a separate room so as to save their readings as digital files. The recordings were conducted on an individual basis in a quiet room so that sufficient sound quality could be secured for later acoustic analysis. A computer application called *SoundEngin*,² which is specifically designed to record, edit, and visualize sound, was used for recording learners' readings. In this experiment their speech samples, with a sampling frequency of 44.1 kHz, were used for the acoustic analysis with the aid of a speech analyzer called *WaveSurfer*.³ In this section, the speech duration of a whole sentence and that of each chunk in the sentence was measured acoustically. The location of stressed or unstressed words was also identified by utilizing a computer speech analyzer called *WaveSurfer*. The English spoken by Japanese EFL learners was recorded and then evaluated by three raters.

Table 3-2

Practice and Test Sentences

No.	English Sentences	No. of words	No. of stressed words	words per sec.
1	I should have sent her a message.	7	3	4.5
2	I look for roses in the garden.	7	3	4.8
3	The woman waited for her son at the window.	9	4	4.8
4	She wanted to buy a pair of tennis shoes.	9	4	4.8

Note. Sentence #1 was used as a test sentence.

The speech rates of all sentences used in this experiment were controlled to be from 4.5 to 4.8 words per second, taking into consideration the number of words, syllables and speech rate to meet the standard of the “Perceptual Sense Unit” (PSU) (Kohno, 2001). Kohno defines a PSU as a chunk of 7 ± 2 or less syllables which consists of a fundamental unit in both speaking and listening processes.

3.4 Results and Discussion

3.4.1 Evaluation by Raters

In order to have the naturalness of the English produced by EFL learners evaluated, two male Americans and one Japanese male were asked to be the raters for this experiment. The two Americans were both in their fifties and came from San Diego and Long Beach respectively. Since they have been teaching English at Japanese universities for more than 10 years, it was speculated that they would be getting used to Japanese English. Nevertheless, the level of exposure to English spoken by Japanese learners does not necessarily correspond to raters’ leniency or severity of the judgment. Suenobu, Kanzaki & Yamane (1992), who investigated the relationship between the judges’ scores and other parameters such as their age, educational background and experience with Japanese English, found no significant correlations between the two. The Japanese rater, who had studied English in England for eight years, was sixty-five years old at the time of this experiment.

The speech samples of the learners were randomized and copied on audio CDs. The raters were asked to listen to each speech sample three times to give pronunciation evaluation scores. The evaluation was made on a 10-level Likert scale based on the raters’ total impression as to the naturalness of English pronunciation. The lowest point of ‘one’ corresponds to *it sounds like very unnatural pronunciation to me*, and the highest point of ‘ten’ is equivalent to *it sounds like very natural pronunciation to me*, which is considered to be native-like. The mean evaluation score of the three raters on a 10-level Likert scale was 5.90 with the *SD* 0.99. The inter-rater reliability score among the three raters was Cronbach $\alpha = 0.817$, which is considered to be fairly high.

3.4.2 Objective Measurements

The left half of Table 3-3 lists the duration of each word in the test sentence, while the right half indicates the percentage of word duration occupied in the whole sentence. Absolute duration, measured in milliseconds (msec.), is necessary to observe the actual length of each word. The percentage of each word illuminates the relative length of time each word occupies in the whole sentence, which eventually tells us whether the sentence has rhythmic alternation with stressed / long syllables and unstressed / short syllables. When the distribution pattern of each word duration in learners' speech is similar to that of a NS, or more specifically, when the stressed words, "should", "sent", and "message" are pronounced longer than the other unstressed words, the English rhythm tends to be natural.

Table 3-3
Word Duration and the Percentage of Word Duration in the Test Sentence

words	Duration (msec.)								Percentage of Word Duration (%)							Total
	<i>I</i>	<i>should</i>	<i>have</i>	<i>sent</i>	<i>her</i>	<i>a</i>	<i>message</i>	Total	<i>I</i>	<i>should</i>	<i>have</i>	<i>sent</i>	<i>her</i>	<i>a</i>	<i>message</i>	
NS. M	79	190	114	248	129	41	574	1375	5.7	13.8	8.3	18.0	9.4	3.0	41.7	100
No1M	96	327	152	222	135	183	553	1668	5.8	19.6	9.1	13.3	8.1	11.0	33.2	100
No.2 F	173	223	459	236	477	296	678	2542	6.8	8.8	18.1	9.3	18.8	11.6	26.7	100
No.3 F	151	269	226	354	289	52	620	1961	7.7	13.7	11.5	18.1	14.7	2.7	31.6	100
No.4 F	152	298	192	397	338	66	729	2172	7.0	13.7	8.8	18.3	15.6	3.0	33.6	100
No.5 M	147	348	256	263	351	74	649	2088	7.0	16.7	12.3	12.6	16.8	3.5	31.1	100
No.6 F	80	212	329	198	216	52	687	1774	4.5	12.0	18.5	11.2	12.2	2.9	38.7	100
No.7 F	106	353	188	263	392	82	686	2070	5.1	17.1	9.1	12.7	18.9	4.0	33.1	100
No.8 F	102	236	178	489	209	76	543	1833	5.6	12.9	9.7	26.7	11.4	4.1	29.6	100
No.9 F	252	336	453	666	464	151	688	3010	8.4	11.2	15.0	22.1	15.4	5.0	22.9	100
No.10 M	134	213	179	306	295	71	683	1881	7.1	11.3	9.5	16.3	15.7	3.8	36.3	100
No.11F	107	215	329	180	215	69	680	1795	6.0	12.0	18.3	10.0	12.0	3.8	37.9	100
No.12 F	250	414	465	800	517	142	772	3360	7.4	12.3	13.8	23.8	15.4	4.2	23.0	100
No.13 F	165	302	338	485	249	58	609	2206	7.5	13.7	15.3	22.0	11.3	2.6	27.6	100
No.14 F	120	377	332	324	204	136	625	2118	5.7	17.8	15.7	15.3	9.6	6.4	29.5	100
No.15 F	155	231	282	430	402	131	704	2335	6.6	9.9	12.1	18.4	17.2	5.6	30.1	100
No.16 F	183	240	394	353	321	45	682	2218	8.3	10.8	17.8	15.9	14.5	2.0	30.7	100
No.17 F	73	287	338	161	190	135	544	1728	4.2	16.6	19.6	9.3	11.0	7.8	31.5	100
No.18 F	211	318	386	330	310	115	708	2378	8.9	13.4	16.2	13.9	13.0	4.8	29.8	100
No.19 F	118	277	159	224	327	118	371	1594	7.4	17.4	10.0	14.1	20.5	7.4	23.3	100
No.20 F	141	297	282	279	159	156	583	1897	7.4	15.7	14.9	14.7	8.4	8.2	30.7	100
No.21F	116	194	469	198	295	120	601	1993	5.8	9.7	23.5	9.9	14.8	6.0	30.2	100
mean	144	284	304	341	303	111	638	2125	6.7	13.6	14.2	15.6	14.1	5.3	30.5	100

Note. F signifies the female and M means the male participants

Figure 3-1 graphically illustrates the absolute duration of each word and the whole test sentence when read by a NS from the U.S. who had been teaching English at Japanese university for

twelve years and 21 NNSs. We can see from this graph that the NS was the fastest speaker and that it took more than twice as long for participants #9 and #12 to produce this test sentence.

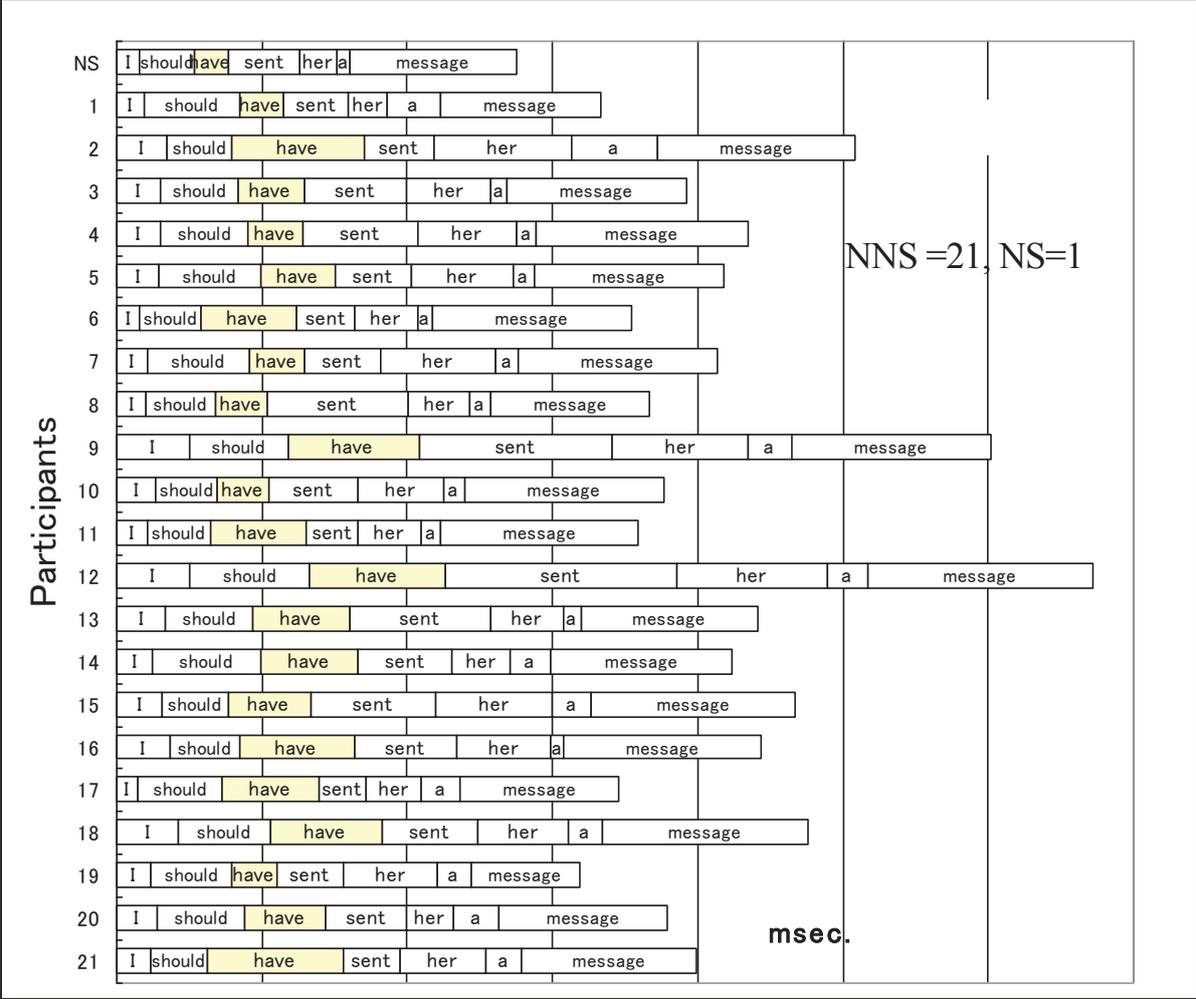


Figure 3-1. Comparisons in speech duration between a NS and Japanese EFL learners.

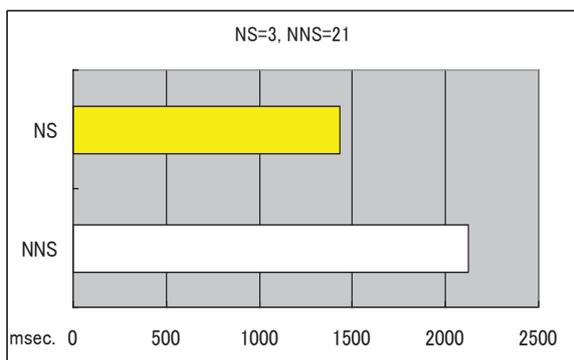


Figure 3-2. Comparison of the sentence duration between NSs and NNSs

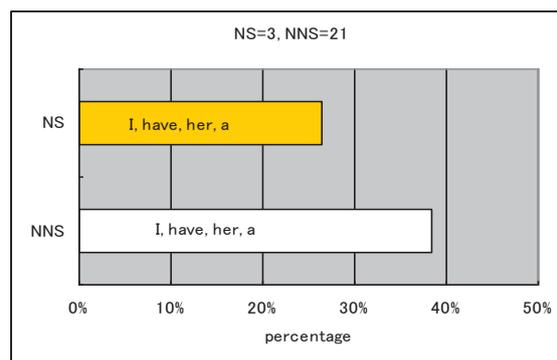


Figure 3-3. Ratio of unstressed words to the whole sentence duration: comparison between NSs and NNSs

Two additional NSs who were Americans and teaching English at Japanese university cooperated with this acoustic analysis. Figure 3-2 compares the mean sentence duration spoken by three NSs (1,431 msec.) with that of 21 NNSs (2,125 msec.). The sentence duration of NSs is approximately 32.6% shorter than the means for the NNSs, i.e., the mean sentence duration of NNS is 1.48 times longer than that of NSs. Figure 3-3 contrasts the difference between NSs and NNSs in the ratio of unstressed words to the whole sentence duration. In an unmarked reading of this sentence, the strong stress falls on “should”, “sent”, and “message.” Please note that the duration of unstressed words produced by NSs accounts for approximately 27.6% of the whole sentence duration, though the number of syllables in the unstressed words (4 syllables) consists of exactly half the total number of syllables (8 syllables) in this test sentence. In non-native speech, on the other hand, it occupies 38.5% of the whole sentence duration. Put another way, the speech duration of functional words (“I”, “have”, “her”, and “a”) pronounced with weak forms by NSs was 10.9% shorter than that produced by NNSs.

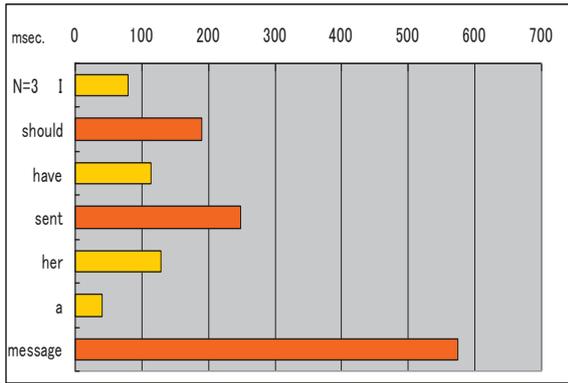


Figure 3-4. Word Duration spoken by NSs.

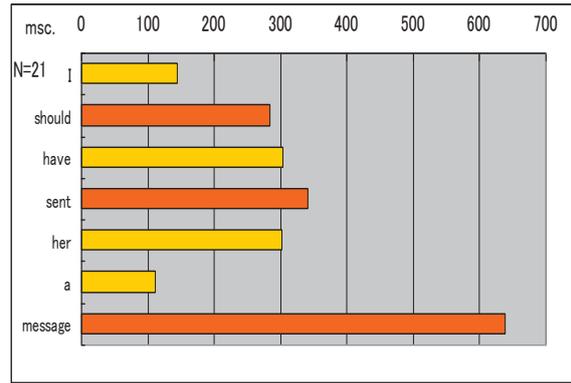


Figure 3-5. Mean Word Duration spoken by NNSs.

Displayed in Figure 3-4 above is the mean word duration of the test sentence spoken by three NSs, while illustrated in Figure 3-5 is the counterpart produced by 21 Japanese speakers. You can see from these figures that the combined duration of the four function words, “I”, “have”, “her” and “a” occupies only 27.6% of the whole sentence duration, whereas the duration of those four words in non-native speech accounts for 38.5% of the total sentence duration. These two figures imply that Japanese EFL learners who are influenced by the mora-timed rhythm of their mother tongue tend to produce unstressed syllables with a similar length to stressed ones. Non-native speech, which often sounds unnatural especially in its rhythm, is caused by the learner’s lack of temporal control when speaking English.

Figure 3-6 graphically demonstrates the mean duration of the “stress groups” in this test sentence when produced by the NSs and the NNSs. The stress group, or “foot”, can be defined as “a group of syllables, the first of which has the primary stress and the remaining are unstressed or secondary stressed syllables” (Brøndsted, 1997, p.1). A stressed syllable is an obligatory element in a stress group, while unstressed ones are optional. For example, the word sequence of “sent her a” constitutes a stress group, with a strong stress falling on “sent” followed by two unstressed syllables. The stress group “sent her a” is phonetically realized as /sent/ as if it were one word with “her” being pronounced with a weak form /r/ instead of a strong form /hɜr/. One outstanding feature in a stress group is that the unstressed syllables of function words next to a stressed word will be

pronounced weak and short. You can see from Figure 3-6 that the function words which follow the stressed words have very short duration in native speech. For example, the stress group “sent her a” with three syllables is even shorter than the two-syllable stress group “message” in native speech. One also notices that the mean duration of “sent her a” is longer than “message” in non-native speech. This clearly shows that unstressed syllables adjacent to stressed syllables were prolonged by the NNSs, demonstrating the influence of the mora-timed rhythm of the Japanese language, as was explained in detail in the previous chapter. The prolonged “her” of “sent her a” in non-native speech distorts the naturalness of English rhythm.

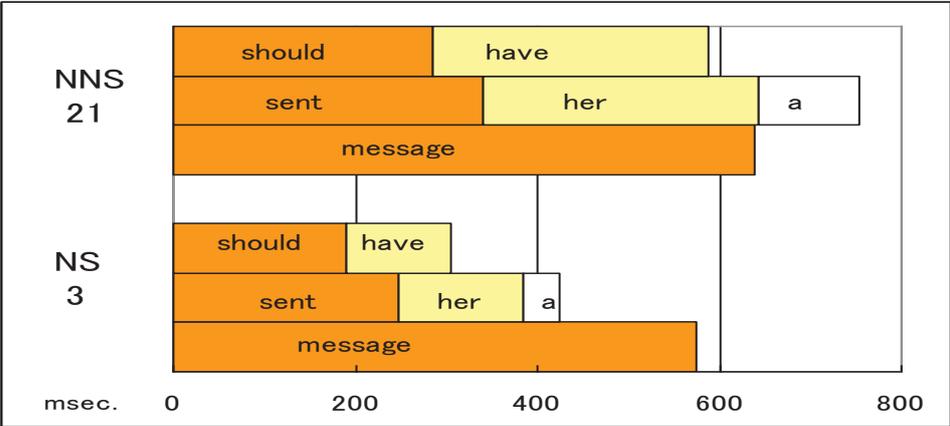


Figure 3-6. Comparison of stress groups between NSs & NNSs.

Another noteworthy phenomenon is that weak stressed function words are grouped together with a stressed word to form a stress group. For instance, the stress group “should have” /ʃʊd hæv/ sounds just like one word with “should” linked together with the /h/ sound of “have” being deleted. Please note that “should have” in non-native speech is almost twice as long as its counterpart in native speech. This stems from the fact that Japanese learners failed to link the two words together as the NSs did, thus it took much more time for them to produce this sequence. When “have” is pronounced with a strong form /h: v/ instead of a weak form /v/, it may also damage natural English rhythm.

The duration of the one-word stress group “message” is about the same in both native and non-native speech. Since the word “message” /mésids/ is often used as a Japanized English word, it

is often pronounced as /mese:dsi/, with the second vowel /i/ as a prolonged /e:/ and with a strong stress like a Japanese word. The placement of a strong stress on the wrong syllable may affect the naturalness in English rhythm and speech duration.

Table 3-4

Spearman Correlation Coefficients Between Raters' Evaluation Scores and Durational Parameters

	Durational Parameters	Correlation Coefficients
1	Sentence duration	-0.475*
2	Duration of each word in the test sentence	-0.405
3	Duration of the stress groups	-0.448
4	Ratio of unstressed words to the whole sentence	-0.415

Note. N=21

* $p < .05$

Table 3-4 summarizes correlations between raters' mean evaluation scores and four different durational parameters in non-native speech. A significant moderate negative correlation ($r = -.475$, $p < .05$) was detected between the means of raters' assessment scores and the mean sentence duration. There were also moderate negative correlations between the raters' mean evaluation scores and other durational parameters such as the duration of each word, the duration of stress groups, and the ratio of unstressed words to the whole sentence. These results can be interpreted as implying that the faster the reading speed of the learners, the higher the naturalness of their English is assessed by the raters. These results support the findings made by earlier studies (Sugito, 1996; Yabuuchi & Satoi, 2001; Yamane, Saito, & Yashima, 2004) in which significant correlations between learners' speech duration and the perceived naturalness as English were reported.

3.5 Conclusion

In concluding this chapter, let the author summarize its three major findings.

- (1) The results yielded from Experiment 1 have indicated that there was a significant negative correlation between the duration of the whole sentence produced by Japanese EFL learners and raters' evaluation. This suggests that should learners improve durational control in speaking English by being able to

pronounce the weak forms of function words properly, their English would be assessed more highly.

(2) It became evident from acoustic analysis that Japanese learners are inept at making weak forms of function words short in duration. The duration of function words in non-native speech accounts for 38.5% of the total duration of the test sentence. This typical phenomenon of lengthened function words can be explained by the fact that Japanese EFL learners are prone to add unnecessary vowels after consonants, influenced by the phonological structure of a Japanese mora, which in many cases consists of a sequence of a consonant followed by a vowel.

(3) What is called a “stress group” is formed when a stressed word is joined with unstressed syllables. In this word group, a stressed word is obligatory but unstressed syllables are optional. Thus, a stressed word alone can form one stress group. In the test sentence, a stressed word, “sent”, is followed by two unstressed syllables, “her a”, to form one stress group, “sent her a.” These unstressed syllables are usually squeezed together with the preceding stressed ones in a so called stressed-timed language like English. The duration of these stress groups is considered to be isochronous regardless of the number of unstressed syllables. Thus, the duration of those unstressed syllables are kept short in order to maintain the isochronism of a stress group. It was found in this experiment focusing on speech duration that Japanese EFL learners tend to produce fairly long duration in unstressed syllables influenced by the mora-timed rhythm of their mother tongue, in which each mora is produced with approximately the same duration. Learners should be encouraged to pronounce function words short and weak to make their English sound more natural.

(4) Japanese learners of English are not accustomed to temporal control of segmental duration which best suits the English stress pattern. As was previously

mentioned in Chapter 2, stressed syllables are pronounced longer, while unstressed ones tended to be shorter. The results of Experiment 1 revealed a high negative correlation between raters' mean evaluation scores and learners' speech duration.

Before concluding this chapter, the author should note three limitations that need to be acknowledged and addressed regarding Experiment 1. First, the number of participants, both NSs and NNSs, was quite small. Secondly, the time allocated for pronunciation practice was short. A longitudinal study is necessary to observe possible improvement in temporal features in English spoken by Japanese learners. Lastly, only one English sentence was acoustical analyzed to compare the speech duration produced by NSs and NNSs. Accumulation of objective data is necessary to endorse the findings.

Experiment 1 elucidated the shortcomings of Japanese learners in the temporal control of English. In order to help learners solve these problems, the author keenly felt the necessity of introducing multimedia applications aimed at improving learners' English rhythm. In the next chapter, Experiment 2 will be conducted to analyze whether Japanese EFL learners become able to acquire proper English durational control with the aid of animated learning materials which visually appeal to learners and make them aware of sentence and phrase duration.

Notes.

1. Critical Period is defined as a specific and limited period for language acquisition (Light & Spada, 1999, p.19).
2. SoundEngine is a program developed by Nicolas Juillerat and it allows users to play one background audio track, and multiple simultaneous sound effects.
3. WaveSurfer is an Open Source tool for sound visualization and manipulation developed at the Center for Speech Technology (CTT) at KTH in Stockholm, Sweden.

4. The Effects of Pronunciation Practice with Animated Materials Focusing on Duration

4.1 Introduction

Chapter 4 will discuss the effects of pronunciation practice with animated materials focusing on duration. This chapter further examines speech duration, which has already been looked at in the previous chapter. Experiment 2, which will be introduced in this chapter, will scrutinize the positive influence exerted by pronunciation practice involving animated materials on learners' pronunciation abilities. Japanese EFL learners, who are often influenced by the mora-timed rhythm of their mother tongue, tend to use Japanese rhythmic patterns when they speak English. In other words, novice¹ Japanese EFL learners are likely to pronounce every syllable with the same length regardless of its stress levels, as has already been discussed in the previous chapter. This interference from the L1 makes it difficult for learners to exert proper durational control over their spoken English. As Shimizu (1987) and Yamane (1999) insisted, intentional pronunciation practice is essential for learners to acquire the stress-timed English rhythm.

In order to assist learners to understand English rhythm and let them acquire proper English durational control, the author created multimedia teaching materials equipped with several functions focusing on durational control. The materials were put up on a website, so that the learners could practice English pronunciation while taking advantage of the visual and audio information provided by the multimedia application.

An experiment was conducted to examine the usefulness of this multimedia application available on the Internet. The participants in this experiment were divided into two groups, a treatment group and a contrast group. The learners in the treatment group (Group A) were shown animated letters which could move and be enlarged to highlight the location of stressed syllables when they practice English pronunciation. Learners in the contrast group (Group B), on the other hand, used still letters on the screen. The sentences were shown on a PC screen but the letters did not show any animated movement. The experiment was carried out once a week for five weeks in a CALL classroom as a part of regular class activities. In order to examine the effects of pronunciation

practice, raters were asked to assess learners' English and the learners' spoken English was also acoustically measured. By analyzing these two sets of data, i.e., the raters' evaluation scores and acoustic measurements of learners' speech, the effects which the multimedia application exerted on the improvement in pronunciation were evaluated.

Learners' speech was recorded both before and after the pronunciation practice to examine their improvement. It became evident from the results of a *t*-test that there was no significant difference between the raters' evaluation on learners' speech recorded on the two different occasions. A correlation was found between learners' speech duration and the means of assessment scores. A cluster analysis was also conducted to elucidate the relationships between the raters' evaluation scores and the learners' speech duration. A questionnaire was given to the participants to extract their opinions regarding the animated materials. On the whole, their comments as to the usability of the multimedia application were favorable.

4.2 Importance of Speech Duration in English Prosody

Earlier studies have ascertained the importance of speech duration in English for Japanese EFL learners, referring to a remarkable difference which exists in rhythmic structures between English and Japanese. It seems that Japanese learners have better manifest ability in perceiving durational differences in English than producing them. In perceiving English duration, Shimada (2005) suggested that Japanese EFL learners are able to differentiate long syllables from shorter ones, though they are less capable of making differences between stressed and unstressed syllables in speaking. His assertion coincides with the claim made by Hakuno (2002) that Japanese learners perceive speech duration with relative ease but tend to misperceive the locus of pitch and stress. Sugito (1996) contended that "Japanese rhythm is closely related to the duration of the word, while English rhythm is closely related to pitch and duration" (p.180). It is easy to speculate from the above argument that Japanese learners are equipped with innate potential ability to acquire durational control in English. The author believes that only intentional pronunciation practice makes the

acquisition of temporal control possible.

The most important finding made in Chapter 3 was that Japanese EFL learners are not good at controlling speech duration when speaking English. This is because their L1 is a mora-timed language, in which each mora, usually consisting of a consonant followed by a vowel, is pronounced with the same duration. Yamane (2001) mentions that “vowel addition is one of the signs influenced by Japanese phonological structure, in which a mora is completed with a vowel at the end of it.” Japanese learners who are influenced by this phonological structure of their L1 tend to lengthen English syllables by adding extraneous vowels.

4.3 Creating Multimedia Applications Focusing on Duration

The author felt the necessity for creating and developing multimedia applications for the purpose of teaching English pronunciation to Japanese EFL learners. He was especially motivated by the results of Experiment 1 conducted in the previous chapter and the earlier researches made by Lambacher, (1997), De Bot, (1983) and Reid (1987). Experiment 1 revealed the weakness in the temporal control of Japanese English learners in speaking English. The author believes that multimedia applications will be of assistance to remedy this weakness. The previous studies conducted by aforementioned researchers disclosed that the visual image of sound displayed on a computer screen assists students in visually grasping the duration of vowels and consonants, and thus facilitates the acquisition of the rhythm pattern.

de Bot (1983) also found that students exposed to both audio and visual feedback production tended to repeat sentences more often and made more effort to correct their mistakes than students exposed to only audio feedback. Reid (1987) mentioned that Japanese ESL students have strong visual and auditory preferences in their learning style.

Taking the effectiveness of visual information into consideration, the author developed the multimedia applications. Appendix I lists the OS environment and the names of the software applications which were used to create the multimedia teaching materials. The initial image of the

practice sentence is shown in Figure 4-1. With a click on the speaker button just below the sentence, a learner can listen to the model reading along with animated words. They have been revised with the following steps. The stressed words 1) were represented in red; 2) were animated so as to be stretched to the right direction; and 3) were stretched to the right and upward to make them larger. After taking the above-mentioned steps and incorporating learners' comments elicited from the questionnaire as to their impression on animated words, the author arrived at the final version which will be used in Experiment 2: stressed words were stretched upward and shown in red to highlight the location of the stress. Figure 4-2 is a visual image of one of the practice sentences used in Experiment 2, which will be elaborated on below.



Figure 4-1. Original phase.



Figure 4-2. Animated phase.

4.4 Experiment 2

4.4.1 Participants

The participants were divided into two groups, A and B. Group A was a treatment group consisting of 21 students, aged 18 to 20, while Group B was assigned as a contrast group with 18 students whose age ranged from 18 to 19. They were all Japanese university students who were non-English majors and their anonymity was secured. The teaching materials were provided with the two groups of learners using two different presentation modes on the Internet. Learners in Group A were assigned the presentation mode of “animated letters with audio,” while the students in Group B learned pronunciation using the “still letters with audio” mode. The learners in both groups logged onto the designated Internet site where each presentation mode was available.

Table 4-1.

Descriptive Data of the Two Groups

	Group A (<i>n</i> = 21)	Group B (<i>n</i> =18)
Presentation Mode	Animated Letters with audio	Still Letters with audio
Listening Test	<i>M</i> =12.57, <i>SD</i> = 3.60	<i>M</i> = 11.78, <i>SD</i> = 5.06
Independent <i>t</i> -test (two-tailed)	<i>df</i> = 37, <i>p</i> = 0.5820, <i>t</i> -value = 0.5554 n.s.	

Note. The full mark of the listening test is 20. *N*=39

Prior to the experiment, a listening test, comprising 10 questions selected from the second, the pre-second, and the third grade of the STEP test, was conducted on both groups to examine their English proficiency levels. A *t*-test was administered to confirm the homogeneity of the learners' English ability across the two groups. The results of the test revealed that there was no significant difference in English proficiency between the two groups. A listening test was used here since it is known that learners' listening ability is highly correlated with the other three linguistic skills, i.e.,

reading, writing, and speaking (DeMauro, 1992; Thompson, 1996). Please refer to Table 4-1 for the details of the descriptive data of the learners in the two groups.

4.4.2 Procedure

The experiment was carried out once a week for five weeks in a CALL classroom from April to May as a part of the regular class activities for both Groups A and B. The pronunciation practice was conducted on the Web for 15 minutes at the beginning of each lesson. The learners in both groups accessed the designated website and practiced using the listen-and-repeat method. They clicked a speaker button on the PC screen to listen to the model speech which was recorded by a Canadian male English instructor at a moderate speech rate and repeated in a CALL room. No explanation on pronunciation was given to the learner while they practiced.

The student's learning history was recorded via a CGI (Common Gateway Interface) program which was designed to accept and return the data on the Internet that conformed to its specifications. This system enabled instructors to trace back the record of the students' learning process. Instructors, for instance, were able to keep a record of the number of times learners repeated, or more precisely, clicked each sentence.

The eight English sentences used for the practice are listed in Appendix C. Words in these practice sentences were chosen from the corpus of the JACET list of 8000 (JACET 8000) basic words, one of the most widely used vocabulary frequency lists in Japan. The words used in the practice sentences were picked up from Level 1 (82.9%). Words belonging to Level 1 represent the most frequently used vocabulary in daily lives and Level 2 (10.5%) represents the second most frequently used words. The vocabulary in Levels 1 and 2 account for 93.4% of the corpus of daily use vocabulary.

The speech rate of the model speech by the NS in this experiment was controlled to be from 4.2 to 5.5 words per second, taking into consideration the number of words and syllables to meet the standard of the "Perceptual Sense Unit (PSU)" suggested by Kohno (2001). Kohno maintains that

holistic information processing will be made possible when units of this length are presented to learners.

Experiment 2 proceeded as follows: Among the eight sentences which were used for pronunciation practice, one test sentence, i.e., “I should have sent her a message.” was chosen for detailed analysis. Before starting the experiment in April, all participants were given five minutes so that they could practice reading it aloud. The readings by the participants were recorded individually by the author. The readings were saved in WAV digital form and were stored as “pre-test” files. An application program called *SoundEngine*, which was designed to edit audio digital data, was used for the recording. No demonstration readings by an NS, or reading instructions as to how to pronounce the sentence were given to the learners. Their speech samples, with a sampling frequency of 44.1kHz, underwent acoustic analysis using a computer speech analyzer, *WaveSurfer*. Upon completion of a 5-week session of pronunciation practice with the Website, learners in both groups recorded the same test sentence again and their oral readings were saved as “post-test” files to be used for both acoustic and assessment analysis.

The speech files were randomized and the raters (one Canadian, one American and one Japanese English teacher) were then asked to listen to each speech sample and give pronunciation evaluation scores. Both of the two native speakers of English had MA degrees in sociology and had been teaching English in Japan for more than ten years. Both of them were quite familiar with the phonetic features of Japanese. The Japanese rater, who majored in English in university, had worked in England and the U.S.A. for eight years. He had experience of teaching English at a national university in Japan for twelve years. Thus, he was considered to be eligible to rate Japanese learners’ English pronunciation.

The evaluation was made on a 10-level Likert scale based on the raters’ total impression as to the naturalness of English. They were asked to make their ratings on a scale ranging from 10: “It sounds perfectly natural.” to 1: “It sounds least natural.” The inter-rater reliability score among the three raters was Cronback $\alpha = 0.851$, which is considered to be very high.

4.5 Results and Discussion

4.5.1 Evaluation by the Raters

Learners' oral readings, which were recorded before they began the pronunciation practice with two different presentation modes, were examined by comparing the rater's mean assessment scores from both groups. In order to compare the difference in evaluation scores between Groups A and B, a *t*-test was conducted on the evaluated mean scores by the raters. As demonstrated in Table 4-2 below, no significant difference was detected in the mean scores between the two groups ($t = 0.305$, $df = 37$, $p = .761$). This result suggests that learners in the two different groups were homogeneous in terms of perceived naturalness of English before they initiated the pronunciation practice.

Table 4-2

Pre-practice Evaluation Results of the Test Sentence "I'll have her give you a call as soon as she is here."

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>		<i>d</i>
A	21	5.14	0.76	37	0.305	.761	n.s.	.099
B	18	5.05	1.00					

Notes. Skewness for A & B: 0.06 & -0.1114, Kurtosis for A & B: -1.04 & -0.42

After the learners practiced in two different presentation modes for five weeks, an independent two-tailed *t*-test was conducted on the raters' scores for both groups. As shown in Table 4-3, the test sentence spoken by the students of Group A was more highly evaluated than that of the students of Group B. There was a statistically significant difference in scores between the two groups ($t = 4.010$, $df = 37$, $p = .011$), this result implies that using the teaching materials with "animated letters with audio" tended to work more effectively in improving learners' English pronunciation than using still letters with audio.

Table 4-3

Post-practice Evaluation Results of the Test Sentence "I'll have her give you a call as soon as she is here."

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>
A	21	5.86	0.73	37	4.010	.011
B	18	5.44	0.78			

* $p < .05$

4.5.2 Acoustic Measurements

As a limitation of Experiment 1 introduced in the previous chapter, it was noted that the time allocated for pronunciation practice was short. Reflecting upon this limitation, in Experiment 2, the author would like to confirm how the participants' spoken English changed after a longer period of five-weeks of practice by measuring their speech acoustically with an aid of the *WaveSurfer* speech analyzer.

Table 4-4

Mean Duration and Evaluation Scores in Pre- and Post-test with the Correlation Coefficient

Group	<i>N</i>	Duration in Pre-test (msec.)	Evaluation scores in Pre-test	Correlation Coefficient	Duration in Post-test (msec.)	Evaluation scores in Post-test	Correlation Coefficient
A	21	4,843	5.14	-.504 *	3,612 (-25%)	5.86	-.717 *
B	18	5,176	5.06	-.506 *	4,520 (-13%)	5.44	-.708 *
NSs	5	2,372					

Note. NSs = native speakers

* $p < .05$

Table 4-4 shows that the participants in both Group A and Group B succeeded in making their sentence duration shorter after the five-week practice. The reduction rate in percentage for the learners in Group A turned out to be almost twice as much as that of Group B, i.e., learners in Group A reduced their sentence duration by 25%, compared with a reduction rate of 13% for learners in Group B. This result indicates that the practice with animated letters proved to be effective in making

learners' speech rate faster and closer to the speed of native model speakers. Furthermore, in post-test, high negative correlation coefficients between the duration and the pronunciation scores were found in both Group A ($r = -0.717$) and B ($r = -0.708$), which suggests that the learners' English was evaluated higher as their speech rate increased.

Table 4-5.

Comparison of Mean Duration in Segments between Group A Learners and NSs

	1st Segment (<i>I'll have her</i>)	2nd Segment (<i>give you a call</i>)	3rd Segment (<i>as soon as</i>)	4th Segment (<i>she's here</i>)	Total
Pre-test (msec.)	1,103	1,216	896	1,078	4,314
Post-test (msec.)	803	873	694	844	3,214
Change rate (%)	-27.2	-28.6	-22.6	-21.7	-25.5
NS(msec.) (N=5)	497	795	420	660	2,372
Ratio of Post-test to NS (%)	161	109	165	127	135

Note. Pauses between each segment have been excluded.

The test sentence, "I'll have her give you a call as soon as she's here", was divided into four segments in order to compare the segment duration between learners' speech and native model speech. Table 4-5 compares the mean duration of segments in the test sentence spoken by learners in Group A and those produced by five NSs. In Experiment 1, the test sentence was separated into "stress groups" focusing on the phonetic aspect, whereas in Experiment 2 the test sentence was divided into meaningful segments such as phrase and group conjunction focusing on grammatical aspects.

Table 4-5 shows that the mean sentence duration was reduced by 25.5% after the five weeks of practice. Another noteworthy result displayed in this table is that the duration of all four segments in the post-test was reduced by more than 20%. It also should be noted that the first and the second segments showed a larger reduction rate than the other two segments.

The duration ratio of learners' post-test speech when compared to NSs' was 161% in the first segment (*I'll have her*) and 165% in the third one (*as soon as*). The duration of these two segments in learners' reading remained a lot longer than that of NSs even after the 5-week session of pronunciation practice. Possible explanations are as follows. Learners failed to pronounce "her" with a weak form /ə/. They used a strong form /hɜːr/ instead, which made the segment duration long. When this sentence is read at a natural speed, "have" and "her" are usually linked together, resulting in [hævɜː] with a [h] sound being deleted. In the third segment, linking takes place between "soon" and "as", and both "as" are pronounced with a weak form of /əz/ instead of a strong form of /:z/. If the learners had succeeded in taking advantage of phonetic features such as linking and the weak form, the segment duration would have been made shorter.

The correlation coefficient between the raters' score and the number of practice repetition was .550 in Group A, which implies that learners' speech is more highly evaluated the more they repeat practicing the sentence.

4.5.2.1 Comparison of the Segment Duration Between NSs and NNSs

Figure 4-3 compares the segment duration of the test sentence as spoken by five NSs. They were four male American English teachers and one male Canadian English teacher. The four American teachers were from the west side of the United States and were all BA and MA degree holders. The one Canadian was from Vancouver and taught English at Japanese University on a part-time basis. As you can see from figure 4-3, the distributional pattern of segment duration was almost the same in the readings of NSs, even though the whole sentence duration was different. The common durational pattern, which can be observed in Figure 4-3, is that the second segment (*give you a call*) is the longest, while the third one (*as soon as*) is the shortest in all native speech, though the total sentence duration of each is different. NS-1, who usually speaks at a slow pace, turned out to be the slowest speaker among the five NSs, but his distributional pattern remained the same as the rest of NSs. This implies that the speech rate does not hinder the consistent distributional pattern of

segments, i.e., grammatical chunks.

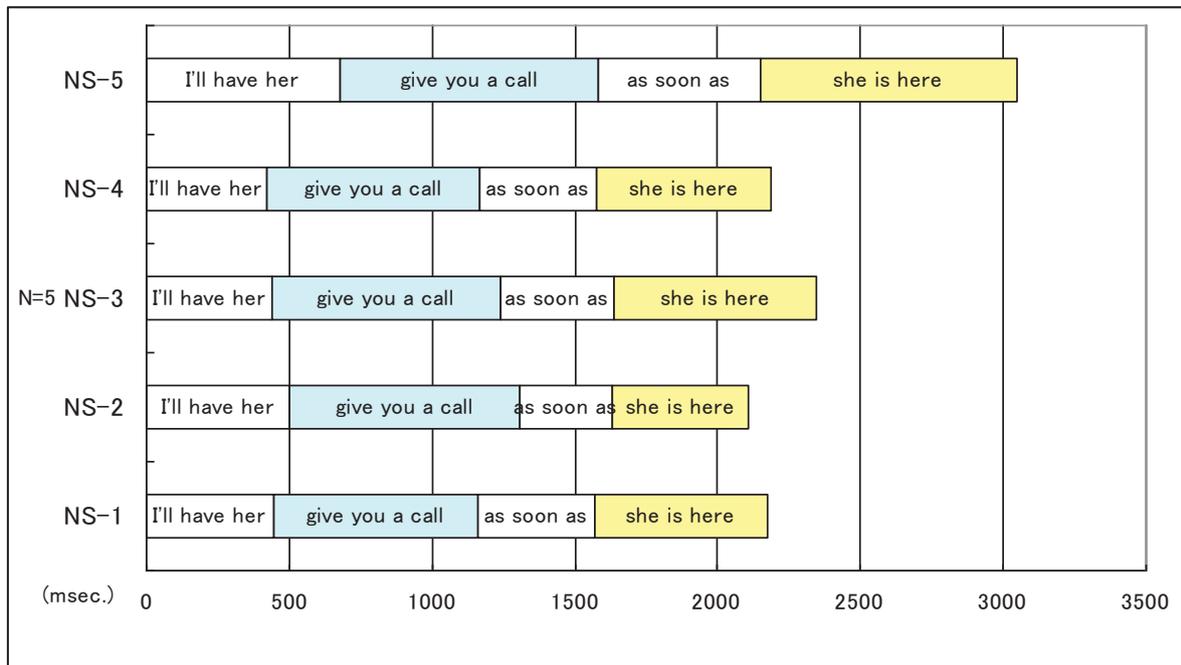


Figure 4-3. Duration comparison spoken by five NSs.

Table 4-6 compares the mean duration ratio of the four segments in the sentence readings by five NSs and 21 NNSs. As you can see from this table, the first, the second, and the fourth segment in learners' speech show a similar durational ratio of around mid-20 percent. On the other hand, uneven durational distribution is a distinctive feature found in the readings by native speakers of English. The second segment, *give you a call*, where two syllables are stressed, is the longest one (33.5%) in the readings by the NSs. On the other hand, an intriguing tendency was identified in the segment duration ratio of the whole sentence in learners' reading, i.e., the duration ratio of the three segments (the 1st, the 2nd, and the 4th segment) all centered around 26%.

Table 4-6.

Mean Percentage of Each Segment Duration in the Test Sentence

	1st segment (<i>I'll have her</i>)	2nd segment (<i>give you a call</i>)	3rd segment (<i>as soon as</i>)	4th segment (<i>she's here</i>)
NS ($N=5$) %	20.9	33.5	17.7	27.8
NNS ($N=21$) %	25.0	27.2	21.6	26.3

4.5.2.2 A Cluster Analysis

We have seen earlier that the distributional pattern of the segments in the test sentence spoken by NSs is different from that by NNSs, i.e., NSs manifested an uneven distribution, while NNSs showed a similar distribution among themselves. A cluster analysis was conducted for the purpose of examining if there is a correlation between the segment distribution pattern and the rater's mean evaluation scores. This statistical analysis assembles together readings which have a similar distributional pattern in segment duration. It is speculated that if learners' segment distribution was similar to that of NSs, their speech would be highly assessed. Shown as a dendrogram in Figure 4-4 is the result of a cluster analysis statistically examined based upon the durational distribution of segments. You can see from the dendrogram that four clusters have been generated.

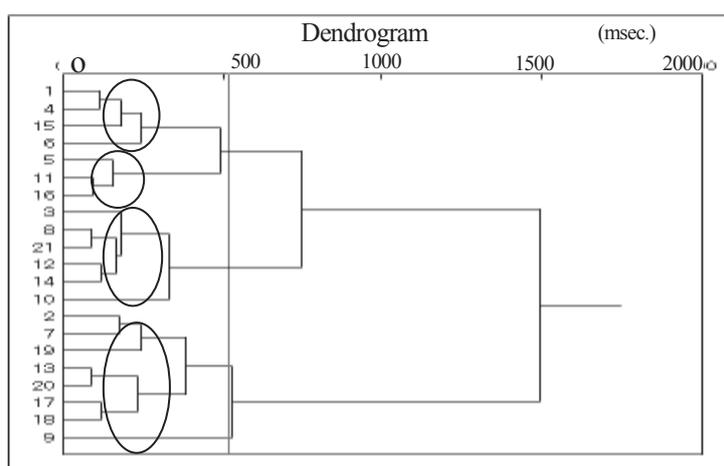


Figure 4-4. Results of a cluster analysis.

Four clusters with a similar distributional pattern yielded by a hierarchical cluster analysis are summarized in Figure 4-5. Indicated above the bars in this figure are the mean evaluation scores of each group. In the group which gained the highest mean evaluation score of 6.7, we can see that the distribution ratio of the second and the fourth segment is larger than that of the first and the third. In the group with the second highest evaluation scores (6.1 points), the third segment (*as soon as*) was long when compared with the same segment in the highest group. The lowest evaluation score (5.7 points) was given to the group of learners who produced the longest sentence duration.

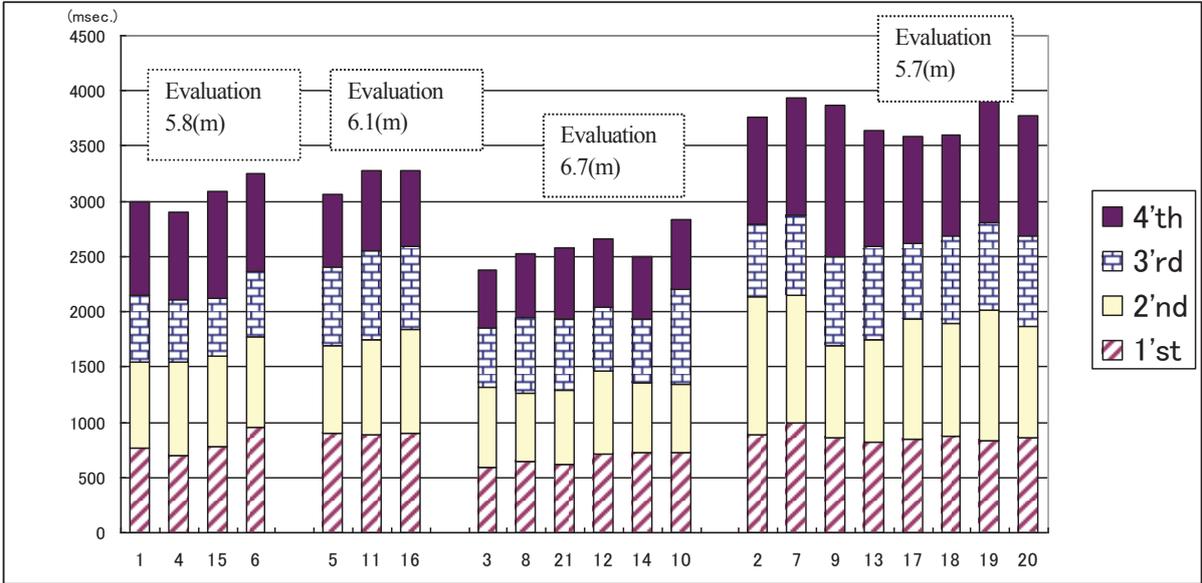


Figure 4-5. Results of a cluster analysis based on segment duration.

4.6 Questionnaire

After the five-week practice, a questionnaire was given to the students of the treatment group to elicit their opinions with regard to language communication skills and the affective factors in learning pronunciation with animated teaching materials. On the whole, as can be seen in Appendices A and B, learners' comments were favorable. In the skill category, the two questions with high scores were "I paid attention to rhythm rather than each word." and "I listened to the English sentences carefully." Learners responded with the lowest mean score to the question, "I

could speak English a little faster.” Their feelings about a slow improvement in speech rate correspond to the long segment duration, as was shown in Table 4-5.

In the category of affective factors, the questionnaire item, “The rhythm practice wasn’t monotonous.” gained the lowest mean scores, followed by “Repeating sentences many times wasn’t a burden.” These results can be interpreted as implying that the participants may have felt this type of repeating practice somewhat monotonous. Simply asking them to repeat the sentences is sometimes too demanding for learners to persevere. A more realistic attainment target should be presented to learners in order sustain their motivation by allowing them to notice how far they have made improvements in their pronunciation. Reflecting upon the feedback from the learners, the author decided to further revise the multimedia application.

4.7 Conclusion and Implication

Let us summarize the findings which have been made clear in this chapter.

(1) Taking visual effects into consideration, the author introduced multimedia applications with colored letters and stretchable words on a PC screen. This multimedia application specifically aimed at notifying learners of durational changes in English. The results of a *t*-test conducted to compare the difference in evaluation scores between Groups A and B revealed that there was a significant difference between these two groups after the five-week pronunciation practice. The results indicated that the learners in Group A, who took advantage of animated teaching materials, were more successful in improving their pronunciation than those in Group B, who were given only still letters.

(2) From the results of objective measurements using a speech analyzer, it became apparent that the learners in Group A were more successful in shortening sentence duration than those in Group B. From the results conducted in Experiment 2, a high correlation coefficient was confirmed between the raters’ scores and the learners’

speech duration, suggesting that their spoken English was highly evaluated as they came to speak faster. This endorsed the finding made in the previous chapter.

(3) The test sentence was divided into four segments and each segment duration was acoustically measured with a speech analyzer, *WaveSurfer*. The participants in both Groups A and B succeeded in making their sentence duration shorter after practice. The reduction rate of the learners in Group A (25%) was almost twice as high as that in Group B (13%). This result indicates that the practice with animated letters proved to be effective in making learners' speech rate faster and closer to the reading speed of the native model speakers. Furthermore, post-test, high negative correlation coefficients between the duration and the pronunciation scores were found in both groups.

The mean ratio of segment duration in a whole sentence remained almost the same in native speech, although there was a difference in the total duration of a sentence depending on the speakers. This implies that native speakers of English are able to change their speaking rate while keeping the durational ratio of each segment constant. When Japanese learners try to accelerate their speaking rate, their English often sounds awkward with the naturalness of English rhythm being impaired.

(4) A cluster analysis divided the participants into four groups based upon the similarity in distributional patterns of sentence segment duration. The group which gained the highest mean assessment score of 6.4 was the one which manifested a higher durational ratio in the second and the fourth segment than the first and the third one.

(5) A questionnaire with a 5-point scale administered to the students in Group A received favorable scores on the whole. Out of 10 question items which asked about learners' pronunciation skills, 6 items were answered with the mean rating of

above 4 points. Questions in the affective section focused on how learners felt while they practiced pronunciation using the multi-media application. It turned out that half of the question items gained above 4 points in the mean score.

This chapter found that after a 5-week session of pronunciation practice, learners who used “animated letters with audio” showed a greater improvement in their English pronunciation than those who used only still letters with audio, although statistical significance was not detected between the two groups of students. This result indicates: (1) the training period of 5 weeks was long enough to achieve visible improvement in learners’ pronunciation and (2) the multi-media application was appealing enough for the students to keep their motivation. However, as questionnaire items revealed, some of the learners might have felt this practice monotonous and others might regard it as a burden. Thus, the author felt the necessity for further improvement in this multi-media application, a revised version of which will be introduced in the next chapter.

Note

1. Novice learners in this dissertation are defined as those learners whose spoken English is strongly influenced by mora-timed Japanese.

5. The Effects of Pronunciation Practice with Animated Materials Focusing on English Prosody

5. 1 Introduction

Chapter 5 will demonstrate how the revised version of the original animated Web materials focusing on English rhythm enables Japanese EFL learners to improve English pronunciation. With this purpose in mind, Experiment 3 attempts to elucidate how the improvement in learners' pitch use affects the naturalness of their spoken English.

According to previous studies (Dauer, 1993; Sugito, 1996; Suzuki, 1992; Takefuta, 1990; Yamane, 1999), prosody, encompassing rhythm, accent, intonation and duration, contributes greatly to the production of natural-sounding English and also plays an important role in successful communication. Some of the prosodic features of the English produced by non-native speakers tend to be influenced by the phonetic characteristics of their mother tongue. For example, Japanese EFL learners tend to use Japanese rhythm when they speak in English (Sugito, 1996). Also, it is likely that Japanese EFL learners have difficulty with stress-related durational control since they are not accustomed to exercising such control in Japanese, as we have already seen in Chapter 2.

The author has further improved his original animated materials for use with the experiment contained within this chapter. For the purpose of remodeling the multimedia application, the computer software *Flash* was used to provide visual and audio information to the Japanese EFL learners. The letters were made to change in color and size, and rhythmic sound effects were added with a clapping sound to represent the stress and duration patterns of native speakers' utterances in order to assist learners in understanding English rhythm.

After a five-week period of training on the Web, the naturalness of the learners' spoken English was evaluated by three raters, including one Japanese professor.

Acoustic analyses were also conducted to measure the duration and pitch of the Japanese speakers' utterances by using the computer speech analyzer *Praat*. Many researchers agree that a combination of auditory and visual feedback can be highly effective in teaching segmentals, suprasegmentals, and other aspects of pronunciation (Anderson-Hsieh, 1992; Lambacher, 1999; Spaai & Hermes, 1993; Stenson, Downing, Smith, & Smith, 1992). Through statistical analyses, Experiment 3, conducted in this chapter, demonstrated that English pronunciation training using visual information helped to improve learners' pronunciation.

In oral communication, suprasegmental features such as rhythm and intonation are regarded as substantial elements (Suzuki, 1992; Takefuta, 1990; Yamane, 1999, 2006). For instance, a high correlation was detected between the values of maximum F0 and the perceived naturalness of English (Yabuuchi & Satoi, 2001). The results obtained from these studies imply that English spoken with a wider pitch range is judged to be more natural than that spoken with monotonous intonation.

The "intelligibility" of English spoken by EFL learners was measured by the ability to transcribe the actual words uttered. Their comprehensibility was evaluated by an overall rating of how easily it was understood by native speakers of English (Derwing & Munro, 1997; Derwing, Munro & Thomson, 2008; Munro & Derwing 1995). In other words, the intelligibility was judged objectively as the number of words native speakers of English were able to correctly transcribe when they listened to the non-native speech, while comprehensibility was rated subjectively as to how comprehensible the non-native speech was for them. In the author's research, however, learners' spoken English will be assessed by raters based upon perceived naturalness of English, since it is speculated that NS judges would easily transcribe the short sentences which will be used in Experiment 3 when they are read aloud by Japanese learners. De Bot (1983) also observed that learners were more motivated to correct

pronunciation errors when visual feedback was supplied.

Shirai et al. (2001) found from their experiment that Japanese EFL learners who learned only text-based English sentences with Japanese translations are less likely to use broader pitch patterns than native speakers of English. It was also reported in their study that Japanese learners who were only repeating English sentence spoken by NSs did not broaden their pitch range. However, learners who were provided with visual information such as the pitch patterns of NSs as they were speaking succeeded in widening their pitch ranges which became more like those of native speakers of English as a result. Their results suggest that giving visual information is far more effective than simply having them repeat the sentences.

5. 2 The Present Study

5. 2. 1 Purposes

The aims of the study in this chapter are (1) to demonstrate how the authors' original animated Web materials focusing on English rhythm enables learners to improve their English pronunciation; (2) to determine which phonetic features of English spoken by Japanese EFL learners contribute to a more positive evaluation by raters; and (3) to measure learners' improvement in pronunciation objectively through the use of an acoustic analyzer.

5. 2. 2 Creating Two Different Types of Web Materials

Two different types of pronunciation practice materials for Japanese EFL learners were created and were posted on the Web using *Flash*. Please refer to Appendix I for details of the OS environment and a list of software programs which were used to create the multimedia teaching material. The following describes the process of making the first type of learning materials:

- 1) As a first step, the color of the letters in each sentence was changed in order to make the stressed syllables stand out.
- 2) The letters on the screen were stretched out to further attract the learners' attention to the stressed syllables.
- 3) A function was added to count the number of times learners repeated each practice sentence.
- 4) The letters and signs were then animated by stretching and shrinking them, so that the learners could visualize the stressed and unstressed words as well as the correct rhythm. As Figure 5-1 below illustrates, the letters carrying stressed syllables are enlarged in sync with the actual pronunciation and those with weak stress are reduced in size to convey English rhythm to the learners. The slur signs () were also added to get learners accustomed to the linked pronunciation of adjacent words. In cases where an unstressed syllable (word) with an adjacent stressed syllable (word) is pronounced as one word, the space between the two words was deleted.
- 5) Clapping sounds were added to represent the stressed syllables.
- 6) On the Web screen, a mini-sized video clip of a native speaker of English pronouncing sentences was added to provide learners with better visual clues.

Wagener (2006) suggests that speakers' lip movements, facial expressions and body language can significantly facilitate understanding of the language used and help the learner to recognize important paralinguistic signals. Hill (1999) also mentions that the moving image is an "inherently attractive and compelling medium with great potential for motivating learners" (p.5). The author believes that the use of video film supports language learners and motivates the learning of pronunciation.

In the second type of pronunciation learning materials, only still letters are

provided along with the model pronunciation of a native speaker of English. As shown in Figure 5-2 below, by clicking the speaker button below the practice sentence on the screen, learners are able to listen to a model voice while looking at the still letters. The model pronunciation was provided by an American male in his early 40s from Chicago. Figure 5-1 below represents the first type of practice materials with animated letters synchronized with model pronunciation while Figure 5-2 illustrates the second type of materials with still letters plus model pronunciation.



Figure 5-1. Condition A (animated letters + sound).

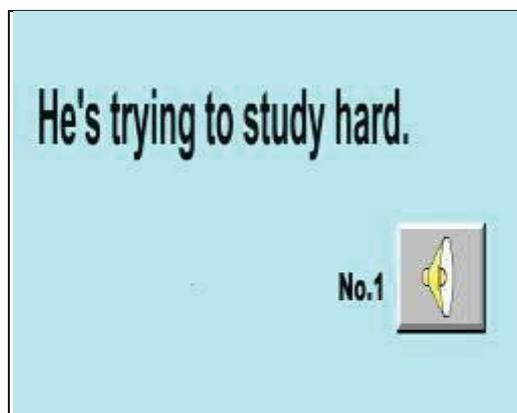


Figure 5-2. Condition B (still letters + sound).

5.2.3 Three-Step Learning System

The following is a description of the three-step learning system adopted for this pronunciation practice. Before beginning their practice on the Web, learners need to log in by filling out their assigned number and name in the blank boxes. This three-step learning system is followed by learners using the first type of materials.

(1) The first step is “a listening phase”: learners just listen to each sentence 10 times while concentrating on English rhythm. As shown in Figure 5-3 below, sentences are masked during this phase. At this step, learners try to hold in their working memory both the auditory image from the model pronunciation and the visual image from the video clip showing how to articulate the target phrase while watching the mouth

movement for correct pronunciation.

(2) The second step is “a parallel reading phase”: learners read the sentences aloud 30 times each while watching animated letters as shown in Figure 5-4. With a click of the speaker button on the right of the screen, the color of the English sentence turns blue, signaling that it is ready for the learner to begin reading. Without this function, learners would not be able to control the lag time before launching their parallel reading. The learners must try to speak aloud in synch with the model speech, without going ahead or falling behind.



Figure 5-3. Listening and articulation planning phase.



Figure 5-4. Parallel reading phase.

(3) Shown in Figure 5-5 is the third step: “a sign shadowing phase”. Here, learners repeat the sentences 10 times each, synchronizing with the model voice while watching circles of different sizes on the screen, with the larger circles representing the accented syllables. Letters did not appear on the screen during this practice phase. Since learners have visual information about the articulators from the video clip, this may have helped them to correctly pronounce consonant and vowel sounds.

In total, learners are required to repeat each sentence 50 times. It takes at most four minutes to complete all three phases for one sentence. With a click of the speaker button on the Web, the repetition data is sent to the instructor and recorded

automatically. Whenever learners successfully finish three phases of their pronunciation practice for each sentence, they are automatically rewarded visually by receiving an “excellent job mark” as well as information about how many times they have practiced the sentence, as shown in Figure 5-6.



Figure 5-5. Sign Shadowing Phase.



Figure 5-6. Excellent Job Mark.

5.3 Experiment 3

5.3.1 Basic Data of the Two Groups

The subjects are divided into two groups, A and B. Group A is a treatment group of 27 Japanese university learners, aged 18 to 22. Group B is a contrast group consisting of 21 Japanese university learners, aged 18 to 19. None of the learners in either group majored in English. The mode of viewing for treatment Group A involved “animated materials plus sound” and for contrast Group B, “text based materials plus sound.” Learners in Group A followed the three-step practice system described above, while participants in Group B practiced using the Web site where only still letters and model pronunciation were provided.

Prior to the experiment, a listening test was conducted on both groups to determine their English abilities. The results of a *t*-test showed that there was no significant difference between these two groups. A listening test was employed as a

homogeneity test here as it has been suggested by researchers that learners' listening ability is highly correlated with the other three skills, i.e., reading, writing, and speaking (DeMauro,1992; Thompson, 1996). This preliminary data is summarized in Table 5-1 below.

Table 5-1.

Basic Data of Two Groups

	Group A $N = 27$	Group B $N = 21$
Presentation mode	Animated letters + Sound	Still letters +Sound
Listening Test	$M = 15.59, SD = 5.806$	$M = 13.33, SD = 4.629$
Independent t -test (two-tailed) $df = 46, p = .1517, t$ -value = 1.45	n.s.	$d=.439$

5.3.2 Procedures

The experiment was carried out once a week for five weeks in a CALL classroom as part of the regular class activities for both Groups A and B. The learners in both groups practiced English pronunciation on the Web for 20 minutes at the beginning of each lesson. They recorded their learning history via a CGI program which was designed to accept and return data on the Web that conformed to CGI specifications. The 12 English sentences used for practice are listed in Appendix F. They were arranged according to the number of words and stressed syllables to allow for learners to practice shorter sentences earlier in the practice sequence. The speech rates of all sentences used in this experiment were controlled to be from 3.8 to 5.5 words per second, taking into consideration the number of words, syllables and speech rate to meet the perceptual sense unit standard (Kohno, 2001).

In order to evaluate the naturalness of the English produced by the EFL learners, two American males and one Japanese male were asked to be the raters for this study. All three had been teaching English at Japanese universities for several years. The

speech samples of the learners, recorded at random on CDs, were given to the raters for evaluation. The evaluation was made on a 10-level Likert scale based on the raters' total impression of the naturalness of the English. The inter-rater reliability score among the three raters was Cronbach $\alpha = 0.8867$, which is considered to be very high.

The experiment proceeded as follows: in April, prior to beginning the experiment, all learners in both groups were asked to read aloud the two test sentences below which were recorded and saved as pre-test samples. In May, upon completion of the assigned practices, they were again given the same two sentences to pronounce as a post-test. In both the pre- and post-tests, no demonstration reading was offered. Their speech samples, with a sampling frequency of 44.1kHz, were used for the acoustic analysis in this study. The sample sentences used for the pre- and post-pronunciation tests were as follows:

- 1) *I sold the watch she gave me.*
- 2) *I love to look at the moon and stars at night.*

The first sentence is one of the twelve sentences used for pronunciation practice, but the second sentence was not included among the sentences for pronunciation practice. The two test sentences were analyzed in Experiment 3, reflecting on the experiment procedures used in both Experiments 1 and 2 in which only one sentence was used for data analysis.

5.4 Results and Discussion

5.4.1 Evaluation by raters

Before initiating our experiment, test sentence 1 was given to all learners in both Groups A and B in order to compare the pronunciation abilities of the two groups. A *t*-test was conducted on the evaluated mean scores by the raters, indicating no significant difference between the two groups. Learners in both groups had almost the

same pronunciation ability at the outset of the experiment.

Table 5-2.

Pre-practice Evaluation Results of Sentence 1 ("I sold the watch she gave me.")

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	
A	27	4.54	.91	46	.75	.455	n.s.
B	21	4.71	.60				

After the learners practiced under different conditions for five weeks, an independent two-tailed t-test was conducted on the raters' scores for both groups. As shown in Table 5-3, with regard to sentence 1, the speech of the Group A learners was much more highly evaluated than that of the learners in Group B ($t = 4.02$, $df = 46$, $p < .01$).

Table 5-3.

Evaluation Results of Sentence 1 ("I sold the watch she gave me.")

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
A	27	6.29	0.61	46	4.02	.000 **	.745
B	21	5.51	0.72				

** $p < .01$

As for the second sentence, which was not included in the practice, a positive transfer was observed. As Table 5-4 indicates, the learners' pronunciation in Group A was highly evaluated by the raters with a statistically significant difference in score between the two groups ($t = 3.61$, $df = 46$, $p < .01$). This result suggests that once good pronunciation habits were acquired through this rhythmic practice, they were transferred to other English sentences as well. The rhythm pattern found in a sequence where a weak syllable is followed both by a stressed and a weak syllable in such

practice sentences like “I looked for ...” and “I need a ...” was transferred to a new combination of “I love to ...” in Sentence 2. A typical stress pattern they learned in phrases such as “in the garden” and “on Sunday” was also applied to the new phrase of “at night” in Sentence 2.

Table 5-4.

Evaluation Results of Sentence 2 (“I love to look at the moon and stars at night.”)

Group	<i>N</i>	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
A	27	6.43	0.65	46	3.61	.0007 **	1.840
B	21	5.76	0.60				

** $p < .01$

5.4.2 Objective Measurements

It is often suggested that the English spoken by Japanese EFL learners lacks in pitch variation and is often dubbed monotonous (Sugito, 1996). In this section the author would like to explore how the participants’ spoken English changed after the five-week practice by measuring their speech objectively with the aid of a speech analyzer. In order to examine the prosodic features of English spoken by Japanese EFL learners, the pitch values and duration were measured using special software for acoustic analysis called *Praat*, since significant correlations have been suggested between these two acoustic parameters and the perceived naturalness of the English produced (Sugito, 1996; Yabuuchi & Satoi, 2001; Yamane, Saito, & Yashima, 2004). For the two test sentences, both the pitch range of the stressed words carrying the nucleus and the duration of the sentence were measured.

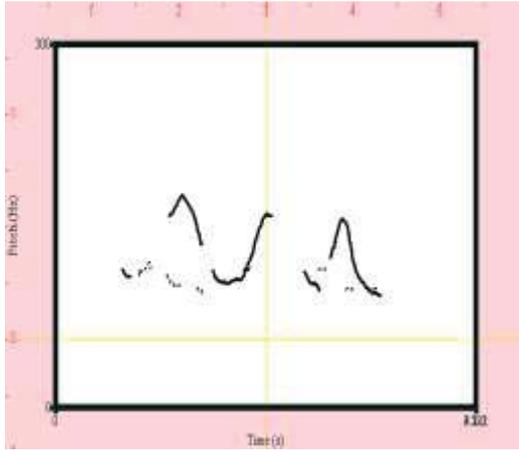


Figure 5-7. Intonation curve of sentence 1 produced by a learner from Group A.

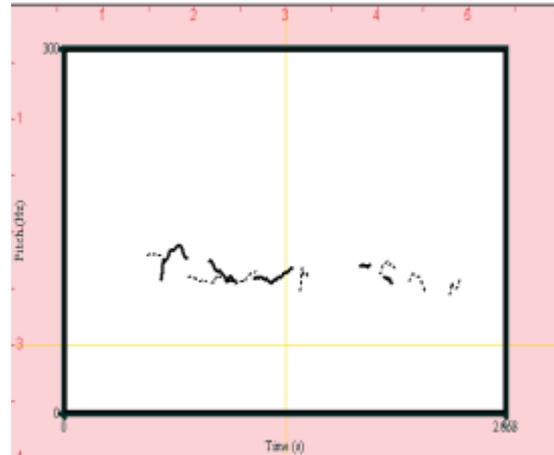


Figure 5-8. Intonation curve of sentence 1 produced by a learner from Group B.

Figure 5-7 indicates the intonation curves of sentence #1 produced by a learner from the treatment group. The dotted line shows the intonation curve in pre-practice speech and the solid one, in post-practice speech. The solid line indicates significantly broader fluctuation in intonation movement. This result coincides with the improvement in mean speech evaluation scores rated for this learner from 3.3 to 6.3. We can observe from Figure 5-7 that the pitch range of the first stressed word, “sold”, became broader in post-practice speech. Japanese EFL learners tend to start an English sentence with high pitch even if the first word of the sentence is an unstressed function word (Mori, 2005). This sentence-initial pitch pattern was improved by the learners in the treatment group.

Figure 5-8 indicates the pitch changes produced by a learner from Group B. As in Figure 5-7, the dotted line shows the pre-practice intonation curve and the solid one, post-practice. This male learner is typical of the participants from the contrast group in that his speech rate became faster, i.e. his sentence duration became 2,220 msec to 1,780 msec, while the pitch range remained narrow. His mean score as evaluated by the raters improved slightly from 5.3 to 5.7.

The pitch ranges of the three stressed words in test sentence #1 were measured. The pitch range in hertz (Hz) was calculated by subtracting the lowest pitch value from the highest value for each stressed word. Stressed words in a sentence carrying nuclei are usually pronounced with prominent pitch change. The English spoken by Japanese EFL learners tends to be judged as monotonous since it lacks these pitch variations. This typical phonetic feature is not highly evaluated as natural spoken English (Takefuta, 1990; Yabuuchi & Sato, 2001; Yamane, 2006). The author was thus interested to know the changes in the pitch range of the stressed words.

Figure 5-9 illustrates the change in the pitch range measured in sentence #1. The three arrows in this figure represent the mean pitch ranges of the three stressed words as pronounced by two native speakers of English. The two sets of bars on the left show the mean pitch range in pre- and post-practice for Group A and the next two sets to the right, for Group B. As you can see from these results, learners in the treatment group were more successful in widening the pitch range than those in Group B.

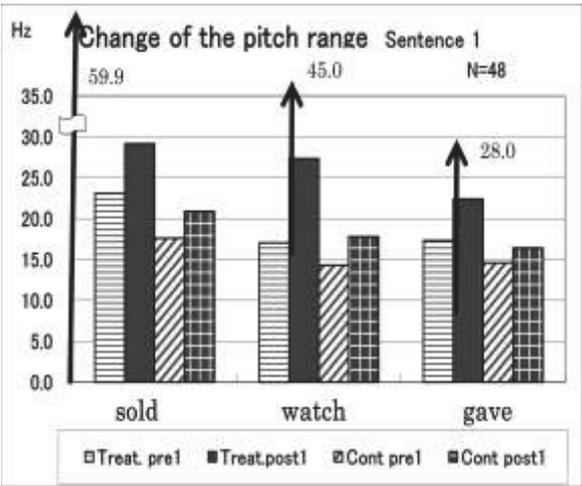


Figure 5-9. Change of mean pitch range in sentence #1.

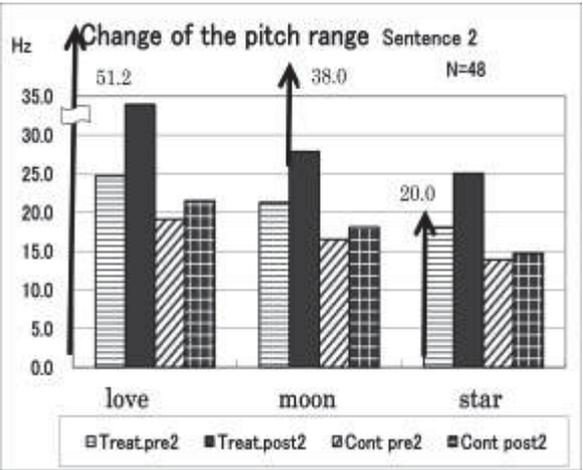


Figure 5-10. Change of mean pitch range in sentence #2.

Although test sentence #2 was not practiced by either group of learners, it is clear from Figure 5-10 that the learners in the treatment group learned to use a broader

pitch range compared with their counterparts in the contrast group. In other words, Japanese EFL learners who practiced with the animated materials were able to successfully apply the English rhythm patterns they had acquired to a new sentence they encountered for the first time.

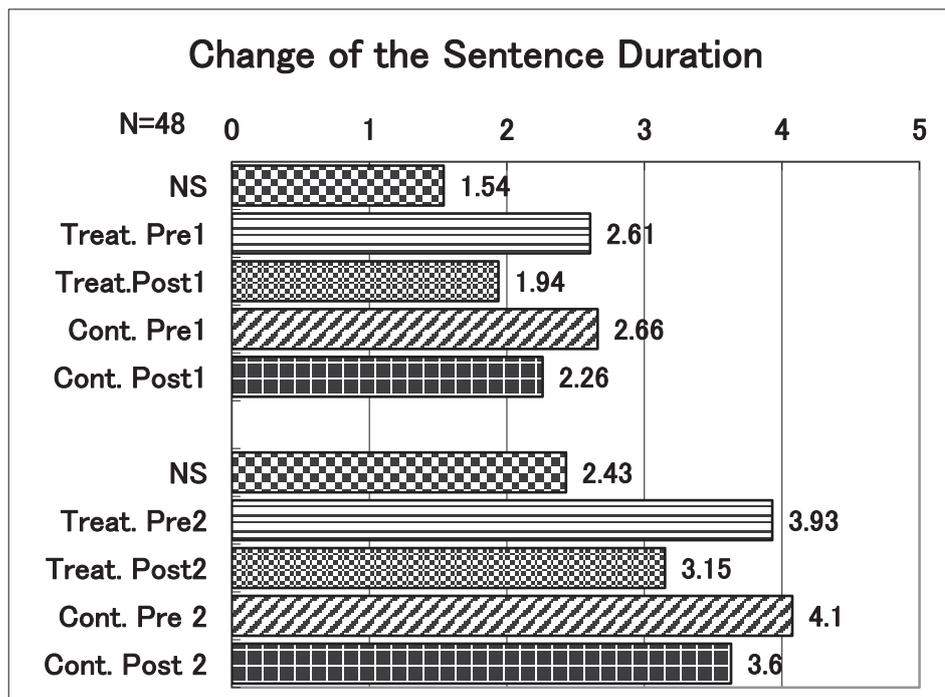


Figure 5-11. Changes of mean sentence duration in milliseconds

Figures 5-9 and 5-10 show the difference in pitch range measured for the two test sentences between pre- and post-practice for the two groups of learners. It is clear that the average difference of pitch ranges between pre- and post-practice for Group A is greater than that of Group B. The mean difference in pitch range in Group A was 10.4 Hz, while for Group B it was 3.5 Hz. This result implies that if learners practice repeating the same sentences just by looking at still letters, with no attention to prosodic features, minimal improvement is expected in the widening of the pitch range.

Figure 5-11 summarizes the mean duration of the two test sentences. The mean duration of sentence #1 is indicated by the four upper sets of bars and for sentence #2,

by the lower sets. The topmost bar in each of the four sets indicates the mean duration of two model speakers. The next two sets show the durational change between pre-test and post-test in the treatment group while the remaining two sets compare the change in the contrast group.

Figure 5-11 shows that the learners in both groups succeeded in shortening their speech duration for both test sentences, indicating that they were able to read the test sentences faster after the five-week practice. We can also see in Figure 5-11 that the learners in Group A made both sentences shorter than those in the contrast group. Mochizuki-Sudo and Kiritani (1991) pointed out difficulties in acquiring the temporal organization of English among native speakers of Japanese since Japanese EFL learners are not accustomed to controlling segmental duration. The result of Experiment 3 implies that the learners in Group A who acquired English rhythm patterns were able to shorten their speech duration by weakening and shortening the unstressed syllables and were thus able to control segmental duration.

5.4.3 Correlations Between Raters' Evaluation and Objective Measurements

Presented in Table 5-5 are the correlation coefficients between the raters' scores and the speech duration. A relatively high negative correlation for the treatment group ($r = -.6809$) and a moderate correlation for the contrast group ($r = -.451$) were found. These results imply that a faster rate of speech correlates with a higher evaluation. The above results suggest that learners in Group A were more highly evaluated by improving their speaking rate than those in Group B.

As suggested in Table 5-5, the correlation coefficients between the raters' score and the number of practice repetitions were as low as 0.0122 for the treatment group and -.0072 for the contrast group, respectively. No correlation was found between the raters' evaluation and the number of repetitions. These results may be interpreted as

showing that the maximum number of repetitions required was sufficient to improve their speech performance.

Table 5-5.

Correlation between Raters' Scores and Objective Measurements

Correlation	Group A	Group B
Raters' scores vs. duration	-.681 *	-.451 *
Rater's scores vs. number of repetition	.012 n.s.	-.007 n.s.

Note. N=48

* $p < .05$

5.5 Questionnaires

A few weeks into the experiment, a questionnaire was given to the learners of both groups to determine which of the viewing conditions, A (animated letters + sound) or B (still letters + sound), was easier to understand. In A, clapping sounds were added to represent the stressed syllables and a mini-sized video clip of a native speaker of English pronouncing sentences was also added to provide learners with better visual clues. As Figure 5-12 illustrates, the combination of animated letters with sound was clearly favored.

After the five-week practice, another questionnaire was given to the learners in the treatment group to ascertain their opinions regarding the animated materials as well as their communicative skills and other affective factors in learning. On the whole, as can be seen in Appendices D and E, the comments were favorable. In the skill category, about 80% of the learners reacted positively to such statements as “*I listened to the English sentence carefully.*” and “*Practice using a PC was fun.*” Seventy percent claimed that they “*came to understand English rhythm*” and the same percentage “*came to understand the difference between stressed words and unstressed words.*” These favorable reactions to the Web practice can be interpreted as indicating that

animated moving letters synchronized with English rhythms motivated the learners to learn English pronunciation.

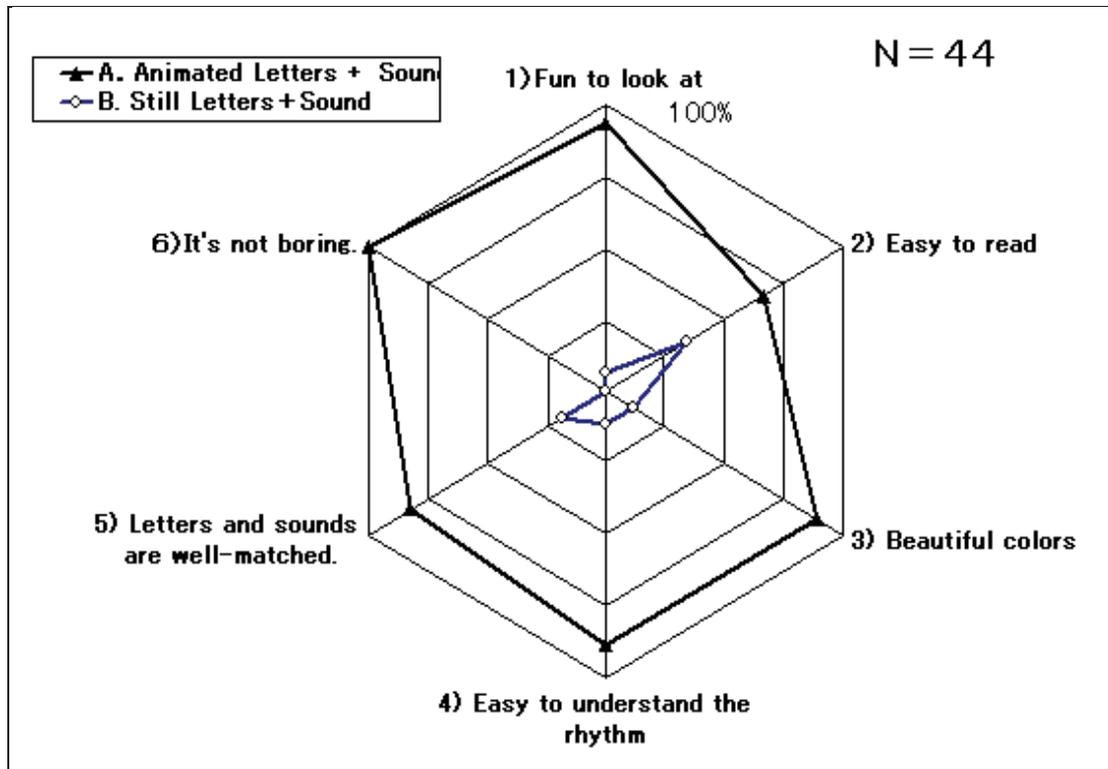


Figure 5-12. Two different modes of pronunciation practice

5.6 Conclusion and Implications

Let me summarize the findings highlighted in this chapter.

(1) The learners' pronunciation in Group A was more highly evaluated by the raters than Group B, showing that the learners in the treatment group made greater progress in pronunciation through rhythmic practice with animated materials than those in the contrast group who practiced with still letters. This can be interpreted as evidence that the animated Web materials helped to improve the pronunciation of the Japanese EFL learners.

(2) It became clear that once good pronunciation habits were acquired

through this rhythmic practice, they were transferable to new English sentences.

(3) The learners in the treatment group learned to use a broader pitch range compared with their counterparts in the contrast group, which lead to higher evaluations by the raters.

(4) The learners who succeeded in producing sentences faster through this rhythmic pronunciation practice were more highly evaluated by the raters.

(5) From the questionnaires administered to the participants, it was clear that those learners who practiced with animated moving letters synchronized with English rhythms were highly motivated to improve their pronunciation of English.

Before concluding this chapter, let us consider the limitations in Experiment 3. First, the number of participants and the length of the experiment were limited. A follow-up study will be necessary to confirm the validity of the above findings with a sufficient number of participants and a longer experiment period. Secondly, for Group A, clapping sounds and a mini-sized video clip of a native speaker of English pronouncing sentences were added, as well as animated letters with sound. It would be desirable to conduct a further controlled experiment to elucidate which visual and audio information worked more effectively in improving the pronunciation of Japanese EFL learners. Thirdly, the number of repetitions of the target sentences in Group A and Group B was the same, i.e. 50 times for each sentence. However, since the experiment was conducted on an individual basis on the Web, it was not possible to strictly control the number of repetitions. And lastly, a longitudinal study is needed to verify whether the good pronunciation habits picked up by the learners were maintained and

transferred to other English sentences they went on to encounter.

In order to prove that these animated materials work effectively in improving the prosodic features of the Japanese EFL learners, it is essential to conduct more empirical experiments not only involving on-campus learners but for off-campus learners as well, via the Internet. At the same time, the author intends to further revise his animated materials to better meet the needs of Japanese EFL learners who wish to improve their pronunciation in general. In this dissertation, the author has so far dealt with prosodic elements such as rhythm and speech duration in detail. The other variable in prosody which may contribute to the naturalness of English is intonation. In Chapter 6, original software aimed at improving learners' intonation will be introduced.

6. A Study of Intonation Practice Using Original Software: Improvement in Pronunciation

6.1 Introduction

Through Chapters 3 to 5, the author's research interest has been mainly directed to learners' prosodic features in English such as rhythm, speech duration, and pitch. In this chapter, however, his research interest shifts to the relationship between learners' intonation and their perceived naturalness of English.

It is well known that intonation, represented as the pitch movement of a group of words, conveys important linguistic information. The pitch of the voice depends mainly on the tension of the vocal folds in the larynx and the rate of vibration, which determines the fundamental frequency (F0) of the resulting speech sound (Ladefoged, 2006). For example, an acceleration in the rate of vibrations is perceived as a rising pitch, and a slowing down as a falling pitch. Variations in intonation patterns are used to convey a speaker's psychological state, the relationship to the listener, and the subtle nuance of what he or she is saying (Dauer, 1993). Intonation helps listeners to identify grammatical structures in speech signaling, for example, whether an utterance is a statement or a question. Intonation also has an attitudinal function, signaling the speaker's surprise, disbelief, sarcasm, teasing and so forth (Underhill, 1994). The five major intonation patterns in English are falling, rising, falling-rising, rising-falling and level tones (Shimizu, 1995). Native speakers of English are able to use and respond to various intonation patterns with few problems. However, for Japanese EFL learners, it is difficult to properly control intonation and to convey the intended meaning through the selection of an appropriate intonation pattern in a given situation (Sugito, 1996).

It is important to note that in cases where the pronunciation of a non-native speaker (NNS) is inaccurate or the grammar or use of vocabulary is improper, a native listener is often able to guess the intended message from the context. Inappropriate use of intonation, on the other hand, can sometimes lead to a breakdown in communication

when a native speaker (NS) is interacting with a NNS (Wells, 2006). Wells (2006) suggests that English makes more elaborate use of intonation to signal meaning than do most other languages. The way of saying something may depend on gestures, facial expression and voice quality, but generally, the most significant factor is intonation (Underhill, 1994). These studies suggest that learning intonation is of extreme importance for EFL learners. Within any given context, an utterance can be given a variety of different meanings depending on the intonation patterns chosen by the speaker. The role of intonation is therefore quite important for mutual understanding.

In order to assist Japanese learners in the acquisition of proper English intonation, an original software application has been newly created to provide the learners with both visual and audio information aimed at improving learners' intonation. The purpose in this chapter is to examine if this software, tentatively named *IntoRec*, actually enables students to improve their English pronunciation. The participants in this experiment were junior high school and university students. After one month's practice, the naturalness of the students' English was evaluated by three native speakers of English and three Japanese English teachers. Acoustic analyses were also conducted using *Praat*, a computer program widely used by linguists and phoneticians to analyze speech sound which can be downloaded from the Web for free. Using this software, the pitch ranges, pitch contours, and the duration of English spoken by Japanese EFL learners were measured. From the statistical analyses, it was clear that practicing English intonation with *IntoRec* helped to improve the students' pronunciation.

6.2 Difficulties that Japanese EFL Learners May Face with Intonation Practice

6.2.1 Perception of Sound

Japanese EFL learners, influenced by the phonetic characteristics of their mother tongue, tend to misperceive a high-pitched syllable as a stressed syllable (Sugito, 1996; Watanabe, 1988). For example, imagine Japanese EFL learners listening to the following dialog:

A: Thank you for inviting me for lunch tomorrow. But I'm afraid I have some work to do until five in the evening.

B: Then, can you come to DINner tomorrow?

Japanese students tend to perceive “tomorrow” as receiving the strongest stress, since it carries the highest pitch in this sentence. However, “dinner” is actually spoken with the heaviest stress, i.e., contrastive stress. This is an example of the kind of negative transfer often observed in the perception of English by non-native listeners. Sugito (1996) suggests that for Japanese EFL learners pitch plays an important role in determining the locus of a stressed syllable, while native speakers are more sensitive in their perception of pitch changes.

6.2.2 Tonetic Notations

Japanese EFL learners may find it difficult to understand various ways of describing intonation notation systems suggested by different phoneticians. Several different descriptions have been used to represent the pitch transition and stress allocation. Six varieties of tonetic notations are summarized in Table 6-1; (1) Pike (1945) and Trager-Smith (1951), for instance, used four different scales, with number 1 indicating a relatively low pitch, 2 a mid pitch, 3 a higher pitch, and 4 an especially

6.2.3 Lack of Proper Teaching Materials

There are currently numerous kinds of pronunciation study materials available, most of which focus on segmentals rather than suprasegmentals. Some books give learners useful insight into intonation, but few identify typical pronunciation problems or give advice as to how to improve them. There is an inevitable limitation to paper-based teaching materials as far as pronunciation instructions are concerned. More practical and reliable devices using multimedia applications are necessary to enhance the learners' motivation to improve their pronunciation.

6.3 Purpose of Chapter 6

The main purposes of this chapter are: (1) to determine to what extent this original software helps students improve their English pronunciation by examining the raters' evaluation and objective measurement using an acoustic analyzer; (2) to explore which phonetic features of English as spoken by Japanese EFL learners contribute to a higher evaluation; and (3) to show how the visual presentation of English intonation contours on the screen contributes to heightening the learners' awareness of correct intonation.

6.3.1 Previous Studies

According to previous studies (Anderson-Hsieh, 1996; Arimoto, 1997; De Bot, 1983; Lambacher, 1997; Shimizu, 1987; Shirai, 2001; Sonobe, Ueda & Yamane, 2009; Sugito, 1996; Yabuuchi & Sato, 2001; Yamane, 1997), providing audio and visual information to Japanese EFL learners is effective in helping them to use a broader pitch range and eventually enabling them to use more native-like pitch patterns. As earlier research revealed, English spoken with a wider pitch range is judged to be more natural than English spoken with monotonous intonation (Dauer, 1993; Shirai, 2001;

Sugito, 1996; Suzuki, 1992; Takefuta, 1990; Yamane, 1999). According to Shirai (2001), showing the pitch contour patterns to learners was more effective than just having them listen to English. These studies suggest that visual information on English intonation might work effectively in improving learners' often monotonous intonation.

It is often suggested that prosody, consisting of rhythm, accent, intonation, and duration, contributes greatly to the production of more natural-sounding English (Sonobe et al., 2009). Chapter 5 demonstrated how original animated materials focusing on English rhythm enabled students to improve their English pronunciation. The author, as a next step, decided to create a software application capable of displaying intonation contours on a PC screen.

6.4 Procedure

6.4.1 Creating an Original Computer Software Application Focusing on English Intonation

In order to help Japanese learners acquire proper English intonation patterns, an original computer software application was developed which shows the visual intonation patterns of a model speaker synchronized with audio output. The model speaker who assisted with this experiment is a native-speaking male in his early fifties from California.

IntoRec was designed to promote autonomous practice without special pronunciation instructions. This software allows learners to practice pronunciation by comparing their own speech with that of the model speakers. This software enables learners not only to record their speech, but also to move waveforms back and forth in order to synchronize the beginning of their speech with the model voice for clear on-screen comparison. The practice sentence patterns include declaratives, negatives, interrogatives, exclamatory sentences, tag-questions and imperatives, with three

sentences for each pattern. Sound files are automatically produced after each practice and can be attached to an email and forwarded to a designated instructor if desired.

6.4.2 Some Features of *IntoRec*

As for the OS environment and a list of computer programs used to create the multimedia application, please refer to Appendix I.

- (1) It is light enough (16 MB) to be used with a USB (Universal Serial Bus) memory stick.
- (2) It enables the learner to practice English intonation by comparing their own pitch patterns and duration with those of NSs on the computer screen without the aid of an instructor.
- (3) Practice can be repeated as many times as desired. The most recent attempt is saved as a sound file which is automatically overwritten with each subsequent practice.
- (4) Any part of the sound wave can be chosen for partial replay.
- (5) The number of recordings is counted automatically.
- (6) Sound files are named automatically.
- (7) The learner's sound file with intonation contours can easily be displayed by pushing the "read file" button.

6.4.3 Participants

Two groups of learners, A and B, both participated as experimental groups in Experiment 4, which will be explained in detail below. There was no control group. In Experiment 3, no significant improvement was observed in the pronunciation of learners in the contrast group, who practiced English pronunciation with no visual support. Based on these results, the author felt there was no necessity for a control

group in Experiment 4. In this experiment, Group A consists of 32 private junior high school students in their third year, who take seven English lessons per week. Participating in Group B were 24 university freshmen, who were not English majors, with ages ranging from 18 to 21. The male and female ratio in each group is shown in Table 6-2. Learners were chosen in different age groups in order to examine the effectiveness of this original software for both junior high school and university students. A listening test, selected from the second grade of the STEP test was conducted on both groups to measure their English abilities, since learners' listening ability is highly correlated with the other three skills (DeMauro, 1992; Thompson, 1996).

Table 6-2.

Basic Data of two groups

Participants		<i>N</i>	Mean Score of the Listening Test (30)
Group A	Junior High School Students	32 (14 Ms, 18 Fs)	18.5 <i>SD</i> = 5.16
Group B	University Students	24 (4 Ms, 20 Fs)	24.2 <i>SD</i> = 5.04
Independent <i>t</i> -test (two-tailed)		<i>df</i> = 54 <i>t</i> -value = 4.1246	**<i>p</i> < .0001

Note. The full mark of the listening test is 30. Ms stands for Males and Fs Females.

Table 6-2 shows that there was a significant difference between the two groups in the mean scores of the listening test. As expected, the university students had higher results.

6.4.4 Practice Procedures

After installing the software, learners must log in by filling out their assigned number and name in the boxes on the computer screen. Learners can choose the sentence they want to practice from among the six different patterns shown on the lower part of the practice screen as indicated in Figure 6-1. With a click of *Hei*

(“Declarative Sentence”), for instance, learners can initiate their practice with one of three declarative sentences. By clicking the Navi Hyoji (“present the model”) button, the intonation contour of the model speaker will appear. The model sentence can be read by clicking the Navi Saisei (“play”) button. Learners can record their reading by clicking the Rokuon (“record”) button, and their intonation contours will be shown by clicking the Hyoji (“present the learner”) button. The waveform and intonation contour of the native speaker will be shown on the upper part as a model and the learner’s waveform and intonation contour will appear on the lower part of the practice screen.

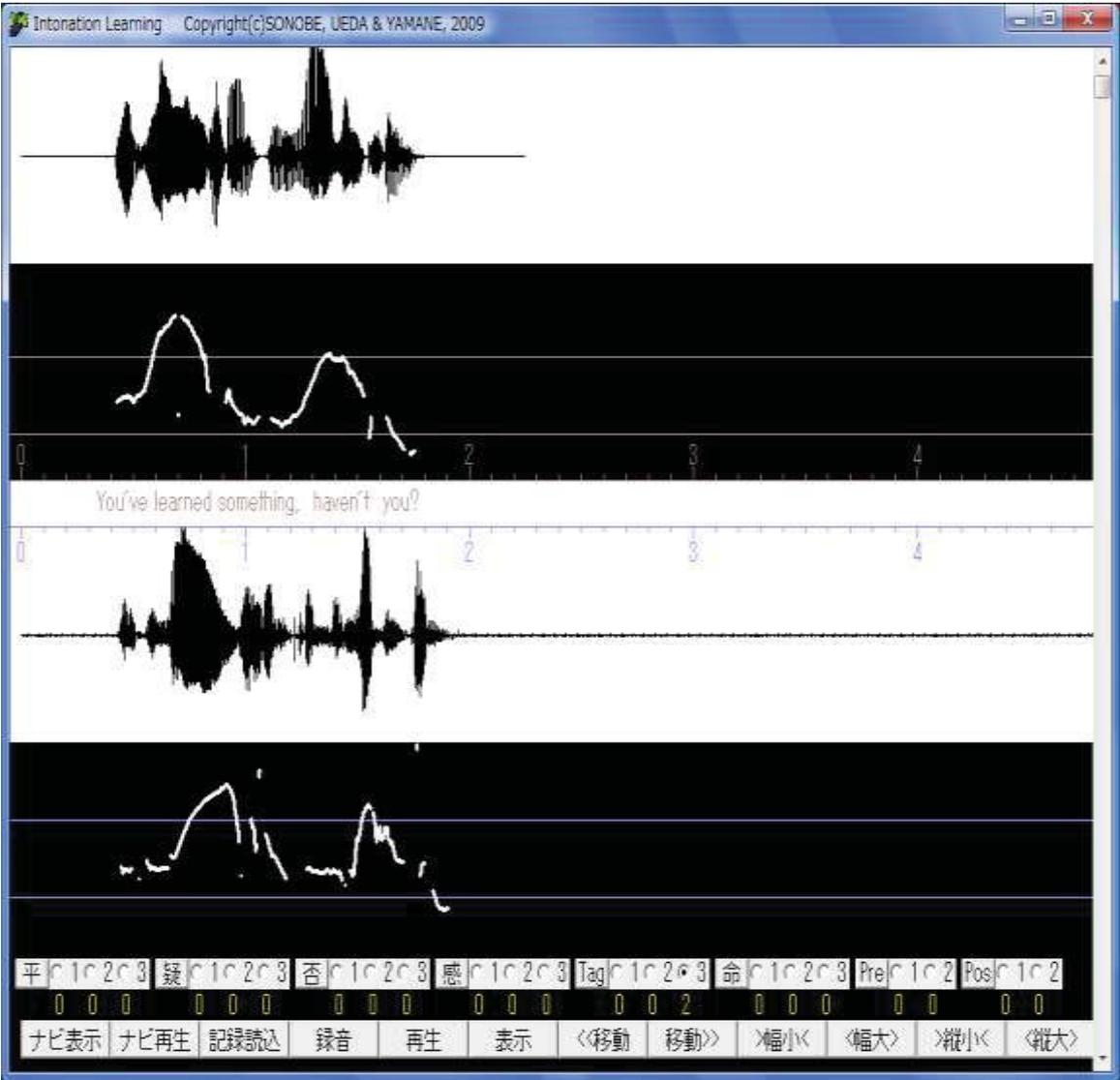


Figure 6-1. Practice Screen of IntoRec

Learners can move the position of their own waveforms back and forth on the screen to synchronize with the initial position of the model speech and facilitate comparison with the model. Amplification and reduction of the learners' wave form is also possible so as to adjust voice input levels. The model intonation contour is displayed in white on the upper part of the screen, and the learner's contour appears below also in white, providing instant feedback. The dual screen shows the pitch movement and the speech duration, considered in that order to be the two important prosodic features for better pronunciation (Makino, 2005; Watanabe, 1994).

Students are expected to note the differences between their own intonation patterns and the native speaker's model by visually comparing a stretch of their own speech with the target model speech. Visual presentation of speech enables learners to be aware of the location of the intonation peak, its bottom, and the duration of a whole sentence. As soon as the learner's pronunciation is recorded, it is digitally saved as a WAV file with a file name plus the ID number initially used to log in. To see a pitch contour from sound file previously saved in the folder merely requires clicking the Rokuon-yomikomi ("read the file") button at the bottom and pushing the Hyoji ("present the learner") and then Saisei ("play the learner") buttons. Learners can access the previous sound files to confirm their progress.

6.5 Experiment 4

The experiment for the junior high school students (Group A) was carried out six times a month as part of a regular class activity. This experiment was conducted using the first 15 minutes of each lesson. As for the university students (Group B), they practiced either at school or at home during their free time. This software is digitally so light that it can be installed on their PCs and used without being logged on to the Internet. The experiment proceeded as follows: Before beginning intonation practice,

all students in both groups were asked to click on the Pre-1 and Pre-2 buttons on the screen to read two test sentences aloud which were recorded and saved as pre-test samples. After recording the two samples, learners in Group A started their intonation practice at school, while Group B worked independently outside of class. After one month, upon completion of the assigned practice, both groups were told to click the Pos-1 and Pos-2 buttons and read the same two test sentences aloud. The sentences were then recorded and saved as post-test samples. In both pre- and post-tests, only the test sentences are shown on the screen and no demonstration readings, model intonation contours, or pronunciation instructions were offered. In other words, they read the sentences aloud with no NS model.

Listed in Appendix G are the 18 sentences prepared for intonation practice: six sets of three sentences, each set having different grammatical structures. The amount of strong stress and sentence duration were controlled in order to make learners pay more attention to the intonation contours. The mean speech rate of all sentences used in this experiment was controlled to be 4.70 syllables per second, taking into consideration the number of syllables and the speech rate to meet the Perceptual Sense Unit standard (Kohno, 2001). The mean duration of the practice sentences was controlled at 1.55 seconds to allow for smooth aural data processing in the learners' short-term memory using the phonological loop (Baddeley, Thomson & Buchanan, 1975). After all participants in both groups practiced the 18 sentences for one month, the post-tests were administered, in which the students were asked to record the same two sentences used in the pre-test. The sound files of the pre- and the post-tests for the university students were collected by the authors via an attached file to an e-mail.

The two test sentences used for the pre- and post-tests were:

- 1) *Could you help me carry these bags?*
- 2) *What an exciting time that was!*

When the learners record the pre- and post-test sentences, no model readings or intonation contours were shown. Prior to their intonation practice, simple tutorial instructions about how to operate this software as well as basic knowledge on acoustic phonetics were provided. To familiarize them with basic acoustics, learners were taught what sound waves meant, as well as how to interpret the intonation contour and speech duration. Their speech samples, with a sampling frequency of 44.1kHz, were used for the acoustic analysis in this study.

Three Americans, two males and one female, and three male Japanese English teachers participated in this experiment as raters to evaluate the naturalness of English produced by the students. One of the male Americans and all three of the Japanese males had been teaching English at Japanese universities. The remaining two Americans have recently moved to Japan, and were less accustomed to English as spoken by the Japanese. The speech samples were arranged in random order and recorded onto CDs for the raters' evaluation. The evaluation was made on a 5-level Likert scale based on the raters' total impression as to the naturalness of English. The inter-rater reliability score among the six raters was Cronbach $\alpha = .8821$, signaling very high reliability. These scores are shown in Table 6-3.

Table 6-3.

Mean Inter-rater Reliability Scores Among Six Raters

		Group A	Group B
Sentence 1	Pre	0.7619	0.8354
	Post	0.8213	0.7705
Sentence 2	Pre	0.8532	0.8909
	Post	0.7675	0.7703

6.6 Results and Discussion

6.6.1 Evaluation by Raters

A paired t-test (two-tailed) was conducted on the mean evaluation scores by the six raters between the pre- and post-tests for the two test sentences. The six raters were asked to evaluate every sound file, so a total of 192 samples were rated for the junior high school students and 144 for university students. Table 6-4 shows the results of the t-test on the mean evaluation scores for test sentences 1 and 2. As for test sentence 1, it was found that the mean scores for the post-test were more highly evaluated than those for the pre-test and the difference was statistically significant ($t = 5.088, p < .01$). When the judges’ evaluations for the pre- and post- tests for sentence 2 were compared, the mean score for the post-test was found to be higher than for the pre-test by a statistically significant margin ($t = 5.931, p < .01$). These results suggest that the junior high school students made improvements in their English pronunciation after practicing with this software for one month.

Table 6-4.
Results of the Paired t-test (two-tailed) Between Pre- and Post- Evaluation of Sentences 1 and 2 (Group A)

Group A	N=32 (18 females)	M	SD	df	t	p	d
sentence 1	pre	3.24	.445	31	5.088	.000**	.724
	post	3.55	.415				
sentence 2	pre	2.94	.516	31	5.931	.000**	.949
	post	3.37	.401				

** $p < .01$

Table 6-5 shows the results of the t-test on the mean evaluation scores in test sentences 1 and 2 for the university students. The mean scores for the post-test were higher than the pre-test, with a statistically significant difference ($t = 8.759, p < .01$). As shown in Table 6-5, in test sentence 2, it was found that the mean score for the post-test was higher than that for the pre-test and the difference was statistically

significant ($t = 8.076, p < .01$). From the results of this experiment, *IntoRec* was found to be effective in improving the pronunciation of learners in both groups with different English proficiency levels.

Table 6-5.

Results of the Paired t-test (two-tailed) Between Pre- and Post- Evaluation of Sentences 1 and 2 (Group B)

Group B	N=24 (19 females)	M	SD	df	t	p	d
sentence 1	pre	3.486	.558	23	8.759	.000**	.911
	post	3.979	.547				
sentence 2	pre	3.347	.478	23	8.076	.000**	1.060
	post	3.812	.432				

** $p < .01$

Tables 6-6 through 6-9 compare the means of the evaluation scores between Pre 1, 2 and Post 1, 2 for the junior high school students and the university students as judged by three native speakers and three Japanese teachers. For both test sentence 1 (interrogative) and 2 (exclamatory), these tables indicate that the mean evaluation scores given to Group A and B by the NSs and Japanese teachers (JTs) increased after the intonation practice. It is interesting to note that the evaluation scores given by the NSs were all higher than those of the JTs in all four cases. The increased rate between the pre- and post-tests as judged by the NSs was more than three times higher than that judged by JTs.

These results suggest two things: first, this intonation practice proved to be effective in improving the pronunciation of not only younger learners whose English proficiency level is still at the beginning stage, but also of more advanced students. Secondly, our results imply that NSs are more sensitive than JTs in evaluating intonation improvement. The American raters tended to evaluate the non-native speech higher when the intonation contour was closer to the model's, as our acoustic analyses in the next section of this study will indicate. This coincides with the results found in earlier studies (Sugito, 1996; Watanabe, 1988).

Table 6-6.

Mean Evaluation Scores of Sentence 1 (Group A)

Rater	NS 1	NS 2	NS 3	JT 1	JT 2	JT 3
Pre-test	2.78	3.47	3.38	2.91	3.38	3.56
Post-test	3.16	3.97	3.72	3.25	3.59	3.63
Increase rates (%)	13.6	14.4	10.0	11.7	6.2	1.9
Increase rates (NS & JT)		12.6			6.6	

Table 6-7.

Mean Evaluation Scores of Sentence 2 (Group A)

Rater	NS 1	NS 2	NS 3	JT 1	JT 2	JT 3
Pre-test	2.47	3.25	3.16	2.63	3.00	3.16
Post-test	2.91	3.69	3.66	3.19	3.34	3.47
Increase rates (%)	17.8	13.5	15.8	21.3	11.3	9.8
Increase rates (NS & JT)		15.7			14.1	

Table 6-8.

Mean Evaluation Scores of Sentence 1 (Group B)

Rater	NS 1	NS 2	NS 3	JT 1	JT 2	JT 3
Pre-test	3.21	3.33	3.38	3.54	3.38	4.08
Post-test	3.67	4.04	4.33	3.71	3.70	4.30
Increase rates (%)	14.3	21.3	28.1	4.8	9.5	5.4
Increase rates (NS & JT)		21.2			6.6	

Table 6-9.

Mean Evaluation Scores of Sentence 2 (Group B)

Rater	NS 1	NS 2	NS 3	JT 1	JT 2	JT 3
Pre-test	2.88	3.33	3.29	3.21	3.29	3.96
Post-test	3.47	4.21	4.08	3.50	3.50	4.21
Increase rates (%)	20.5	26.4	24.0	9.0	6.4	6.3
Increase rates (NS&JT) (%)		23.6			7.2	

6.6.2 Acoustic Analyses

The perceived naturalness of English read aloud by Japanese EFL speakers can be influenced by various phonetic parameters such as the articulatory correctness of segmentals, rhythm, pitch range, intonation, and speech rate. Earlier research suggests that among those factors, pitch range plays the most influential role in influencing the

perceived naturalness of spoken English. English spoken with a wider pitch range is judged to be more natural than English with flat intonation (Dauer, 1993; Shirai, 2001; Sugito, 1996; Suzuki, 1992; Takefuta, 1990; Yamane, 1999).

In order to examine the prosodic features of English as spoken by Japanese EFL learners, the pitch range and the sentence duration were measured using an acoustic analyzer called *Praat*. The pitch range is measured by subtracting the highest F0 values in Hz from the lowest values in speech. The sentence duration is defined as the length of time it took for a speaker to utter a sentence. These two parameters were selected for acoustic analysis, since significant correlations have been suggested between them and the perceived naturalness of spoken English (Sugito, 1996; Yabuuchi & Satoi, 2001; Yamane, Saito, & Yashima, 2004). For the two test sentences, the mean semitone ranges and the mean duration of the test sentences were measured, and their means were compared between pre- and post- pronunciation tests. Both Groups A and B have males and females mixed together. It is known that the average pitch range of females is in general higher than male, for physiological reasons. Thus, the conversion from F0 values to semitone values was necessary to offset sex and individual differences for objective acoustic analyses. At first, accurate F0 values were measured by “Praat” and time-normalized F0 values were expressed as semitones in the logarithmic scale.

Figures 6-2 and 6-3 summarize the change in mean pitch range in test sentences 1 and 2, respectively between the pre- and post-tests for Group A. As you can see from these results, for both test sentences, the maximum value of the semitone became higher and the minimum value, lower. We tend to pay more attention to the maximum semitone values, but a drop in the minimum value was also observed in the post-tests. Over time, the semitone range became broader by as much as 43% for sentence 1 and 42% for sentence 2. Shown in Figures 6-4 and 6-5 are the changes in mean pitch values

expressed in semitones in Group B. It should be pointed out that the junior high school students in Group A were more successful in widening the semitone range for both test sentences than the university students. These results suggest that *IntoRec* worked more effectively for younger students in widening the semitone range.

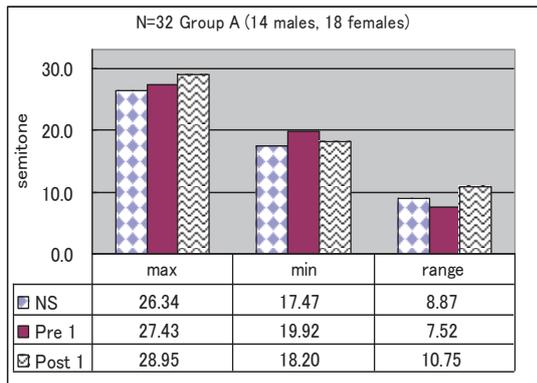


Figure 6-2. Changes of semitone range in Group A (Sentence 1).

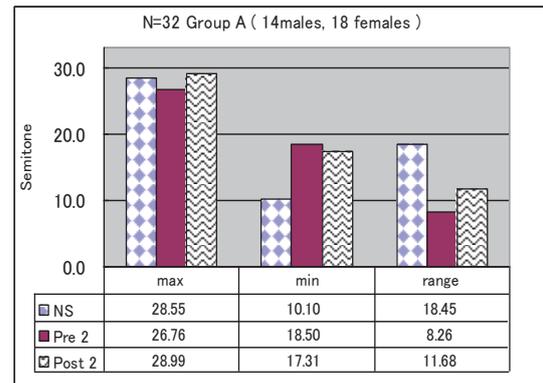


Figure 6-3. Changes of semitone range in Group A (Sentence 2).

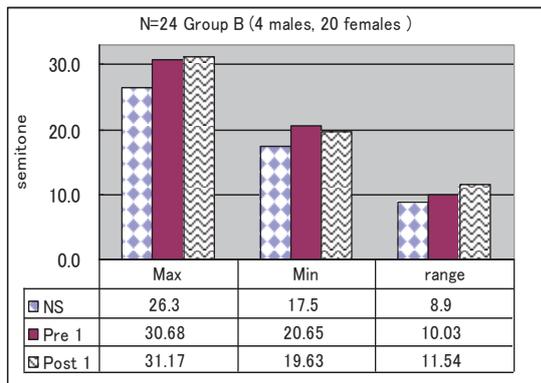


Figure 6-4. Changes of semitone range in Group B (Sentence 1).

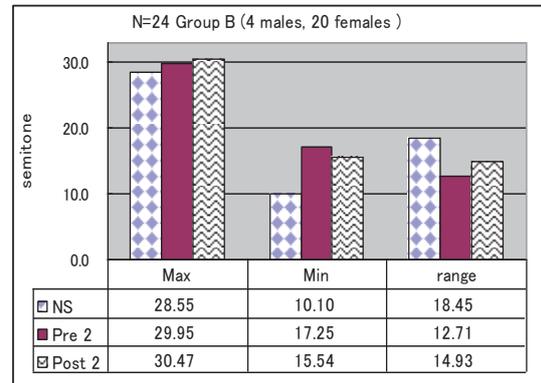


Figure 6-5. Changes of semitone range in Group B (Sentence 2).

A paired *t*-test was performed to assess the significance of the observed difference in means of the semitone ranges between pre- and post-tests for two test sentences in Group A. The results of this analysis, given in Tables 6-10 and 6-11, revealed that the mean semitone ranges for the post-test were higher than for the pre-test with a statistical significance ($t = 4.3521$, $p < .01$) for sentence 1 and ($t = 2.8310$, $p < .01$) for sentence 2. These results suggest that the junior high school

students were successful in widening their semitone ranges for both interrogative and exclamatory sentences after practicing with this software for one month.

Table 6-10.

A Paired t-test (two-tailed) Between Pre- and Post- Semitone Range in Sentences 1 and 2 (Group A)

Group A	N=32 (18 females)	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
sentence 1	pre	7.52	3.384	31	4.3521	.000**	.937
	post	10.75	21.229				
sentence 2	pre	8.26	22.759	31	2.8310	.008**	.689
	post	11.47	22.187				

** $p < .01$

A paired *t*-test was also conducted on the semitone ranges between the pre- and post-tests for group B for two test sentences. As shown in Table 6-11, it was found that the semitone ranges of both sentences 1 and 2 in the post-test were higher than in the pre-test. The difference yielded statistical significance in both sentences ($t = 2.390$, $p < .05$, $t = 3.336$, $p < .01$). These results imply that university students could also broaden their semitone ranges. *IntoRec* was therefore shown to be effective for both beginners of English and more advanced students.

Table 6-11.

A Paired t-test (two-tailed) Between Pre- and Post- Semitone Range in Sentences 1 and 2 (Group B)

Group B	N=24 (19 females)	<i>M</i>	<i>SD</i>	<i>df</i>	<i>t</i>	<i>p</i>	<i>d</i>
sentence 1	pre	10.03	11.613	23	2.390	.025*	.401
	post	11.54	18.207				
sentence 2	pre	12.71	29.368	23	3.336	.003**	.420
	post	14.93	29.294				

* $p < .05$ ** $p < .01$

Stressed words in sentences carrying nuclei are usually pronounced with prominent pitch change. The English spoken by Japanese EFL learners tends to be

judged as monotonous since it lacks these pitch variations. Further acoustic measurements were made on the F0 range of stressed words which carry nuclei in sentences 1 and 2. The nucleus in sentence 1 falls on “bags” and in sentence 2 it falls on “time”. These two words containing stressed syllables should stand out as being more prominent than the other syllables according to the end focus rule. Table 6-12 shows the mean F0 values (Hz) of the highest pitch (maximum F0), the lowest pitch (minimum F0), and the difference between the two (pitch range) in the two prominent words: “bags” in sentence 1 and “time” in sentence 2 in Group A. Table 6-13 lists the same statistics in Group B. As can be seen in Table 6-12, for example, the junior high students successfully added pitch prominence to “bag”, widening its pitch range by 20 percent over time. It is evident from these results that both the junior high and university students succeeded in widening their pitch range and broke away from their monotonous intonation.

Table 6-12.

Mean Changes in F0 Range of “bags” in Sentence 1 and “time” in Sentence 2 (Group A)

N=32	“bag”			“time”			
	max	min	range	max	min	range	
Pre 1 (F0)	209.6	170.5	38.2	Pre 2 (F0)	205.5	186.1	19.5
Post 1 (F0)	209.9	164.2	45.7	Post 2 (F0)	223.3	197.6	25.7
Difference (%)	100	96	120	Difference (%)	109	106	132

Table 6-13.

Mean Changes in F0 Range of “bags” in Sentence 1 and “time” in Sentence 2 (Group B)

N=24	“bag”			“time”			
	max	min	range	max	min	range	
Pre 1 (F0)	255.2	191.4	63.8	Pre 2 (F0)	238.7	186.6	52.1
Post 1 (F0)	273.8	194.6	79.2	Post 2 ((F0)	263.3	198.6	64.7
Difference (%)	107	102	124	Difference (%)	110	106	124

Figure 6-6 below graphically illustrates the intonation contours of 14 junior high school students from Group A for when they read Sentence 1 aloud, prior to beginning their intonation practice. Figure 6-7 shows the intonation contours of the same sentence spoken by the same learners after the 5-week intonation practice with *IntoRec*. The thick line in each figure represents the intonation contour produced by the model speaker. Though the learners were not provided with a model voice when they recorded this test sentence, the contours were added to the figures for the purpose of intonation comparison. Comparing these figures, it can be noted that the monotonous low-pitched intonation typical of Japanese male students improved somewhat after the practice.

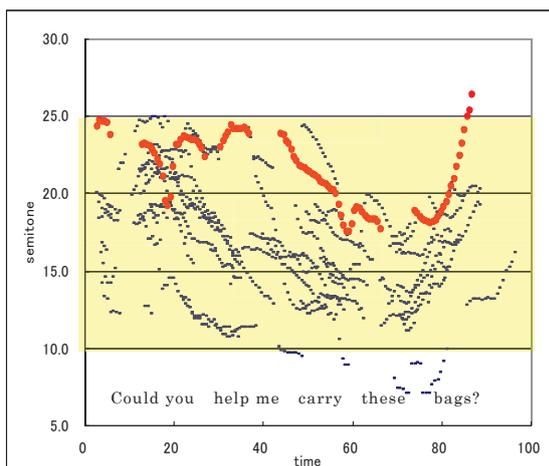


Figure 6-6. Pitch contours of Pre 1 in semitone ($N=14$, Junior High Male Students).

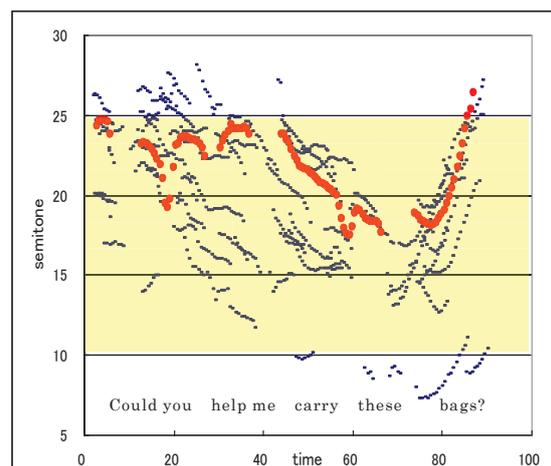


Figure 6-7. Pitch contours of Post 1 in semitone ($N=14$, Junior High Male Students).

6.6.3 Correlations Between Raters' Evaluations and Other Parameters

Table 6-14 shows the correlation coefficients between the means of the raters' scores and the recording frequency. A relatively high correlation was found between the raters' scores and the recording frequency for both junior high and university students. When a correlation coefficient is between $\pm.04$ and $\pm.70$, the correlation is regarded as "relatively high." The recording frequency indicates the number of times

they recorded their speech during the practice period. This result suggests that repetition is an important factor for learners to improve their pronunciation, regardless of their English proficiency level.

Table 6-14.

Correlations between Raters' Evaluation and other Parameters

Group	Sentence	Recording frequency	Correlation	Post Duration (sentence)	Correlation	Pre Duration (sentence)	Correlation
Group A	1	6.9 (times)	0.5221 *	1.84 (sec)	0.064	1.90 (sec)	0.0211
	2	6.7	0.4657 *	1.90	0.272	2.03	0.2755
Group B	1	6.4	0.5204 *	1.96	0.038	1.96	0.1186
	2	6.5	0.4416 *	1.97	-0.056	2.03	0.0093

* $p < .05$

As shown in Table 6-14, no correlations were found between the raters' mean evaluation scores and the learners' speech duration, which suggests that the speech rate did not play a vital role in determining the perceived naturalness of the sentences. This contradicts the results of Experiment 1, 2, and 3, in which a faster speech rate by the learners was evaluated higher. These results are interpreted as implying that even when learners' speech is slow, they may be highly evaluated as long as their intonation is correct.

Figure 6-8 and 6-9 are the results of duration and pitch range comparison between NS and 10 selected good students of Group A for six different sentence patterns. After 20 minute-intonation practice at the first lesson using *IntoRec*, they could manage to control sentence duration though they had difficulty in widening the pitch ranges in interrogative, exclamatory and tag questions.

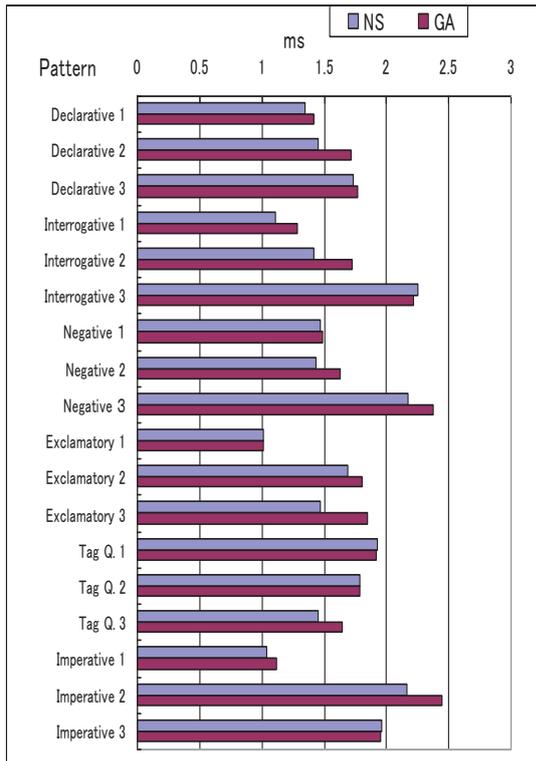


Figure 6-8. Duration comparison between NS and 10 students of GA.

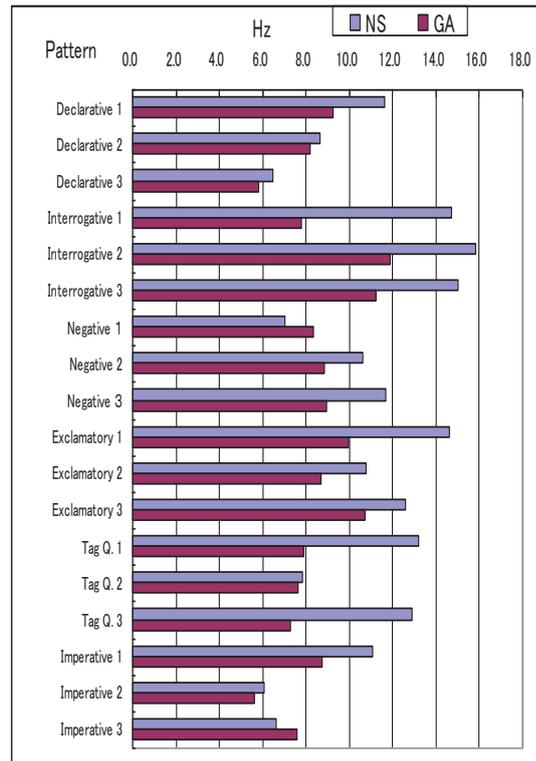


Figure 6-9. Pitch range comparison between NS and 10 students of GA.

6.7 Questionnaires

After the five-week practice, a 16-item questionnaire, created by the author, was administered to 32 junior high students and 8 university students to inquire about the usability of the software, their motivation in learning intonation, and their general impressions of this practice. As Figures 6-10, 6-11, and 6-12 show, most of the questions received favorable replies. Participants gave the highest mean point, 4.5 out of 5, when asked about the ease of software installment (*Q1*). The lowest mean point of 3.2 was given to *Q2*, which concerns the software operations. Items which received more than 4.0 were *Q3*, *6*, *11* and *13*. Those four questions all addressed the visual presentation of the intonation contours. They are: *Q3: The visual intonation contour represented by the white line makes it easier to understand English intonation than when I just listen to a sentence.* *Q6: I tried to make my intonation contour as close to that of the model.* *Q11: I could improve my oral English.* *Q13: I came to be aware of*

the importance of intonation in a sentence as a whole.

These questions were directly related to the description of the intonation contours in color and made us realize that showing the intonation contour played a very important role in the students' acquisition of English intonation, as they were actively trying to make their pitch contours closer to those of NS's.

The results of the questionnaire, shown graphically in Figures 6-10, 6-11, and 6-12, reveal that displaying the intonation contour on the screen played a very important role in generating and heightening the learners' awareness of correct intonation. In other words, the visual presentation of intonation contours enabled them to make good use of their metacognition in intonation. In the classroom we often ask students to listen and repeat model English. However, it is sometimes hard for Japanese EFL learners to reproduce a sentence with correct intonation. Since *IntoRec* allows learners to use a trial-and-error strategy by letting them record their speech as many times as they like and compare the intonation contours, they gradually learn to control their pitch movement to match the model intonation pattern, which was a clear target for them to aim at. The immediate presentation of intonation contours motivated them to adjust their intonation to make it as close as possible to the model speech.

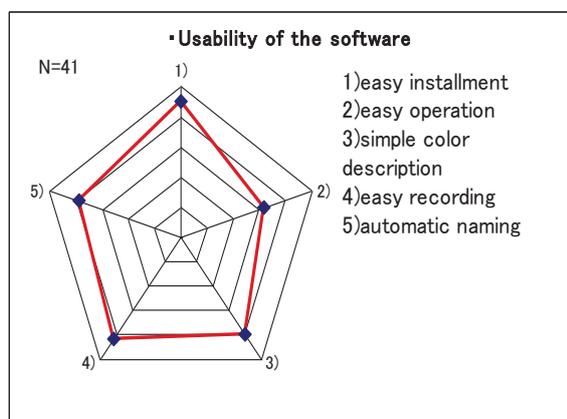


Figure 6-10. Usability of the software

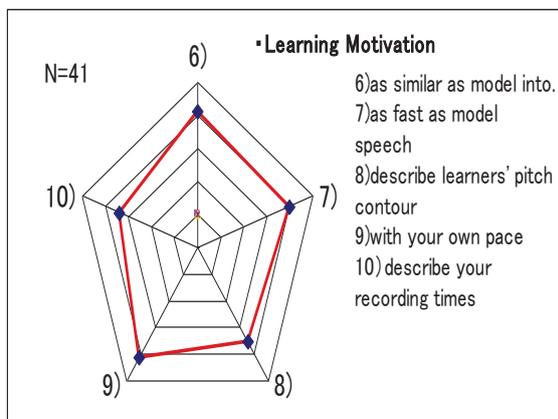


Figure 6-11. Motivation for learning

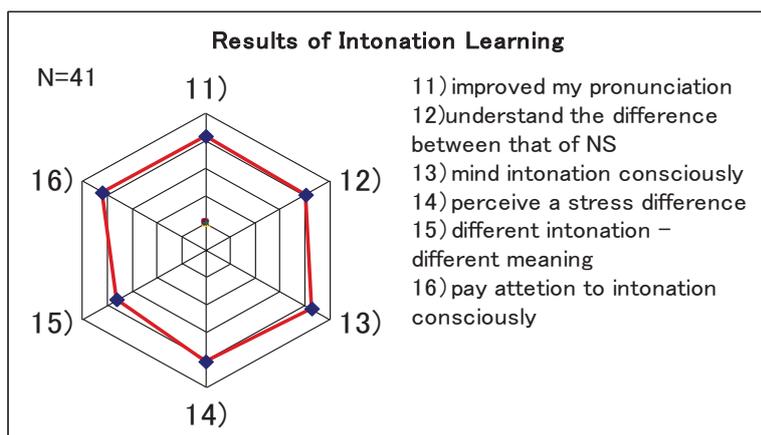


Figure 6-12. Intonation learning

6.8 Conclusion and Implications

Let the author summarize the findings made in this chapter.

(1) After the 5-week practice with 16 sentences, both the junior high school and university students learned to use a broader pitch range than when they began the pronunciation practice and got over their flat intonation.

(2) For both interrogative sentences (sentence 1) and exclamatory sentences (sentence 2), the results of the paired *t*-test revealed that the mean evaluation scores, for the junior high school and university students as rated by the NSs and Japanese teachers (JTs), increased as a result of the intonation practice to a degree that was statistically significant.

(3) After the practice, the raters' evaluations for the university students turned out to be higher than for the junior high school students. This can be ascribed to the fact that the junior high school students were not as good as the university students in pronouncing segmentals, since they are beginners of English.

(4) The increase in the rate of the scores between the pre- and post-tests

as judged by the NSs was more than three times higher than that judged by the JTs. This suggests that the NSs are more sensitive than JTs in evaluating intonation improvement, while JTs are more lenient in rating learners' pronunciation.

(5) A relatively high correlation was detected between the raters' scores and the number of recorded practices for both junior high and university students, which suggests the importance of repeated practice for pronunciation improvement.

(6) As mentioned earlier, in recording both the pre- and post-tests, only the test sentences were shown on the screen and no demonstration readings, model intonation contours, or pronunciation instructions were offered. In other words, the students read aloud the test sentences on their own. These two test sentences were not included in the 16 practice sentences. Despite this, the pronunciation of these sentences was improved, judging from both raters' assessment and acoustic analysis. This implies that repeated intonation practice for five weeks with 16 sentences had a positive influence in improving the learners' intonation overall, and that once good pronunciation habits were acquired through this intonation practice, they were transferable to other English sentences.

(7) *IntoRec* was found to be an effective multimedia learning tool in raising students' awareness of English intonation. Visual presentation of the learners' speech made them aware that their intonation patterns were often different from those of the model speech. This awareness promoted their metacognitive skills and helped learners to take conscious control of their English intonation.

Before concluding this chapter, let the author consider the limitations of the current study. First, the possibility remains that the evaluation of the learners' pronunciation improvement was partially based on phonetic parameters other than intonation. After the five-week practice, the learners' pronunciation of segmentals as well as their English rhythm and fluency may have improved. It would be desirable to conduct a further controlled experiment to elucidate which parameters contribute more to the raters' high assessments. Second, the number of participants and the length of the experiment were limited. A follow-up study will be necessary to confirm the validity of the above findings. And last, a longitudinal study, longer than one month, is needed to verify the author's assertion that good pronunciation habits are transferable to new English sentences the students encounter and whether those habits are sustainable.

In Chapter 7, the author will summarize the major findings made in the previous chapters and discuss "blended learning", combining traditional face-to-face learning with e-learning using multi-media applications, whose positive effects on learners' pronunciation has been verified in previous chapters.

7. Conclusion and Implications

With the development of recent computer technology, learning management systems (LMS) are rapidly gaining ground in the field of higher education. This trend has resulted in a lot of researches dealing with the effectiveness of Web-based learning, e-learning, and learning with multimedia applications. There have also been many studies exploring the phonetic characteristics of English spoken by Japanese EFL learners, as we discussed in Chapter 2. Nevertheless, the author believes that very few studies have shed light on the effectiveness of multimedia applications focused on English prosody and how they contribute to the improvement in the pronunciation of Japanese EFL learners.

This final chapter examines if the four research questions raised at the outset of this dissertation were all sufficiently answered by summarizing the major findings made in the previous chapters. This dissertation will conclude with teaching implications, a list of limitations, and suggestions for possible future development in this field of research.

7.1 Research Questions

The first research question was: *what are the typical prosodic features displayed by Japanese EFL learners?*

(1) As was expected, the speech rate of NNSs is lower than that of NSs. The learners practiced reading short sentences aloud for 30 minutes with no demonstration reading by a NS offered to them. The mean sentence duration produced by these NNSs turned out to be 1.48 times longer than that of NSs (Experiment 1). Slower speech rate is one of the typical prosodic features of non-native speech.

(2) It became evident that Japanese learners are inept at making weak forms of function words short in duration. The duration of unstressed words produced by NSs

accounted for approximately 27.6% of the whole sentence duration, whereas, in non-native speech, it occupied 38.5%. This implies that Japanese EFL learners, who are influenced by the mora-timed rhythm of their mother tongue, tend to produce unstressed syllables with a similar length to stressed ones (Experiment 1).

(3) The duration ratio of the segments in a sentence as spoken by NNNs is approximately the same, whereas uneven durational distribution is a distinctive feature found in readings by native speakers of English (Experiment 2). This is because the learners are incapable of durational control when speaking English.

(4) Learners' pitch range is generally smaller than NSs' (Experiment 3). It was found that Japanese learners are not good at using a wide pitch range in their spoken English.

(5) English intonation produced by Japanese EFL learners' tends to be monotonous, and it is especially low-pitched in the case of Japanese male students (Experiment 4).

The second research question was: *Will the original multi-media application help improve learners' pronunciation?* Two versions of programs focusing on English rhythm with animated letters were created. The first version, used in Experiment 2, had several revisions made to it in order to make it more appealing to the learners by highlighting English rhythm. The revised version was then used in Experiment 3.

(1) The first version used in Experiment 2 with "animated letters with audio" tended to work more effectively in improving learners' English pronunciation than using still letters with audio with a statistical significant difference in scores between the two groups (Experiment 2). The learners who practiced with the revised version were highly evaluated by the raters with a statistically significant difference in assessment scores (Experiment 3), indicating that the learners in the treatment group made greater progress in pronunciation through rhythmic practice with animated materials than those in the contrast group who practiced with still letters.

IntoRec, a multimedia application designed to focus on the improvement of learners' intonation, was found to be effective in improving the pronunciation of students with different English proficiency levels (Experiment 4). On top of that, this intonation practice proved to be effective not only for younger learners whose English proficiency level is still at the beginning stage, but also for more advanced students (Experiment 4).

To sum up, the second research question could be responded to as follows: yes, the three versions of the original multimedia application all improved learners' pronunciation.

The third research question was: *Should it be improved, and in what aspects of pronunciation in particular has it been improved?*

(1) The practice with animated letters proved to be effective in making learners' speech rate faster and closer to the speed of native model speakers (Experiment 2). The learners who used animated letters reduced their sentence duration by 25%, compared with the reduction rate of 13% by learners who used still letters (Experiment 2). The Japanese EFL learners were able to read the test sentences faster after the five-week practice (Experiment 3), since they became capable of shortening their speech duration by weakening and shortening the unstressed syllables and were thus able to control segmental duration.

(2) The learners who used the multimedia teaching material were more successful in widening their pitch range than those who used still letters (Experiments 2, 3).

(3) After the one-month practice with *IntoRec*, the learners became capable of using a wider pitch range in their English, i.e., their semitone range became broader by as much as 42% or 43% (Experiment 4), suggesting that both the junior high and university students succeeded in widening their pitch range and broke away from their

monotonous intonation. *IntoRec* worked more effectively for younger students in widening the semitone range (Experiment 4). The monotonous low-pitched intonation typical of Japanese male students was also improved after the practice (Experiment 4).

The fourth research question was: *which aspect in prosodic improvement influences the evaluation of Japanese learners' speech as assessed by raters?*

(1) The faster the reading speed of the learners, the higher the naturalness of their English is assessed by the raters (Experiment 1). The results of Experiment 1 revealed a high negative correlation between raters' mean evaluation scores and learners' speech duration. The practice with animated letters proved to be effective in making learners' speech rate faster and closer to the speed of native model speakers (Experiment 2). As a result, their spoken English was highly evaluated as they came to speak faster. A faster rate of speech correlates with a higher evaluation (Experiment 3).

(2) No correlations were found between the raters' mean evaluation scores and the learners' speech duration, which suggests that speech rate did not play a vital role in determining the perceived naturalness of the sentences (Experiment 4). This contradicted the results of Experiment 1, 2, and 3, in which learners' faster speech rate was evaluated higher. This can be interpreted as implying that even when learners' speech is slow, they may be highly evaluated as long as their rhythm and intonation is correct (Experiment 4). When we watched an individual student closely, it was found that a learner with a faster speech rate is not necessarily assessed highly if his / her speech uses a narrower pitch range (Experiment 3).

(3) A cluster analysis revealed that, if learners' segment distribution in sentence duration was similar to that of NSs, their speech would be highly assessed (Experiment 2), which implies that learners' speech would be highly evaluated if their English was spoken with stress-timed rhythm.

(4) When learners were successful in widening the pitch range, their speech tended to be assessed highly (Experiment 3).

It is possible to make the following suggestions for the improvement of learners' pronunciation based upon the results of this study.

(1) Learners should be encouraged to pronounce function words short and weak to make their English sound more natural (Experiment 1). A negative correlation between the duration of the whole sentence produced by Japanese EFL learners and the raters' evaluation suggests that, should learners improve the durational control in speaking English by being able to pronounce the weak forms of function words properly, their English would be assessed more highly (Experiment 1). Encouraging learners to pronounce function words short and weak after stressed words is one tactic for them to acquire correct English rhythm.

It was found that Japanese EFL learners tend to produce fairly long duration in unstressed syllables influenced by the mora-timed rhythm of their mother tongue, in which each mora is produced with approximately the same duration. The duration of unstressed syllables should be kept short in order to maintain the isochronism of a stress group (Experiment 1).

(2) If the learners had succeeded in taking advantage of phonetic features such as linking and the weak form, the segment duration would have been made shorter (Experiment 2). A linking should be taught systematically to the students so that they can maintain correct English rhythm.

(3) Once good pronunciation habits were acquired through the rhythmic practice, they would be transferred to other English sentences as well (Experiment 3), in other words, Japanese EFL learners who practiced with the animated materials were able to successfully apply the English rhythm patterns they acquired to a new sentence

encountered for the first time. On the other hand, when learners practiced repeating the same sentences just by looking at still letters with no attention to prosodic features, minimal improvement was made (Experiment 3).

(4) Repetition is an important factor for learners to improve their pronunciation, regardless of their English proficiency level (Experiment 4). A relatively high correlation was found between the raters' scores and the recording frequency for both junior high and university students.

7.2 Pedagogical Implications

7.2.1 Importance of Visual information for Pronunciation Training

The author feels from his experience of teaching English pronunciation to Japanese university students for more than two decades that many of them have serious difficulties in controlling English prosody. The author, who has tried to provide visual information by displaying intonation contours in class, found that after a few attempts at using visual information his students were able to effectively correct their mistakes by themselves. He believes that learning with visual information will reduce students' biased idea that good English pronunciation is difficult to acquire. The learners who participated in the experiments of this study seemed to be fascinated with the visual display of English prosody (Questionnaires in Experiments 3, 4). Using visual information is an excellent way of raising students' awareness of English prosodic characteristics. The learners who actually *look at* model pronunciation will be motivated to improve their prosody. By comparing the visualized image of English prosody displayed by model speakers with their own, learners are able to "notice" the problems of their prosody.

7.2.2 Metacognition

Postulated in Figure 7-1 is a tentative model for illustrating the process of pronunciation improvement in learning with multimedia applications. It was found in this study that the animated letters (Experiments 2, 3), as well as intonation contours (Experiment 4), worked effectively in improving learners’ English prosody. This visual information, which was designed to highlight English prosody such as rhythm, pitch, intonation, and duration, assisted learners to process incoming audio information. They can look at the visualized speech, while simultaneously listening to it. This creates what is called the “modality effect,” which enhances the learner’s smoother process of audio information. In other words, by actually *looking at* the model speech on the screen and by listening to it at the same time, the learners are able to self-monitor the differences between the model speech and their own. This eventually promotes the learners’ self-correction process. As Figure 7-1 graphically explains, an automatic self-correction process fosters autonomous learners.

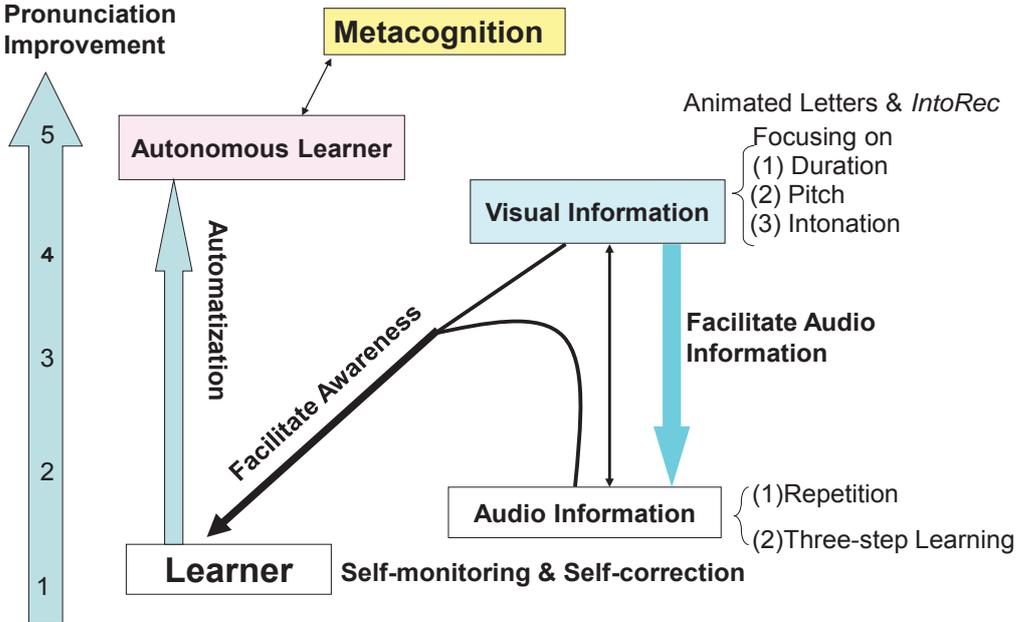


Figure 7-1. Pronunciation Improvement in Metacognition

For autonomous¹ learners, instant feedback on their learning history is indispensable. Yamane (1999, 2006), who examined the effectiveness of visually-aided feedback on the improvement of prosodic features of pronunciation, recommended utilizing speech analyzing software in pronunciation instruction. He contended that this type of speech analyzer can provide learners with instant feedback in the form of precise displays of their pronunciation on a PC screen.

With the traditional “listen and repeat” practice, in which learners are required to listen to model speech and thereby correct their pronunciation errors, it is very difficult for learners to “notice” their problems in pronunciation and correct them by themselves. The results of the questionnaire (Experiments 3, 4) revealed that visualized speech on the screen played a very important role in generating and heightening the learners’ awareness of correct prosody. In other words, the visual presentation of English speech enabled them to make good use of their metacognition in pronunciation.

Metacognitive strategies have functions of planning, monitoring, controlling and evaluating learning (Takeuchi, 2003). Since English prosody can be relatively easily visualized on a PC screen with animated letters or intonation contours, this visual information helped monitor their pronunciation. Oxford (1990) maintains that in order to make metacognitive strategy successful, setting the feasible and visible target is quite important. The author believes that a visualized image of the model prosody became a feasible and visual target of pronunciation for the learners.

7.2.3 Blended Learning

As the author mentioned in Chapter 1, although face-to-face instruction is essentially indispensable for pronunciation learning, various factors prevent sufficient use of this individual learning system in Japan. The following is a summary of the

main reasons cited in Chapter 1; (1) It is difficult to allocate enough time to pronunciation learning in junior and senior high schools in Japan. The first priority is given to lessons preparing for the entrance examination, thus pronunciation practice is downgraded to secondary or even tertiary importance; (2) English teachers are not necessarily keen on teaching pronunciation to students face to face; and (3) As far as pronunciation learning is concerned, there is an inevitable limitation to paper-based teaching materials suited for individual learning. Practical and reliable devices using multimedia applications which provide learners with audio and visual information are necessary to enhance their motivation for pronunciation improvement.

To help solve these problems in pronunciation teaching, the author would like to propose the introduction of “blended learning” to pronunciation teaching, in which traditional face-to-face learning is combined with e-learning using multi-media applications, whose positive effects on learners’ pronunciation were verified in this study.

From the results of evaluations by raters and objective acoustic measurements conducted in experiments 1 to 4 in previous chapters, it became clear that the original teaching materials all worked effectively in improving prosodic features of the Japanese EFL learners. The author would like to use these original software applications to maximize the effectiveness of pronunciation learning and suggests using them under the supervision of an instructor as part of blended learning. Blended learning is defined as a learning method in which traditional face-to-face instruction is combined with computer-mediated instructions (Kay, Johnson, & Hinkelman, 2007; Rink & Yamauchi, 2007).

One barrier to face-to-face instruction is that it is difficult to allocate enough time to pronunciation learning in schools in Japan. Blended learning can help solve this time constraint, which the author cited as an initial problem. Blended learning was

made to expand learning opportunities. Pronunciation learning materials can be available for on-campus students, as well as off-campus students. For instance, the author's students, who come from different parts of Japan, were given opportunities for face-to-face pronunciation training when they come to the campus, while off campus they were able to continue leaning at their own pace. The authors' teaching materials, available on the Internet, will spare teachers a lot of the time that is normally necessary for pronunciation instruction.

The second problem was that English teachers are not necessarily keen on teaching pronunciation to students face to face. The multi-media applications which the author has developed are designed to assist those instructors who are not knowledgeable enough to teach pronunciation with confidence. The application, incorporated with authentic pronunciation with visual information, would be of useful assistance for language teachers. Instructors who are not able to provide learners with enough pronunciation practice can make up their lesson using the Web-based materials.

The third problem was that there is an inevitable limitation to paper-based teaching materials suited for individual learning. The multimedia teaching materials the author created are designed so that learners can establish their learning target easily and on an individual basis. While using these materials, learners are able to literally *look at* their gradual improvement and gain confidence in their pronunciation. Their motivation will be enhanced by effectively taking advantage of audio and visual information provided by the software. Setting a feasible and visible target is important in making learners continue their practice.

The ideal use of blended learning is as a support to classroom-based activities, not as a web-only form of learning. Simply establishing a website for pronunciation learning is like giving texts to students and it does not guarantee effective learning on

the part of the students. Blended learning should be provided to learners as a complement to face-to-face instruction. As Rink and Yamauchi (2007) said, “although blended learning can be a great benefit, mixing traditional instruction with computer-mediated instruction alone cannot guarantee its success - a successful blend is indispensable.” This implies that teachers and learners should understand the role of its function for supporting face-to-face instruction and work together to improve the quality of learning and teaching.

Before concluding this dissertation, the limitations of the current study should be stated. First, the possibility remains that the evaluation of the learners’ pronunciation improvement was partially based on phonetic parameters other than prosody. It is plausible that the learners’ pronunciation of segmentals also improved after the practice with the multimedia application. It would be desirable to conduct a further controlled experiment to elucidate which parameters contribute more to the raters’ high assessments.

Secondly, the number of sentences used for acoustic analysis was limited. Though the learners who participated in this experiment practiced with many different sentences, a fewer number of sentences were acoustically analyzed, i.e., one sentence in Experiment 1 and 2, two sentences in Experiment 3, and two sentences in Experiments 4. In the field of research of acoustic phonetics, it is common that not many original samples are analyzed, since even a small amount of speech samples produce abundant data to examine. Further acoustic analysis is required to endorse the findings made in the current study.

And lastly, the length of the experiment period was limited. The results of Experiment 1 were based on one-shot pronunciation practice. The students continued their practice for five weeks in Experiments 2, 3, and 4. It was found that the learners

became capable of reading new sentences they have never practiced with better pronunciation, i.e., good pronunciation habits are transferable to new English sentences that students go on to encounter. A longitudinal study is needed to verify whether these habits are sustainable. A follow-up study will be necessary to confirm the reliability of the findings made in this study.

One of the author's jobs as an educator is to lead learners so that they can produce better English prosody. In this dissertation, the author wished to demonstrate how his learners improved their pronunciation with the aid of a multimedia application. The author is now working on revising his multimedia application so that learners can practice both English rhythm and intonation simultaneously. In closing, the author hopes that animated materials and *IntoRec* will go on to be used by a whole host of language teachers interested in the blended learning method.

Note

1. Autonomous learners in this dissertation are defined as those learners who can practice English pronunciation by themselves and evaluate their own pronunciation objectively without any constant instructional advice by utilizing multimedia applications.

References

- Abercrombie, D. (1967). *Elements of general phonetics*. NY: Aldine
- Anderson, N.J. (2002). The role of metacognition in second language teaching and learning. Retrieved from October 2, 2003 from <http://www.cal.org/resources/digest/0110anderson.html>
- Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20, 51-62.
- Anderson-Hsieh, J., Johnson, R. & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529-555.
- Arimoto, J. (1997). *Eigo hatsuon kyousei no tactics: Intonation*. [*Tactics of English Pronunciation Stress*], *Sonoda Gakuen Jyoshi Daigaku Ronbunshu*, 32, 87.
- Arimoto, J. (2002). Teaching materials in English pronunciation. *Research Institute for Communication Kansai University of International Studies*, 3, 1-13.
- Arimoto, J., & Kochiyama, M. (2006). *Eigo hatsuon shido heno kokoromi "Eigo hatsuon clinic"*. [*Teaching English Pronunciation*], *Research Institute for Communication Kansai University of International Studies*, 4, 45-47.
- Ashby, M. & Maidment, J. (2005). *Introducing phonetic science*. Cambridge: Cambridge University Press.
- Baddeley, A.D., Thompson, M. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575-589.
- Brondsted, T. (1997). Intonation contours "distorted" by tone patterns of stress groups and word accents. In A. Botinis, G. Kouroupetroglou, & G. Carayiannis, (Eds.). *Intonation: Theory, models and applications* (pp. 11), Athens: Athanasopoulos.
- Celce-Murcia, M., Brinton, D., and Goodwill, J. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. New York: Cambridge University Press.
- Cenoz, J. & Lecumberri, M.L.G. (1999). The acquisition of English pronunciation: Learners' views. *International Journal of Applied Linguistics*, 9, 3-17.
- Collins, B., & Mees, I. M. (2008). *Practical phonetics and phonology: A resource book for students*. London: Routledge.
- Date, T. (2001). *Eigo dorama kara manabu eigo onpo dokuhon*. [Learning phonetics from films and drama], Tokyo: Aoyamasha.

- Dauer, R. M. (1993). *Accurate English*. Englewood Cliffs, NJ: Prentice Hall Regents.
- De Bot, C. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26, 331-350.
- DeMauro, G. (1992). Examination in the relationships among TSE, TWE, and TOEFL scores. *Language Testing*, 9, 149-161.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2008). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359-380.
- Eiken (2010). Unpublished manuscript. Retrieved January 6, 2010, from <http://www.eiken.or.jp/faq/index.html#q37>
- Gimson, A. C. (2008). *Gimson's pronunciation of English*. (7th ed.). Revised by A. Cruttenden. London : Hodder Education.
- Hakuno, M. (2002). *Sakkaku ga koukateki na onseininchi wo kanounisuru [Illusion makes it possible to perceive sounds effectively]*, *Gengo*, 31, 11.
- Hasebe, K. (1989). Using a speech trainer to improve the English of Japanese learners. *Bulletin of Tokai Women's Junior College*, 15, 85-91.
- Higgins, M.D. (1997). Visualising speaking. *Journal of Audio-Visual Education*, 11, 1-8.
- Hill, B. (1999). Video in language learning. *Centre for Information on Language Learning*, 5-6.
- Homma, Y. (1981). Durational relationship between Japanese stops and vowels, *Journal of Phonetics* 9, 273-281.
- Ichizaki, K. (2002). A prosodic study of English inverted sentences: Intonation patterns, nuclei, and rate of utterance. *Miyazaki Women's Junior College Kiyō*, 28, 1-9.
- Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.
- Jenkins, J. (2003). *World Englishes*. London & NY: Routledge.
- Jenkins, J. (2007). *English as a Lingua Franca*. Oxford: Oxford University Press.
- Jones, D. (1960). *An Outline of English phonetics*. Cambridge: Cambridge University Press.
- Kachru, B. (1986). *The alchemy of English: The spread, functions and models of non-native Englishes*. New York: Pergamon Press.
- Kashiwagi, A., Snyder, M, & Craig, J. (2006). Suprasegmentals vs. segmentals: NNS phonological errors leading to actual miscommunication. *JACET Bulletin*, 43, 43-57.
- Kawagoe, I. (1999). *Eigo no onsei wo kagaku suru. [Doing science on the pronunciation of English]*, Tokyo: Taisyukann Shoten.
- Kenworthy, J. (1987). *Teaching English pronunciation*. London: Longman.

- Kohno, M. (2001). *Onseigengo no ninshiki to seisei no mekanizumu: kotoba no jikanseigyokikou to sono yakuwari. [The mechanism of recognition and production of speech: The mechanism of durational control and the role of speech]*. Tokyo: Kinseido.
- Koster, C.J. & Koet, T. (1993). The evaluation of accent in the English of dutchmen. *Language Learning*, 43, 69-92.
- Ladefoged, P. (2006). *A Course in Phonetics*. New York: Harcourt Brace
- Lambacher, S.G. (1997). Developing English pronunciation skills of Japanese learners using computerized instruction. *Journal of the Phonetic Society of Japan*, 1, 45-52.
- Lambacher, S.G. (1999). A CALL tool for improving second language acquisition of English consonants by Japanese learners. *Computer Assisted Language Learning*, 12, 137-156.
- Lehiste, I. (1973). Rhythmic units and syntactic units in production and perception. *Journal of the Acoustical Society of America*, 54, 1228-1234.
- Lightbown, P.M. & Spada, N. (2006). *How languages are learned* (3rd ed.). Oxford: Oxford University Press.
- Makino, T. (2005). *Nihonjin no tameno Eigo onseigaku lesson. [Lessons of English Phonetics for Japanese Learners]*. Tokyo:Taisyukann Shoten.
- McGurk. H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264, 746-748.
- Mochizuki-Sudo, M., Kiritani, S. (1991). Production and perception of stress-related durational patterns in Japanese learners of English. *Journal of phonetics*, 19, 231-248.
- Mori, Y. (2005). The initial high pitch in English sentences produced by Japanese speakers. *English Linguistics*, 22, 23-55.
- Munro, M. J., & Derwing, T.M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45, 73-97.
- Nema, H. & Suzuki, S. (2000). *Hatsuon to rizumu wo masutasuru eigo onseigaku [English Phonetics to master Pronunciation and Rhythm]*. (p.141),Tokyo: Eihosha.
- Nishihara, M. (2003). An acoustic-based analysis of English rhythm. *English Phonetics*, 6, 91-108.
- Oxford, R.L. (1990). *Language learning strategies: What every teacher should know*. NY: Newbury House.
- Pennington, M. (1987). Pronunciation revisited. *TESOL Quarterly*, 20, 207-225.
- Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Joty of*

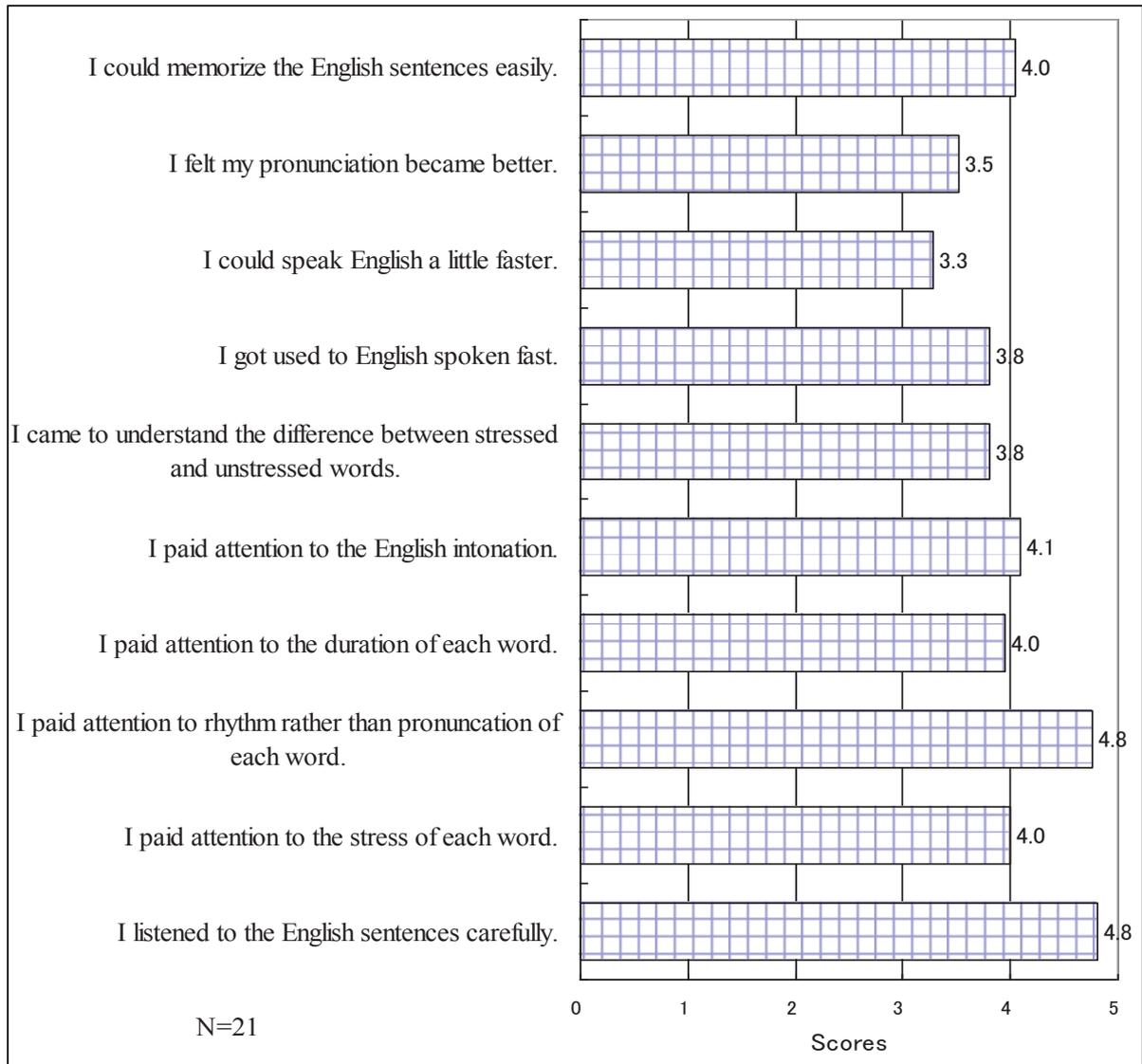
- America*, 32, 6.
- Pimsleur, P., Hancock, C. & Furey, P. (1977). Speech rate and listening comprehension. In M.K. Burt, H.C. Dulay & M. Finocchiaro, (Eds.), *Viewpoints on English as a Second Language* (pp. 27-34). New York: Regents Publishing Co.
- Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, 21, 87-111.
- Richards, R. (2002). The most difficult phoneme for the Japanese learner's of English. *Bulletin of Kyushu Women's University, Humanities and Social Sciences* 38, 3, 27-41.
- Rink, L., & Yamauchi, M. (2007). Integrating blended learning into the language classroom. *Kobe Kaisei College Kiyo*, 46, 195-211.
- Roach, P. (1991). *English Phonetics and Phonology* (p.143). Cambridge: Cambridge University Press.
- Roach, P. (2009). *English phonetics and phonology: A practical course* (4th ed.) Cambridge: Cambridge University Press.
- Shimada, T. (2005). *Eigono jyakuboin ni tuite. [About English weak vowels]. Murokoudai Kiyo* 55, 1-7.
- Shimizu, K. (1987). A study of visual feedback effect on the teaching of English intonation. *The Phonetic Society of Japan, The Bulletin*, 184, 11-15.
- Shimizu, K. (1995). *Eigo onseigaku-ron to gakushu [English Phonetics-Theory and Learning]*. Tokyo: Keisoshobou.
- Shirai, K., Naito, T., Hayakawa, T., Watanabe, R., & Kobayashi, T. (2001). Development of pronunciation learning system based on prosodic information. *The Study of implementing Multimedia in Higher Education*, 83-86.
- Shirai, K., Naito, T., Maeki, D., Kondo, A., & Sagisaka, Y. (2003). Koutou kyouiku kaikaku ni shisuru maruchimedia no koudo riyou ni kansuru kenkyu. [Prosody analysis of English spoken by Japanese learners and development of a system for automatic assessment of pronunciation]. *Kagaku kenkyu kenkyu seika houkokusyo*, 120.
- Sonobe, H. (2006). Prosodic features of speech produced by Japanese EFL learners. *Aichi Sangyo Daigaku Tanki Daigaku Kiyo*, 19, 51-60
- Sonobe, H., Ueda, M., & Yamane, S. (2009). The effects of pronunciation practice with animated materials focusing on English prosody. *Language Education & Technology*, 46, 41-60.
- Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *The Calico Journal*, 10, 19-30.

- Stenson, N., Downing, B., Smith, J., & Smith, K. (1992). The effectiveness of computer-assisted pronunciation training. *The Calico Journal*, 9, 5-19.
- Sudo, M. M. & Kiritani, S. (1999). Effects of acoustics fetures in perception of stress related variants. *Proceedings of the 14th International Congress of Phonetic Sciences*, 1, 113-116.
- Sudo, M. M. & Kaneko, I. (2005). Acquisition process of English rhythmic patterns: Comparison between native speakers of English and Japanese junior high school students. *JACET bulletin*, 40, 1-14.
- Suenobu, M., Kanzaki, K., & Yamane, S. (1992). An experimental study of intelligibility of Japanese English. *International Review of Applied Linguistics in Language Teaching*, 30, 146-156.
- Sugito, M. (1996). *Nipponjin no eigo: Nihongo onseino kenkyu [English spoken by the Japanese]*. Osaka: Izumisyoin.
- Sundai • Benesse Daigaku Nyushi Center Shiken Sokuho. (2008). Unpublished manuscript . Retrieved January 6, 2010, from <http://www.sundai.ac.jp/yobi/dn2008/bunseki/eigohik.htm>
- Suzuki, H. (1992). *Gengo gijutsu toshiteno prosody [Prosody as a linguistic technique]*, *Gengo*, 21, 38-45.
- Suzuki, S. (1990). *Eigo no intoneishon to kontekusuto [English Intonation and Contents]*, *Kokusai Junior College Kiyo*, 5., 65-84.
- Takefuta, Y. & Black, J.W. (1967). A study of foreign accent in Japanese English-Communication. *Communication Monographs*, 34, (2), 167-171.
- Takefuta, Y. (1990). *Nipponjin eigo no kagaku [The Behavioral Science of "Japanese English"]*. Tokyo: Kenkyusha.
- Takeuchi, O. (Ed.). (2000). *Ninchiteki apurochi ni yoru gaikokugo kyouiku [A cognitive approach to foreign language teaching]*. Tokyo: Shohakusha.
- Takeuchi, O. (2003). *Yori yoi gaikokugo gakushuhou wo motomete: Gaikokugo gakushu seikousha no kenkyu [In search of good language learning strategies: Studies on good language learners in the Japanese FL context]*. Tokyo: Shohakusha.
- Teaman, B. D. (2001). Japanese English intonational errors: Features and teaching. *Hiroshima Studies in Language and Language Education*, 4, 67-83.
- Thompson, I. (1996). Assessing foreign language skills: Data from Russian. *The Modern Language Journal*, 80, 47-65.
- Underhill, A. (1994). Sound foundations. *Macmillan Heinemann English Language Teaching*, 74-75.
- Wagener, D. (2006). Promoting independent learning skills using video on digital

- learning laboratories. *Computer Assisted Language Learning*, 19, 279-286.
- Watanabe, K. (1988). Sentence stress perception by Japanese students. *Journal of Phonetics*, 16, 181.
- Watanabe, K. (1994). *Eigo no Rhythm intonation no shido [English Rhythm and Intonation Study]*. Tokyo: Taisyukan Shyoten.
- Wells, J.C. (1999). Which pronunciation do you prefer? *IAFTEFL Issues*, 149, IATEFL
- Wells, J.C. (2006). *English intonation: An introduction*. Cambridge: Cambridge University Press.
- William, K., Gemmell, P., Johnson, A. & Hinkelman, D. (2007). Blended language learning: Using wireless notebooks and a project-based approach. *Sapporo Gakuin University Jinbun Gakkai Kiyo*, 82, 45-79.
- Yabuuchi, S., & Satoi, H. (2001). Prosodic characteristics of Japanese EFL learners' oral reading comparison between good and poor readers. *Language Education & Technology*, 38, 99-112.
- Yamada, R. (2004). Second language speech perception and production learning: Relationship between domains and processing levels. *Technical report of IEICE. Thought and language*, 104 (503), 41-46.
- Yamane, S. (1999). *Eibeijin tono hikaku ni yoru nihonjin eigo no prosodi [Comparison of the prosodic features between English spoken by North American speakers and Japanese speakers]*. *Speech and communication*, 3, 19-28.
- Yamane, S. (2001). *Eigoonseki to komyunikeishon [English phonology and communication]*. Tokyo: Kinseido.
- Yamane, S., Saito, E. & Yashima, T. (2004). *Repeating ga Eigo prosody no shuutoku ni ataeru kouka [Some effects repeating gives on prosody learning.] Kotoba no Kagaku Kenkyu*, 5, 45-51.
- Yamane, S. (2006). *Some characteristics of Japanese EFL learners' utterances and their phonetic Features: Observations from a psycholinguistic perspective*. Unpublished Ph.D. Dissertation submitted to the Graduate School of Foreign Language Education and Research, Kansai University.

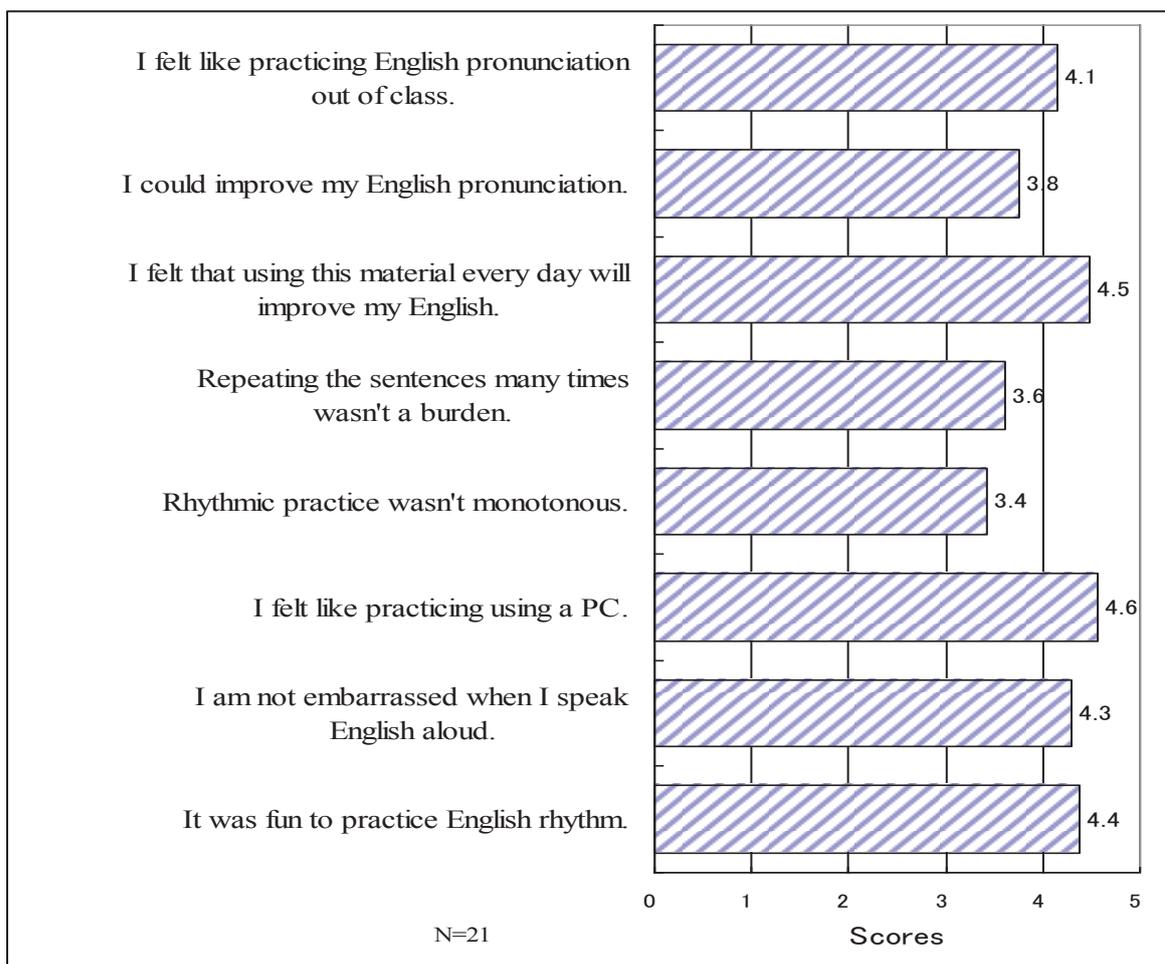
Appendices

Appendix A *Questionnaire: Pronunciation Skills (Chapter 4)*



Note. $N = 21$

Appendix B Questionnaire: On Affective Factors (Chapter 4)



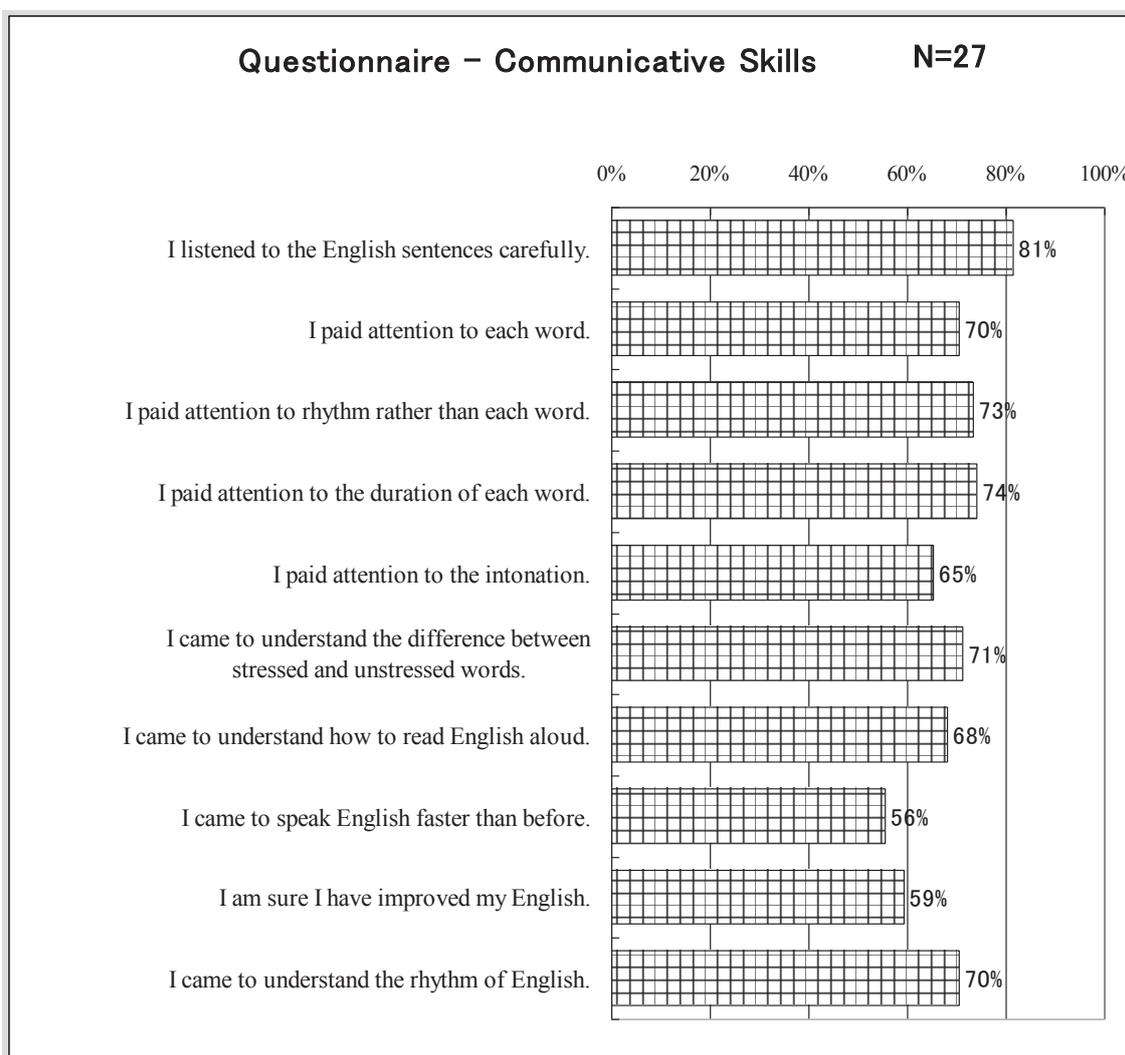
Note. N = 21

Appendix C Sentences for Rhythm Practice (Chapter 4)

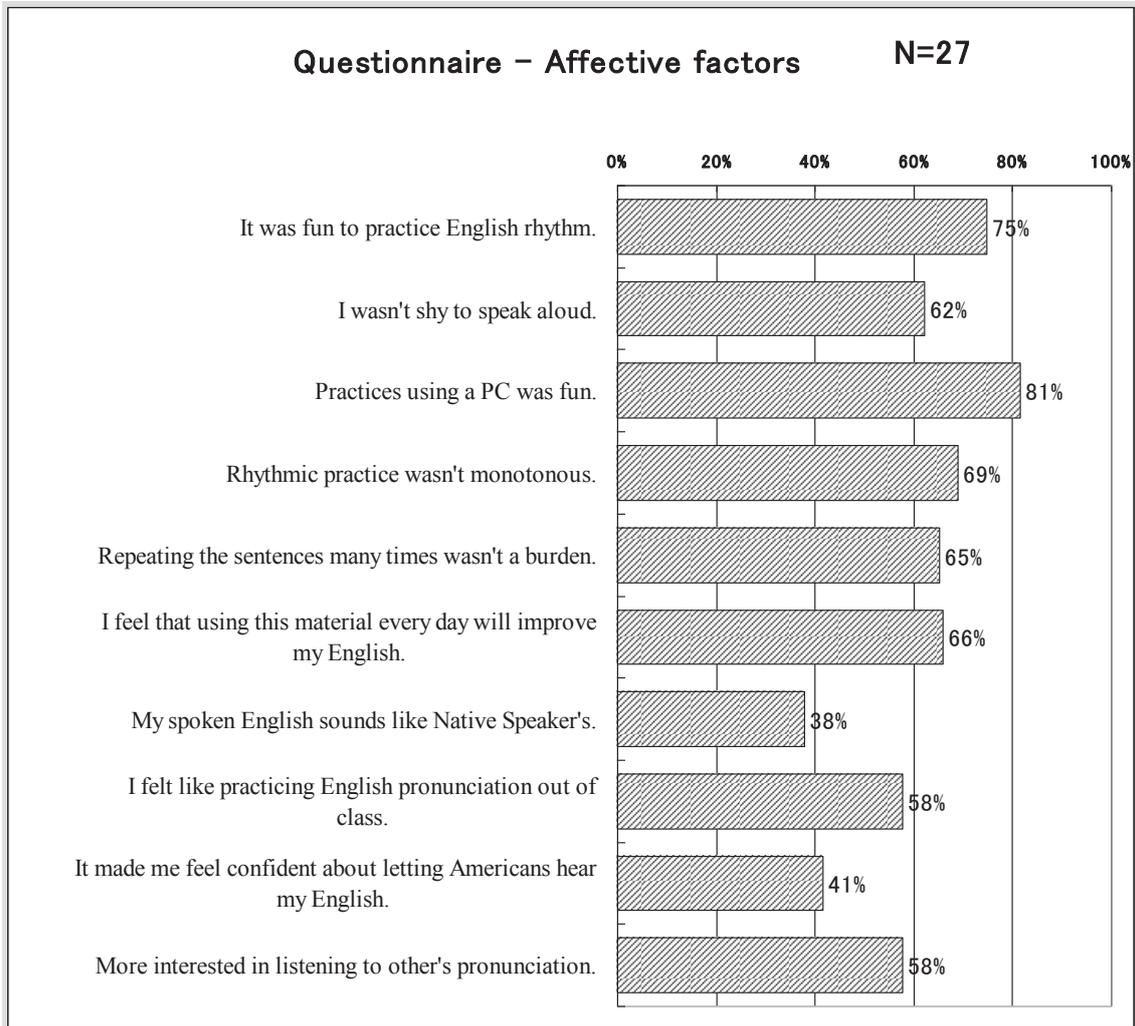
No.	Practice Sentences	number of words	number of stressed syllables	w. p.s.
1	He's <u>trying</u> to <u>study</u> <u>hard</u> .	6	3	4.2
2	I <u>look</u> for <u>roses</u> in the <u>garden</u> .	7	3	4.8
3	Will you <u>make</u> some <u>tea</u> for <u>me</u> ?	7	3	5.5
4	I <u>should</u> have <u>sent</u> her a <u>message</u> .	7	3	4.4
5	There are a <u>lot</u> of <u>schools</u> in our <u>city</u> .	7	3	4.7
6	She <u>wanted</u> to <u>buy</u> a <u>pair</u> of <u>tennis</u> shoes.	9	4	4.8
7	The <u>woman</u> <u>waited</u> for her <u>son</u> at the <u>window</u> .	9	4	4.8
8	The <u>subway</u> in <u>Nagoya</u> is very <u>convenient</u> but I <u>don't</u> like <u>using</u> it.	13	5	4.3

Notes. Based on JACET 8000, the English words used for this practice were from Level 1 (82.9%) and Level 2 (10.5%). w.p.s. = words per second The words with stressed syllables are underlined.

Appendix D Questionnaire: On Communicative Skills (Chapter 5)



Appendix E *Questionnaire: On Affective Factors (Chapter 5)*



Appendix F Sentences for Rhythmic Practice (Chapter 5)

No.	English Sentences	number of words	number of stressed syllables	w. p.s.
1	He's <u>trying</u> to <u>study</u> <u>hard</u> .	6	3	4.2
2	I <u>look</u> for <u>roses</u> in the <u>garden</u> .	7	3	4.8
3	Will you <u>make</u> some <u>tea</u> for <u>me</u> ?	7	3	5.5
4	I <u>need</u> a <u>key</u> for the <u>machine</u> .	7	3	4.4
5	I <u>sold</u> the <u>watch</u> she <u>gave</u> me.	7	3	4.7
6	I'll <u>get</u> you <u>something</u> to <u>drink</u> .	7	3	5.1
7	I <u>should</u> have <u>sent</u> her a <u>message</u> .	7	3	4.5
8	There are a <u>lot</u> of <u>schools</u> in our <u>city</u> .	9	3	4.9
9	She <u>wanted</u> to <u>buy</u> a <u>pair</u> of <u>tennis</u> shoes.	9	4	4.8
10	The <u>woman</u> <u>waited</u> for her <u>son</u> at the <u>window</u> .	9	4	4.8
11	Just <u>think</u> of the <u>play</u> we'll <u>perform</u> on <u>Sunday</u> .	10	4	3.8
12	The <u>subway</u> in <u>Nagoya</u> is very <u>convenient</u> but I <u>don't</u> like <u>using</u> it.	13	5	4.3

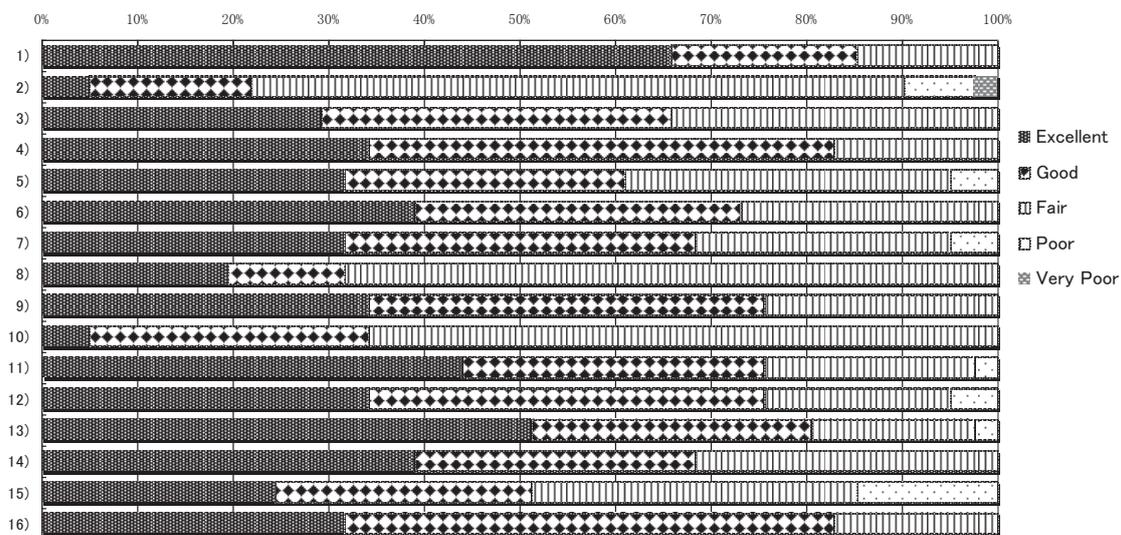
Notes. Based on JACET 8000, the English words used for this practice were from Level 1 (82.9%) and Level 2 (10.5%). Stressed words are underlined. w.p.s = words per second

Appendix G Practice Sentences (Chapter 6)

Sentence Patterns	Speech rate (syll./sec)	Sentence Duration(sec)
Declarative Sentences (Fall)		
1. <i>I'll do it, if I can.</i>	5.04	1.19
2. <i>She's as good as an angel.</i>	4.90	1.43
3. <i>They went for a walk in the park.</i>	5.44	1.47
Interrogative Sentences (Rise, Rise and Fall)		
1. <i>Do you want to come with me?</i>	6.54	1.07
2. <i>Why were you absent yesterday?</i>	4.81	1.66
3. <i>Are you from Hong Kong or Singapore?</i> (Closed Choice)	4.59	1.96
Negative Sentences (Fall)		
1. <i>I didn't say Tuesday.</i>	3.65	1.37
2. <i>I can't stand it any more.</i>	3.80	1.58
3. <i>I don't know how I came to forget it.</i>	5.18	1.93
Exclamatory Sentences (Rise and Fall)		
1. <i>How happy you are!</i>	3.96	1.01
2. <i>What a beautiful garden it is!</i>	5.59	1.61
3. <i>How thoughtful you are to do that!</i>	4.91	1.63
Tag Questions		
1. <i>It's a piece of cake, isn't it?</i> (Declarative, Fall)	5.04	1.39
2. <i>You like to play tennis, don't you?</i> (Question, Rise)	5.06	1.58
3. <i>You've learned something, haven't you?</i> (Declarative, Fall)	4.96	1.41
Imperative Sentences (Fall)		
1. <i>Shut your mouth.</i>	3.06	0.98
2. <i>Study hard, or you'll fail the exam.</i>	3.54	2.26
3. <i>Tell me more what you did yesterday.</i>	4.59	1.96
Test 1. <i>Could you help me carry these bags?</i>	4.71	1.70
Test 2. <i>What an exciting time that was?</i>	4.42	1.81

Appendix H Questionnaires (Chapter 6)

•Usability of this software	
1) I could install this software easily.	
2) I could use this software without a manual.	
3) The visual intonation contour shown in white line makes it easier to understand English intonation than when I just listen to a sentence.	
4) I could record my speech easily.	
5) It was convenient that the sound files were named automatically.	
•Learning motivation	
6) I tried to make my intonation contour as close as possible to that of the model.	
7) I tried to read the sentences as fast as the model.	
8) The immediate display of the intonation contour encouraged me to practice more.	
9) I could practice at my own pace.	
10) I tried to practice as many times as possible.	
•Results of intonation practice	
11) I could improve my oral English.	
12) I came to understand the difference between my intonation and that of the model.	
13) I came to be aware of the importance of intonation in a sentence as a whole.	
14) I came to discriminate between stressed and unstressed words.	
15) I came to understand that different intonation conveys different meanings.	
16) I became to be aware of intonation when listening to English.	



Appendix I List of the Original Software Developed for Each Experiment

Experiment	OS Environment	Software used for Development	Function & Effects of the Software
Experiment 1	Windows XP	Macromedia FLASH 8	Still letters on the web
Experiment 2	Windows XP	Macromedia FLASH 8	1)Color of the letters changed 2)Stressed letters stretched to the right 3)Stressed letters stretched to the right and upward
Experiment 3	Windows XP	Macromedia FLASH 8	4)Count the number of repeats 5) Letters and signs animated by stretching and shrinking 6) With slur signs 7) With clapping sounds 8) With a mini-sized video clip of a native speaker
Experiment 4	Windows Vista	Borland C++Builder	1) Software is light enough (16 M bite) 2) Compare learners' pitch pattern 3) Sound files are automatically saved 4) Partial replay is available 5) Count the number of repeats 6) Sound files named automatically 7) Previous sound files easily displayed with intonation contours

Studies of Improvement in Pronunciation of Japanese EFL Learners by Using Multimedia Applications
SONOBE, Hideyuki