

**A Study on How to Enhance Spoken Word Recognition by
Japanese EFL Learners with Lower Levels of Proficiency**

2017

兵庫教育大学大学院
連合学校教育学研究科
教育実践学専攻
(岡山大学)
米崎 啓和

**A Study on How to Enhance Spoken Word Recognition by
Japanese EFL Learners with Lower Levels of Proficiency**

A Dissertation

Presented to

Hyogo University of Teacher Education

In Partial Fulfillment of the Requirements for the Degree of

Doctor of School Education

By

Hirokazu YONEZAKI

March 2017

TABLE OF CONTENTS

Abstractviii
Acknowledgements xii
List of Tables xiii
List of Figures xvii
List of Abbreviations xix
Chapter 1 Introduction 1
1.1 Background of the Study 1
1.1.1 Importance of Listening 1
1.1.2 Challenging Nature of Spoken Word Recognition 3
1.2 Purpose of the Study 7
1.3 Organization of the Dissertation 7
Chapter 2 The Process of Listening and Spoken Word Recognition 11
2.1 The Process of Listening 11
2.2 Bottom-up and Top-down Processing 15
2.3 Word Recognition and Comprehension for Japanese EFL learners 18
2.4 The Definition of Word Recognition in this Study 21
2.5 Variables and Skills Related to Listening Comprehension 24
2.5.1 Skills Related to Listening 24
2.5.2 Differences between Reading and Listening 27
2.5.3 Speech Rates 33
2.5.3.1 Effects of Rapid Listening 34

2.5.3.2	Effects of Slowing down the Speech on L2 Learners	36
2.5.3.3	Possibility of a Faster Speech Rate as a New Baseline	37
2.6	Significance of Enhancing Spoken Word Recognition for Japanese EFL learners with Lower Proficiency Levels	38
Chapter 3	Spoken Word Recognition and Phonological Features of English	40
3.1	Differences in Phonological Features between English and Japanese	40
3.1.1	Matching Sound with Orthographic Representation	40
3.1.2	The Phonemic System	42
3.1.3	The Syllable Structure	45
3.1.4	Prosodic Features	47
3.2	Segmentation of Continuous Speech and Prosodic Features of the Stress-Timed Language	49
3.2.1	Segmentation of Continuous Speech by L1 Speakers	49
3.2.2	Utilization of Prosodic Information in Segmenting Speech in a Stress-Timed Language	52
3.2.3	A Role of Formulaic Sequences in Speech Perception	54
3.3	Acoustic Challenges of Syllable-Timed L1 Speakers in Perceiving Stress-Timed Language	56
3.3.1	Stress-Timed Rhythm and Speech Rate in English	56
3.3.2	Difference Between Actual Auditory Stimulus and Acoustic Image Created by Japanese EFL learners	59
3.3.3	Recognition of Unstressed Syllables and Function Words in a Stress-Timed Language	63

3.4	Significance of Top-Down Strategies Adopted by Lower-Proficiency Japanese EFL Learners 67
3.4.1	Information from the Bottom-Up and Activation of Top-Down Strategies 67
3.4.2	Prediction and Expectancy Grammar 68
3.5	What is Necessary in Enhancing Spoken Word Recognition by Japanese EFL Learners with Lower Proficiency Levels 72
Chapter 4	Experiment 1: Examining Lower-Proficiency Japanese EFL Learners' Spoken Word Recognition Gap Between Content and Function Words 74
4.1	Introduction 74
4.2	Experiment 75
4.2.1	Purpose 75
4.2.2	Participants 76
4.2.3	Materials 76
4.2.4	Method 77
4.2.5	Procedure 78
4.3	Results 80
4.3.1	Results of the Listening Comprehension Test 80
4.3.2	Results of the Paused Transcription Test 81
4.3.3	Analyses of the Gap in Word Recognition Between the Two Word Categories 86
4.3.4	Correlation Between Listening Proficiency and Word Recognition 88
4.4	Discussion 89
4.5	Further Implications 94

Chapter 5	Experiment 2: Examining Whether Handing out Japanese Translation Beforehand Can Activate Top-Down Strategies 99
5.1	Introduction 99
5.2	Experiment100
5.2.1	Purpose100
5.2.2	Participants101
5.2.3	Pretest and Posttest102
5.2.4	Treatment103
5.2.5	Method of Analyses105
5.3	Results106
5.3.1	Results of ANOVA for Word Recognition in Total106
5.3.2	Results of ANOVA for the Recognition of Content and Function Words109
5.3.3	Results of Chi-Square Tests for Recognition of Each Targeted Word in the Posttest113
5.4	Discussion114
Chapter 6	Experiment 3: Examining Whether Providing Grammatical and Phrasal Knowledge Can Enhance Word Recognition118
6.1	Introduction118
6.2	Experiment119
6.2.1	Purpose119
6.2.2	Participants119
6.2.3	Materials120
6.2.4	Method120

6.2.5 Procedure121
6.2.6 Treatment122
6.3 Results123
6.3.1 The Listening Comprehension Test123
6.3.2 The Paused Transcription Test124
6.3.3 Two-Way ANOVAs for Content and Function Words128
6.3.4 Fisher's Exact Tests130
6.4 Discussion131
6.5 Implications135

Chapter 7 Experiment 4: Examining the Effects of Compressed

Speech Rates on Spoken Word Recognition139
7.1 Introduction139
7.2 Experiment140
7.2.1 Purpose140
7.2.2 Participants141
7.2.3 Materials141
7.2.4 Procedure142
7.3 Results144
7.3.1 Results of Three-Way ANOVA144
7.3.2 Results of Two-Way ANOVAs148
7.3.3 Results of Chi-Square Tests150
7.4 Discussion152
7.5 Implications154

Chapter 8	Experiment 5: Examining the Effects of Explicit Instructions Concerning English Phonological Features on Spoken Word Recognition156
8.1	Introduction156
8.1.1	Learning Habits of Lower-Proficiency Japanese EFL Learners and the Difference in Phonological Features Between English and Japanese157
8.1.2	Perception and Articulation158
8.2	Experiment159
8.2.1	Purpose159
8.2.2	Participants160
8.2.3	Method160
8.2.4	Treatment163
8.3	Results167
8.3.1	Results of the Preliminary Listening Comprehension Test167
8.3.2	Results of the Cloze Tests167
8.3.2.1	Results of a Three-Way ANOVA168
8.3.2.2	Results of Fisher's Exact Tests173
8.3.3	Results of the Paused Transcription Tests174
8.3.3.1	Results of a Three-Way ANOVA174
8.3.3.2	Results of Two-Way ANOVAs for Content and Function Words177
8.3.3.3	Results of Fisher's Exact Tests180
8.4	Discussion181
Chapter 9	Conclusion186
9.1	General Discussion186

9.2	Major Findings of the Study192
9.3	Implications and Limitations of the Study195
9.3.1	Implications for English Education in Japan195
9.3.2	Limitations of the Study197
	References199
	Appendices211

Abstract

The main purpose of this study is to propose effective methods of enhancing spoken word recognition by Japanese EFL learners with lower levels of proficiency. The study first gives theoretical analyses about the listening process and the spoken word recognition. Following this, several experiments were conducted in order to empirically examine what kind of pedagogical methods would be effective in enhancing spoken word recognition by Japanese EFL learners with lower levels of proficiency.

The theoretical study revealed that the listening process consists of three main phases: perception, parsing, and utilization. In addition, when listeners perceive and parse the incoming speech, they utilize bottom-up and top-down processing across all these three phases. In order for listening comprehension to be successful, therefore, both bottom-up and top-down processing must be fully functional.

On the other hand, spoken word recognition is a basic component in listening comprehension, since, unlike in reading, words are not distinctly segmented with spaces. Listeners, therefore, must find by themselves where word boundaries fall and identify words in the continuous speech. Especially in the case of L2 learners, word recognition is not always automatic, and if not, it may well impair comprehension.

Many Japanese EFL learners, especially those with lower levels of proficiency, find it challenging to recognize words in speech, even when they can recognize and understand the same words in the written script. In addition, they are sometimes unable to segment the speech and recognize a word in it which they have no difficulty identifying when the same word is enunciated in isolation.

This is partly due to the difference in phonological features between English and Japanese. Specifically, English stress-timed rhythm and its closed-syllable structure not only brings about a lot of phonetic changes, but also makes the speech quite disproportionate in length with its written version. This causes trouble for Japanese EFL learners, because Japanese is a mora-timed language and is articulated as it is written.

Studies show that the unit for spoken word recognition in English is a stress unit, which contains one stressed syllable with several weak ones. Here, not an individual word but a chunk of words, which form a stress unit such as formulaic sequences, play an important role. Therefore, in order to correctly recognize elusive weak syllables in English speech, it is important to first catch a chunk of words as a whole before segmenting it into individual words.

However, Japanese EFL learners are not accustomed to English natural rhythm as well as natural speech rate, which is one of the greatest variables in listening. Based on these theoretical background, five experiments were conducted in order to search for effective pedagogical methods which would enhance Japanese EFL learners' spoken word recognition.

The first experiment examined whether recognition of function words, which are mostly made up of unstressed syllables, are more demanding than that of content words. The result indicated that function words are more difficult to recognize than content words with speech rate an important variable.

In the second experiment, it was shown that treatment in which Japanese translations were given before dictation practices and instructions were provided to make inferences about the text had positive

effects on spoken word recognition. This might well have resulted from some form of reinforcement on the top-down processing, through application of such strategies as semantic and contextual inferences. In addition, the treatment was no less effective in enhancing the recognition of function words than that of content words.

In the third experiment, it was shown that the treatment of giving learners grammatical and phrasal knowledge had only limited effects on their spoken word recognition. In the case of Japanese EFL learners with lower levels of proficiency, it was only effective on content words for the speech delivered at a moderately slow rate.

In the fourth experiment, learners were provided with treatment in which they listened in class to the material of the textbook at four different compressed speech rates for half a year. The results showed that 1.5 times faster than the normal speech rate had positive effects on their word recognition at the baseline rate. However, effects on recognition of function words were limited.

The fifth experiment focused on the phonological features of English. The treatment involved explicit explanations about English stress-timed rhythm, closed-syllable structure and other phonological features as well as perception and articulation practices using dialogues. In the practice sessions, the participants were asked to stick rigidly to the rhythm and other phonological features proper to English. The results showed that the treatment had been effective for the recognition of both content and function words.

In conclusion, based on these empirical data, the present study gives four major findings concerning the teaching methods to enhance spoken word recognition by Japanese EFL learners with lower levels of proficiency.

First, it would be effective for learners to get accustomed to English phonological features and its stress-timed rhythm through articulation as well as perception practices after explicit explanations. Second, constant exposure to a compressed speech rate of about 67 percent the baseline rate would also be effective. Third, it is important to get listeners to pay more attention to meanings and instruct them to make inferences on the information they perceived. Fourth, phrasal and grammatical knowledge must be effectively complemented by the reinforcement from the bottom-up processing, such as the one related to speech rate or to English phonological and prosodic features, in order to help learners better recognize words in the spoken text.

From these findings, the present study suggest that the use of authentic materials, which fully reflect the English stress-timed nature and other phonological features, not be avoided in the English educational environment.

Acknowledgements

I would like to, first of all, express my deepest gratitude to my supervisor, Professor Shigenobu Takatsuka of Okayama University, for his endless support and encouragement. Without his sincere, constant and insightful advice, this study would not have been completed.

Second, I would like to extend my sincere appreciation to my assistant supervisors, Professor Hiromasa Ohba of Joetsu University of Education and Professor Tatsuhiro Yoshida of Hyogo University of Teacher Education, who guided and assisted me through all stages of the work and provided me with helpful suggestions and warm encouragement.

Third, I am also grateful to Professor Akinobu Tani of Hyogo University of Teacher Education, Associate Professor Naoto Yamamori of Naruto University of Education, and Professor Takafumi Terasawa of Okayama University, who, as examination board members, offered me countless constructive suggestions, which contributed greatly to the completion of this dissertation.

Last but not least, my heartfelt thanks go to the participants of this study, namely the students of National Institute of Technology, Nagaoka College as well as Ritsumeikan Moriyama High School. Without their willing and generous cooperation, the whole study would never have been possible. The work has become an unforgettable experience for me, thanks to their help.

List of Tables

Table 3.1.	Rhythmic Categories and the Units that Work as Cues to Speech Segmentation in Each Language	51
Table 4.1.	The Respective Number of Participants in Each Proficiency Group and the Speech Rate They Listened at	78
Table 4.2.	Sections of Recording Targeted for Transcription	79
Table 4.3.	Descriptive Statistics of the Preliminary Listening Comprehension Test	80
Table 4.4.	Descriptive Statistics of the Paused Transcription Test	81
Table 4.5.	The Results of the Three-Way Mixed ANOVA for the Paused Transcription Test	82
Table 4.6.	Simple Interactions at Each Level of the Three Factors	83
Table 4.7.	Simple Main Effects of Three Factors at Each Level of the Combinations of the Other Two Factors	84
Table 4.8.	The Results of Multiple Comparisons of Proficiency at Each Level of the Speech Rate and the Word Category.....	84
Table 4.9.	The Results of the Two-Way ANOVA for the Word Recognition Gap Between the Two Word Categories	87
Table 4.10.	Simple Main Effects of Proficiency and Speech Rate on Word Recognition Gap Between the Two Word Categories	87
Table 4.11.	Correlations Between Listening Proficiency and Word Recognition for Different Speech Rates	88

Table 5.1.	Descriptive Statistics for the Total Percentage of Correct Word Recognition in the Pretest and the Posttest107
Table 5.2.	The Results of the Two-Way Mixed ANOVA for the Word Recognition in Total107
Table 5.3.	Simple Main Effects of the Group in Both the Tests and Those of the Time for the Three Groups108
Table 5.4.	The Results of Multiple Comparison Between Three Groups in the Posttest108
Table 5.5.	Descriptive Statistics for the Content and Function Word Recognition in the Pretest and the Posttest109
Table 5.6.	The Results of the Two-Way Mixed ANOVA for the Content Word Recognition110
Table 5.7.	The Results of the Two-Way Mixed ANOVA for the Function Word Recognition110
Table 5.8.	Words in Which There Was Significant Difference in Recognition Between the Two Control and One Experimental Groups in the Posttest113
Table 6.1.	Number of Participants in Each Group and the Speech Rates in the Pretest and the Posttest121
Table 6.2.	Sections of Recording Targeted for Transcription in the Posttest122
Table 6.3.	Descriptive Statistics of the Preliminary Listening Comprehension Test124
Table 6.4.	Descriptive Statistics of the Pretest and the Posttest124
Table 6.5.	The Results of the Three-Way ANOVA (Standard Speech Rate)125

Table 6.6.	The Results of the Three-Way ANOVA (Slower Speech Rate)126
Table 6.7.	The Results of the Two-Way ANOVAs128
Table 6.8.	Words in Which There Was Significant Difference in Recognition Between the Experimental and the Control Groups (Standard Speech Rate)131
Table 6.9.	Words in Which There Was Significant Difference in Recognition Between the Experimental and the Control Groups (Slower Speech Rate)131
Table 7.1.	Average Speech Rates in WPM of the Pretest, the Posttest, and Treatment Materials143
Table 7.2.	Descriptive Statistics of the Pretest and the Posttest145
Table 7.3.	The Results of the Three-Way ANOVA145
Table 7.4.	Simple Main Effects in AB Interaction146
Table 7.5.	Simple Main Effects in BC Interaction146
Table 7.6.	The Results of Multiple Comparison Between Four Groups at the Posttest147
Table 7.7.	The Results of Multiple Comparison for Content and Function Word Recognition at the Posttest150
Table 7.8.	Words in Which There Was a Significant Difference in Recognition Between the Control and Three Experimental Groups151
Table 8.1.	Sections of Recording Targeted for Transcription in the Paused Transcription Posttest163
Table 8.2.	Descriptive Statistics of the Preliminary Listening Comprehension Test and the Results of the t-Test167
Table 8.3.	Descriptive Statistics of the Cloze Tests167

Table 8.4.	The Results of the Three-Way ANOVA for the Cloze Tests168
Table 8.5.	Simple Interactions at Each Level of the Three Factors in the Cloze Tests169
Table 8.6.	Simple Main Effects of Three Factors at Each Level of the Combinations of the Other Two Factors in the Cloze Tests169
Table 8.7.	Words in Which There Was Significant Difference in Recognition Between the Control and the Experimental Groups in the Cloze Posttest173
Table 8.8.	Descriptive Statistics of the Paused Transcription Tests174
Table 8.9.	The Results of the Three-Way ANOVA for the Paused Transcription Tests175
Table 8.10.	Simple Main Effects in AB Interaction175
Table 8.11.	Simple Main Effects in AC Interaction176
Table 8.12.	The Results of the Two-Way ANOVA for Content Word Recognition in the Paused Transcription Tests178
Table 8.13.	The Results of the Two-Way ANOVA for Function Word Recognition in the Paused Transcription Tests178
Table 8.14.	Simple Main Effects in the Interaction for Content Word Recognition179
Table 8.15.	Simple Main Effects in the Interaction for Function Word Recognition179

List of Figures

Figure 2.1. Listening model by Rivers 11
Figure 2.2. Listening model adapted from Vandergrift and Goh 13
Figure 2.3. Model of word recognition process by auditory stimuli.....	23
Figure 4.1. Comparisons of the means in percentage of correct content and function word recognition for six groups in the paused transcription test 82
Figure 4.2. Word recognition gap in percentage between content and function words at different speech rates for the three different listening proficiency groups 86
Figure 5.1. Means of total word recognition in percentage for the three groups in the pretest and the posttest107
Figure 5.2. Means of content and function word recognition in percentage for the three groups in the pretest and the posttest110
Figure 6.1. Words successfully recognized in percentage at the standard speech rate124
Figure 6.2. Words successfully recognized in percentage at the slower speech rate125
Figure 6.3. Content and function words successfully recognized in percentage for both standard and slower speech rate groups129
Figure 7.1. Words successfully recognized in percentage for each two of the three factors145
Figure 7.2. Content and function words successfully recognized in percentage at the pretest and the posttest148

Figure 8.1. Words successfully recognized in percentage for each two of the three factors in the cloze tests168
Figure 8.2. Simple interactions between time and word categories in the cloze tests170
Figure 8.3. Simple interactions between groups and word categories in the cloze tests170
Figure 8.4. Simple interactions between groups and time in the cloze tests171
Figure 8.5. Words successfully recognized in percentage for each two of the three factors in the paused transcription tests175
Figure 8.6. Content and function words successfully recognized in percentage at the pretest and the posttest178

List of Abbreviations

ANOVA: analysis of variance

C: consonant

EFL: English as a foreign language

IP: Institutional Program

ITP: Institutional Testing Program

L1: first language; native language

L2: second language; foreign language

MEXT: Ministry of Education, Culture, Sports, Science and Technology

MSS: Metrical Segmentation Strategy

PBT: paper-based test

S: strong syllable

spm: syllables per minute

SVL: Standard Vocabulary List

TOEFL: Test of English as a Foreign Language

TOEIC: Test of English for International Communication

V: vowel

W: weak syllable

wpm: words per minute

Chapter 1

Introduction

This chapter discusses the background and the purpose of the present study as well as the organization of this dissertation. Particularly under discussion is the background of why listening is most important of the four skills and why spoken word recognition is of prime importance in listening. The purpose of the study will then be stated, followed by the organization of this dissertation.

1.1 Background of the Study

1.1.1 Importance of Listening

Of four basic skills, reading, writing, listening, and speaking, there are some reasons to believe why listening is more challenging as well as important than the other three for Japanese EFL learners, especially those with lower levels of proficiency.

In 2009, the new course of study issued by Ministry of Education, Culture, Sports, Science and Technology (MEXT) recommended that all the classes in upper secondary education should, in principle, be conducted in English (MEXT, 2009, p.7). Furthermore, MEXT issued a new implementation plan regarding English education (MEXT, 2013), in which it mentioned the following three things. First, in primary education, focus should be placed on nurturing English communicative competence. Second, in lower secondary level, classes should be conducted primarily in English. Third, in upper secondary level, not only should classes be taught in English, but also students' communicative competence should be brought

to a higher level by using such activities in class as presentation, discussion, and negotiation, thereby enabling them to communicate with a native speaker of English fairly fluently.

In this context, it seems that skills of speaking are regarded as most important. However, as Rivers (1966) mentioned, 'speaking does not of itself constitute communication unless what is being said is comprehended by another person' and 'teaching the comprehension of spoken speeches is therefore of primary importance if the communication aim is to be reached' (p.196).

In addition, in order for speaking skills to be improved, there must be considerable amount of intake to be given (Shirai, 2013), which means that it is very important to give learners sufficient comprehensible input first by listening.

Furthermore, since the new course of study states that English should be taught in English, learners must first understand what teachers say in English. In addition, it would probably take Japanese EFL learners far more time and efforts to understand teachers in class, because they have so far been accustomed to learning English as a written language. Therefore, it is all the more important to improve learners' listening skills.

In communication, the speaker almost always takes the initiative and the listener follows the speaker. In other words, from the speech rate to where to put stresses, the listener has no controllable variables whatsoever over the utterance made between the communicators. Communication ends up in failure, however, if the listener cannot comprehend the message, which in turn leads to inadequate achievement of a goal, proposed by MEXT, that communicative competence should be fully developed in English class. Therefore, to teach how to listen is more important.

When it comes, on the other hand, to the comparison between listening and reading, the former is more challenging than the latter for L2 learners. Although the two skills seem to be alike in that the learner tries to understand the message given to her¹ in written or spoken input, listening is more demanding than reading, because the difference does not end with the one in modality. Since there are no word boundaries in the sound stream, the listener must parse the incoming speech and segment it into words, with acoustic signals she¹ has just perceived held in her limited working memory.

In addition, speech takes place only once. In other words, the linearity of the acoustic signals does not allow listeners to go back along the speech and hear them again (Saussure, 1959; Buck, 2001). Therefore, all these perceiving and parsing must be done very quickly, constantly referring to the listeners' mental lexicon and syntactical knowledge. This would certainly place higher cognitive load on their working memory than in reading.

Furthermore, acoustic signals that listeners hear are often indistinct and ambiguous with speakers modifying the sounds considerably and not all the phonemes clearly encoded (Bond & Garnes, 1980; Buck, 2001; Osada, 2004). Thus, listening involves more complicated processes and variables than reading, and is therefore more challenging. This means that a set of appropriate and focused methods of teaching must be developed.

1.1.2 Challenging Nature of Spoken Word Recognition

Development of listening skills, therefore, is of a primary concern for English teachers in Japan to cope with. In this section, we will focus on the aspect of spoken word recognition.

The process of understanding spoken language can be divided into two parts: recognizing words and understanding their meanings. The first part involves segmenting the undivided continuous speech stream into several chunks of words and the second understanding the speech based on those recognized words (Richards, 1983; Buck, 2001). On the other hand, there are two kinds of processing involved in understanding language. They are bottom-up and top-down processing. The former is a kind of processing in which the listener perceives acoustic signals, then parses them, and finally constructs coherent meanings. In the latter, however, the listener refers to knowledge she already has such as the context, discourse, and pragmatic and prior knowledge, in order to guess the meaning of the signals obtained through the bottom-up processing. These two kinds of processing are happening simultaneously and interactively when the listener tries to comprehend the spoken message (Field, 1999; Buck, 2001; Vandergrift & Goh, 2012).

When the learner perceives an acoustic linguistic input, what is perceived is a mere sequence of sound, if the language is unfamiliar to her (Oller, 1971). The learner may find in it some rhythms or sound pitches, even though she is unable to recognize linguistic signals, still less a meaningful content. Japanese EFL learners with elementary levels of proficiency often say, in answering questionnaires, that they just cannot recognize words, adding that everything sounds like a continuous flow of musical sound. Mitsuhashi (2015) reports the results of a questionnaire in which he asked college students whose first TOEIC test scores after their enrollment into college are below 400 and whose English test scores in the National Center Test for University Admission were 200 or more out of 250 about difficulties of listening in the TOEIC test. Many of them cite the

speech rate and naturalness of pronunciation for reasons why they found it so challenging. One of the most popular comments was, ‘The pronunciation is too “native” for them to understand.’ This illustrates the challenging nature of spoken word recognition for Japanese EFL learners. The gap between spoken and written English is too wide for them to bridge.

This implicates that listeners with lower levels of proficiency have a high hurdle against the first part of listening. They just cannot segment the speech stream into meaningful chunks of words. Lexically non-recognizable, hence incomprehensible, auditory input is nothing but noise (Krashen, 1982). If the listener cannot recognize words, then the listening process does not reach that of a higher level such as recollection of meanings of words, syntactical analyses, and utilization of discourse and prior knowledge, ending in the listener’s failure to construct a mental representation of their understanding of the utterance (Vandergrift & Goh, 2012).

This challenging nature of spoken word recognition results primarily from the facts that the target language is English² and that Japanese EFL learners tend to rely on the written version of the language. Unlike many other languages, English words are not enunciated the way they are written (Narita, 2013).

In addition, the difference in phonological systems between English and Japanese is one of the reasons Japanese EFL listeners are not successful in identifying words in oral texts. English is a stress-timed language in which only stressed syllables are pronounced long and clearly with strength, while the unstressed ones short and quickly, whereas Japanese is a syllable-timed language, in which each syllable, or more accurately each mora³, is articulated evenly stressed, at even intervals and

in the same length (Takei, 2002). This syllable-timed nature of Japanese renders the spoken version of the language quite proportionate in length to the written version (at least in *kana* characters). Japanese EFL learners expect this syllable-timed rhythm in learning English. They are quite embarrassed, therefore, by the fact that English speech is disproportionate in length to the written language, or much shorter than what they see in the written script.

These characteristics of English speech bring about frequent phonetic changes, which is the case more often with unstressed syllables. In order to successfully recognize these weak syllables, listeners must compensate for missing syllables by themselves. It is essential to listen, constantly turning to grammatical and phrasal knowledge as well as the context and background information to fill in the gaps they have failed to bridge. This means utilization of top-down strategies is very important.

In segmenting speech, however, most of the lower-proficiency listeners are said to have more problems with low-level bottom-up processing such as segmentation of the speech, unable to make full use of top-down strategies from grammatical and phrasal knowledge (Goh, 2000; Field, 2003). Considering that there must be a certain threshold level of information that should be picked up through bottom-up processing (O'Malley, Chamot, & Kupper, 1989; Eastman 1993; Vandergrift & Goh, 2012), it might help to give training sessions related to English phonological features in order to shore up the bottom-up processing.

Additionally, considering the fact that shorter duration of English speech than what might be imagined from the written version of the text is what makes listening difficult for Japanese EFL learners, mechanically changing speech rates may affect word recognition. The time required for

the perception of words affects listeners' understanding of the speech and that the rate of speech is a significant variable in the process of listening (Foulke, 1968; Kelch, 1985; Griffiths, 1992). In addition, speech rate is said to be psychologically the most influential factor in listening (Hasan, 2000; Graham, 2006).

Against these backgrounds, this study focused on the ways and effective methods to enhance Japanese EFL learners' spoken word recognition.

1.2 Purpose of the Study

The main purpose of this study is to propose effective methods of enhancing spoken word recognition by Japanese EFL learners with lower levels of proficiency. For this purpose, theoretical background to the listening process and spoken word recognition, from perception of acoustic signals to parsing and to utilization of prior knowledge, will be first reviewed. Following this, both bottom-up and top-down approaches will be empirically examined.

In order to fully develop such skills as lexical segmentation of continuous speech stream and identification of words, both bottom-up and top-down strategies must be effectively taken advantage of. In this study, in search of effective methods of enhancing spoken word recognition, several experiments were conducted.

1.3 Organization of the Dissertation

This study is composed of 9 chapters including this chapter. Chapter 2 deals with the listening process. A mechanism of listening is complicated. After the definition of word recognition, we discuss several phases of the

listening process, from perception of acoustic signals to lexical segmentation to eventual comprehension. Differences between listening and reading as well as relationship between speech rate and comprehension will also be discussed.

Chapter 3 focuses on the challenging nature of English spoken word recognition for Japanese EFL learners. From the perspectives of phonological features of English and learning habits of Japanese EFL learners, we discuss what is challenging about spoken word recognition and why it is so challenging for them. At the end of the chapter, what kinds of teaching methods would be effective will also be implicated and some possible candidates itemized.

In Chapter 4, the effects that the difference in speech rate have on content and function word recognition by learners of different levels of proficiency is empirically examined. There are three explanatory variables: learners' proficiencies, speech rates, and word categories. The criterion variable is correct rates of word recognition. We will also discuss how the listeners with different levels of proficiency adopt bottom-up and top-down strategies.

In Chapter 5, it will be examined whether giving meanings before dictation practices activates top-down strategies and has positive effects on word recognition. In the experiment, only one of the experimental groups was given Japanese translations of the scripts beforehand. The chapter discusses whether the treatment was effective in activating top-down strategies.

In Chapter 6, effects of providing listeners with short-term grammatical and phrasal knowledge on word recognition are examined. The experiment was conducted at two different speech rates and effects of

fortified top-down processing on word recognition will be discussed.

In Chapter 7, we will examine the effects of compressed speech rates on the listeners' word recognition of the baseline speech rate. The participants of the experiment listened to conversations and sentences in a textbook at four different speech rates for half a year. The effects of this treatment on word recognition will be discussed.

In Chapter 8, we empirically examine the effects on word recognition of explicit explanations about English phonological features with some practice sessions of perception and articulation. In the experiment, after listening to a text several times, the participants were instructed, in reading the same text, to adopt and practice "native-like" reading-aloud strategies including linking words, various phonetic changes, and stress-timed rhythms.

Chapter 9 concludes this study. It gives the summary of this study and provides suggestions and implications for English language education in Japan.

Notes

1. This study uses a pronoun 'she/her' to refer to the learner and to the listener.
2. Were the target language to be Spanish, for example, which basically has a combination of consonant and vowel (CV) syllable pattern and allows the learner to articulate as it is written, spoken word recognition might have been less challenging.
3. In Japanese, each mora, rather than each syllable, is evenly stressed and articulated at even intervals for the same length of time. A syllable in Japanese consists of one or two morae. For example, *toyota* is a three-

syllable word which has three morae, while *nissan* is a two-syllable word which has four morae. In Japanese, one *kana* character corresponds to one mora and articulation time depends on the number of morae. Accordingly, in the above example, two-syllable *nissan* is articulated longer than three-syllable *toyota*.

Chapter 2

The Process of Listening and Spoken Word Recognition

This chapter reviews the literature concerning the process of listening: phases from perception of sound to comprehension of the text and bottom-up as well as top-down processing. After reviewing a word recognition model by Japanese EFL learners with lower levels of proficiency, the chapter gives a definition of spoken word recognition in this study. Following this, variables and skills specific to listening comprehension will be discussed.

2.1 The Process of Listening

Even though listening comprehension has held an important place in language teaching, most researches into comprehension has been concerned with reading (Lund, 1991; Osada, 2004; Vandergrift & Goh, 2012). Very few theoretical models, therefore, which elaborate how a cognitive process works in listening comprehension, have been proposed, unlike for reading comprehension.

Rivers (1971) proposes a simple model that will explain how the listener cognitively processes the incoming auditory signals. According to her model, the process of listening consists of three stages: sensing, identification, and rehearsal and recording (Figure 2.1).



Figure 2.1. Listening model by Rivers (1971).

Sensing is a stage of rapid impressions, roughly identified and differentiated, and relatively passive and receptive. At this stage, the listener begins 'rudimentary segmentation' (p. 126) on incoming auditory signals, drawing on her fleeting echoic memory. Rivers says that much of what is heard does not pass on to the second stage because the listener rejects in rapid selection as noise which does not fit in with the initial construction resulting from her familiarity with the phonemic system.

In the next stage of identification, the listener segments and groups what she perceived in the first stage into words as she applies the phonotactic, syntactic, lexical, and collocational rules of the language. This identification stage is active rather than passive, as the listener processes the signal she is receiving sequentially, interrelating the segments already identified with those she is now identifying within the phrasal structure of the utterance. In this way associations are aroused in the listener's information system.

The third stage is rehearsal and recoding of the material, which Rivers says is taking place simultaneously with the other two stages. Rehearsal refers to the recirculating of the material through the cognitive system as the listener makes continuous adjustments and readjustments of her interpretation in view of what has preceded and in anticipation of succeeding segments. The process is entailed by some form of constant anticipatory projection and this adjustive correction takes place every time the utterance does not conform to her expectation.

Finally the process reaches the final part of the stage, recoding. Rivers says that the listener recodes the material of the utterance in a more easily retainable form, in which the basic semantic information will be retained. This recoding, she says, takes place without conscious attention of the

listener, lest she should miss the next part of segmenting and grouping while recoding the previous sections.

A more comprehensive model, which explains how the cognitive processing and processing components are involved in L2 listening, is proposed by Vandergrift and Goh (2012). They developed a model that is based on that of speech production (Levelt, 1989), mirrored by a comprehension processing side. In their model, the listening process involves the following three main phases: perception, parsing, and utilization (Figure 2.2).

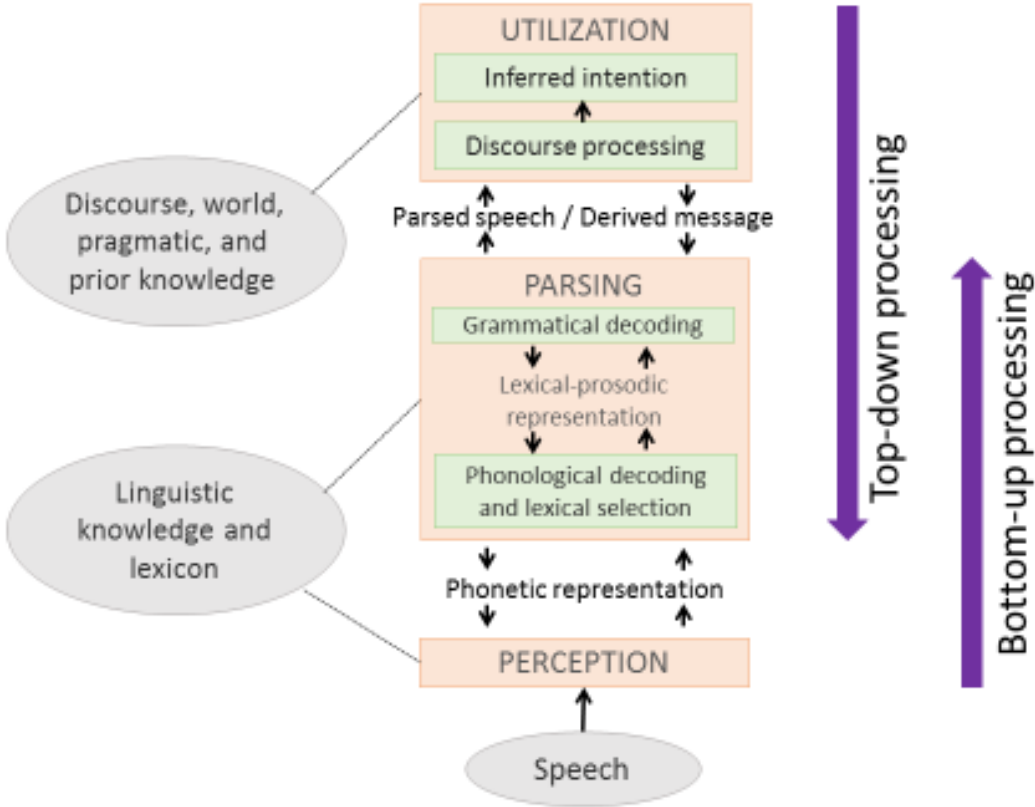


Figure 2.2. Listening model adapted from Vandergrift and Goh (2012).

Perception involves the recognition of incoming sound signals by the listener as words or meaningful chunks of the language. The perception

phase involves bottom-up processing and will depend very much on the listener's L1 and its phonemic system. The degree of perception can also depend on other factors such as the rate of the sound stream. In this stage, the perceived information is active only for a very short period of time so that processing for recognition and meaning must be done almost simultaneously and this information is quickly displaced by other incoming sounds. In addition, the amount of information that can be retained in working memory depends on the listener's language proficiency. Presumably, in the case of learners with lower levels of proficiency, this amount will be quite limited.

Parsing involves the segmentation of an utterance according to syntactic and semantic cues, creating mental representation of the combined meaning of the words. Both bottom-up and top-down processing are involved while the parser attempts to segment the sound stream into meaningful units, through phonological analyses and word retrieval from the mental lexicon. Perception and parsing continue to inform each other until a plausible mental representation emerges. The two processing activities are not linear or happening independently. They are happening at the same time. As the listener perceives new perceptual information, the parser analyzes what remains from what has previously perceived in the listener's working memory.

Utilization, which is top-down in nature, involves creating mental representation of what is retained by the perception and parsing phases and linking this to existing knowledge stored in the listener's long-term memory. During this phase, the meaning derived from the parsed speech is monitored against the context of the message, the listener's prior knowledge, and other relevant information available to the listener in order

to interpret and enrich the meaning of the utterance. The application of prior, pragmatic and discourse knowledge occurs both at a micro level, a sentence or a part of the utterance level, and at a macro level, the level of the whole text or conversation.

2.2 Bottom-up and Top-down Processing

Through these phases, bottom-up and top-down processing are intricately involved and in order for the spoken word recognition to be successful both strategies must be utilized. Bottom-up processing involves segmentation of the sound stream into meaningful units to interpret the message. Listeners segment the sound stream and construct meaning by accretion, using their knowledge of individual sounds or phonemes as well as of patterns of language intonation such as stress, tone, and rhythm (Vandergrift & Goh, 2012).

Field (1999) describes the bottom-up processing as assembling step by step perceptual information until it reaches some coherent meanings. The listener combines groups of acoustic features into phonemes, phonemes into syllables, and syllables into words.

According to Buck (2001), the bottom-up view of language processing is that of starting from the lowest level of detail and moving up to the highest level. Acoustic signals are first decoded into phonemes, which are used to identify individual words. Then the processing continues on to the next higher stage, the syntactic level, followed by an analysis of the semantic content to arrive at a literal understanding of the basic linguistic meaning. Finally, the listener interprets that literal meaning in terms of the communicative situation to understand what the speaker means.

As both research and daily experience indicate, however, it is evident

that the listening process does not occur solely through picking up acoustic signals or in a linear sequence from the lowest to the highest level, but different types of processing may occur simultaneously (Buck, 2001). The processing must involve utilization of information provided by context, the listener's prior or pragmatic knowledge, which is called the top-down processing (Field, 1999; Rost, 2002; Vandergrift & Goh, 2012). The top-down process is rather complicated because the listener must take advantage of various sources of information: knowledge of the world, analogy with a previous situation, or the meaning that has been built up so far (Bond & Garnes, 1980; Field, 1999). It can also be derived from a schema or expectation set up before listening. In addition, as far as contextual information is concerned, it can be invoked before, during and after the perception of auditory signals (Field, 1999). If invoked before the perception, it helps the listener anticipate or predict the incoming words. At other times, these kinds of information will only be available during the perceptual process and, at still other times, it is employed only after the identification of words.

Therefore, it is possible to understand the meaning of a word before decoding its sound thanks to the listener's various knowledge (Buck, 2001). The listener typically has some expectations about what she will hear or has some hypotheses about what is likely to come next. In other words, context helps reduce the number of lexical possibilities and hence enhance word recognition, which is prerequisite in listening comprehension (Grosjean, 1980). For example, in an uncompleted sentence, '*She was so frustrated and angry that she picked up the gun, aimed and ...*' (adapted from Grosjean, 1980), the listener can fill the blank, given very little acoustic information, with a word such as '*fired*' or '*shot*' (Buck, 2001). As

the listener processes the incoming speech, she can naturally expect the following word and all that has to be done is to listen to the sound and confirm the expectation or sometimes she does not have to listen to the last word. In the above example, the listener's background knowledge about guns and possible behaviors by angry people would be enough to predict the word.

Buck (2001) suggests that listening comprehension is a top-down process in the sense that various kinds of knowledge helps the listener understand what the speaker means, even though the knowledge does not applied in any fixed order. These types of knowledge can be used in any order and simultaneously. Where bottom-up decoding fails, top-down strategies can be called in to compensate (Rost, 2002). Nevertheless, the acoustic input, information from the bottom-up process, is no less important, because top-down strategies are nowhere to be applied without any lexical information from the bottom-up process. Accordingly, listening process is an interactive one in which the listener turns to a number of information sources, including acoustic input, different types of linguistic knowledge, details of the context, other related, general, or pragmatic knowledge, and whatever information sources she has available.

However, in L2 listening, utilization of top-down processing, expectations and predictions from the context and general knowledge does not necessarily occur, especially in the case of lower-proficiency listeners (Goh, 2000; Vandergrift & Goh, 2012). The ability to activate various types of knowledge during listening comprehension depends on listeners' language proficiency. Lower-proficiency listeners have greater difficulty processing both contextual and linguistic information, and, therefore, are less able to simultaneously make use of both bottom-up and top-down

strategies.

2.3 Word Recognition and Comprehension for Japanese EFL learners

As has been discussed in the previous sections, there are three phases and two kinds of processing involved in listening comprehension. In L2 listening, however, word recognition is not necessarily automatic (Rost, 2002), and therefore not 'given' unlike in reading. This is the case, especially when listeners' proficiency is lower (Goh, 2000). Naturally, unsuccessful word recognition leads to unsatisfactory comprehension, even though spoken language comprehension can occasionally continue successfully with some words unrecognized (Rost, 2002), only if the listener can make inferences about the meaning of the utterance through the activation of top-down strategies. However, if there are too many words unrecognized, there is no way for these strategies to work.

In order for spoken word recognition to be fully successful, all three phases in listening, perception, parsing and utilization and two kinds of processing, bottom-up and top-down, must be functional, because spoken word recognition is a distinct sub-system providing the interface between all these three phases (Dahan & Magnuson, 2006). It goes without saying that, in recognizing words, perceptual information from the bottom-up process would not be enough. Likewise, only with background knowledge, one cannot be successful in recognition of spoken words. Without any acoustic information, the listening process does not go up through the other higher stages, parsing and utilization, nor can one expect any feedback from the phase of utilization or top-down strategies.

In the case of Japanese EFL learners, especially learners with lower levels of proficiency, what is a cause or causes of the challenging nature of

listening, when they fail to understand a speech which is easy enough for them to understand if they listen to it with the written script at hand? They can understand easily, for example, ‘You are an athlete, aren’t you?’ when they listen to it with the written script. Without the script, however, they are unable to segment the speech stream into words and consequently cannot access the meaning when they hear /jʊrənæθli:tantju:/ without any word boundaries. This shows that they can apparently ‘read’ it visually and that they cannot ‘listen’ when they try to process the incoming speech.

It can safely be said that, of the three main phases involved in listening, what Japanese EFL learners find most challenging is the phases of perception and parsing, especially the segmentation of an utterance, informed of by perception, into meaningful chunks of words, that is, lexical segmentation (Ito, 1989; Hayashi, 1991; Yamaguchi, 1997). Ito (1989) holds that the learners’ auditory vocabulary is much smaller than their visual counterpart. They cannot comprehend an utterance even when they can understand it easily if given the written script. Their abundant grammatical and lexical knowledge are, therefore, of little use in listening (Ito, 1990). Hayashi (1991) claims that failure to comprehend even at a sentential level may be caused by inefficient processing of individual words. Noro (2006) also states that unfamiliarity with native speakers’ pronunciation is yet another big hurdle, which leads to Japanese EFL learners’ failure to recognize words.

When the learner fails to recognize words in connected speech, even though she can understand them in reading its written script, there are several possibilities. One is the case when the learner cannot find a word in the phonological lexicon (Yamaguchi, 1999), even though she has its orthographic representation in the mental lexicon. Her visual vocabulary

is there to be searched for, even though there is no auditory counterpart. Another possibility is that, even when the learner has a word or words both in the phonological and orthographic lexicon (Yamaguchi, 1999) and, if articulated individually, is able to recognize them by accessing to the mental lexicon, she cannot find the boundaries of the words, unable to segment the speech stream into meaningful chunks. This is often the case with many function words. The third possibility is the case when the learner's problem lies in the perception phase. For example, difference in phonemic systems between the learner's L1 and L2 may make her unable to distinguish two different phonemes perceptively (e.g., /l/ and /r/), or sometimes the rate of the sound stream is simply too fast.

All these possibilities might be caused by insufficient intervention of top-down strategies through the phases of parsing and utilization. Sometimes the learner does not have enough syntactic or lexical knowledge, and/or prior general knowledge about the content, and at other times she fails to activate them. However, in order to take advantage of top-down strategies, the learner has to have sufficient amount of information, above threshold level (Vandergrift & Goh, 2012), picked up from the sound stream.

A certain level of word recognition is, therefore, crucial in listening. There is no question about the claims that one of the important roles of listening instruction is to help learners deconstruct speech in order to recognize words and phrases quickly (Vandergrift, 2007), and that the problem in listening is how to match unintelligible chunks of language with their written forms (Goh, 2000; Field, 2008a). In addition, when bottom-up processing is accurate and automatic, it frees working memory capacity and thus allows the listener to build complex meaning representations. However, when it is not, it may limit the listener's ability to form a detailed

and coherent message (Field, 2008a).

2.4 The Definition of Word Recognition in this Study

Buck (2001) divides the process of recognizing words into two stages: that of recognition itself and of understanding their meaning. For meaning, listeners access their mental lexicon and elicit semantic information of the word. He argues that one of the problems about this process in listening is that the incoming signals do not indicate words by putting gaps between them, unlike in writing, so that listeners make use of all clues possible, acoustic as well as contextual.

According to Rost (2002), there are two main synchronous tasks the listener must be engaged in when recognizing words: identifying words or lexical phrases and activating knowledge associated with those words or phrases. He also suggests that the concept of a word itself is different for the spoken and written versions of any language and that the concept of a word in spoken language should be understood as part of a phonological hierarchy, with phonemes the lowest and utterances the highest. Naturally, as there is no auditory equivalent to the white spaces found in a continuous written text or any other reliable cues between word boundaries, he argues that recognition of spoken words is an approximating process marked by continual uncertainty, a process in which lexical units and boundaries must be estimated in larger groupings in the phonological hierarchy.

In reading, on the other hand, word recognition can be defined as a process of decoding continuous graphic sequences and eliciting semantic as well as phonological information of each word (Koda, 2005). When L1 speakers recognize words in reading, especially in the early stages of reading when children learn to read a written version of their native

language, they re-code graphic input into aural input which is eventually decoded for meaning (Goodman, 1973).

In L2 listening, however, the process is quite different. For L2 listeners, especially Japanese EFL learners with lower levels of proficiency, who have been accustomed to learning English in written forms, the decoding process seemingly works in the other direction; if they fail to decode auditory input into orthographic representation, they cannot access the meaning, even though this may not be the case with highly proficient listeners who can automatically recognize words as they listen.

In addition, as has been discussed in the previous section, the number of words which Japanese EFL learners can recognize in listening is smaller than that in reading. This is partly because English as spoken language has long been put on the back burner. They feel uneasy and reluctant when they are required to learn English without any written scripts.

Yamaguchi (1997) proposes a spoken word recognition model by Japanese EFL learners (Figure 2.3). She suggests some of the characteristics or strategies employed specifically by Japanese EFL learners with lower levels of proficiency:

1. Learners try to search for an auditorily perceived word in their mental lexicon after spelling it out in their mind, which is presumably caused by the fact that Japanese EFL learners are primarily accustomed to written language
2. Some learners cannot recognize the same word which they have no problem in recognizing when presented visually
- 3 In the case of some basic and highly concrete words, learners occasionally draw an image of the word in their mind before translating the word into Japanese

4. This illustrates that their recognition of words in listening reaches the semantic level by way of translation of visual or orthographic counterpart into their native language
5. It takes lower-proficiency learners a lot of time to recognize words by auditory stimuli, because they occasionally go a roundabout way along the process from speech perception, to phonological lexicon (optional), to phoneme-grapheme correspondence, to orthographic lexicon, and to semantic system (Figure 2.3).

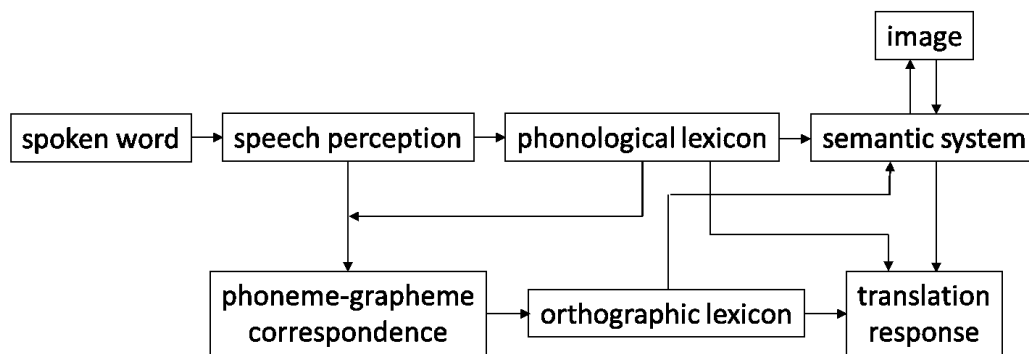


Figure 2.3. Model of word recognition process by auditory stimuli (Yamaguchi, 1997).

Yamaguchi's research as well as learning and linguistic environments of Japanese EFL learners suggests that it is not easy for them, especially beginning learners of English (Takashima, 1998), to recognize spoken words, the written version of which they have no difficulty recognizing and the meaning of which they can understand, including those words which they can recognize and understand easily in reading but graphic sequences of which do not remind them of their acoustic representation.

Against these backgrounds, in this study, success or failure of spoken word recognition means whether or not, after correctly segmenting the speech stream, one can recognize and understand the meaning of a word.

In other words, the focus is on whether the listener is able to decode a given sound stream and recognize a word in it, a word which, if given a written script, she can recognize and have access to her mental lexicon.

In this context, phonological, syntactic, pragmatic, or contextual knowledge all helps recognize a word (Buck, 2001). Strategies to be used are both bottom-up and top-down.

2.5 Variables and Skills Related to Listening Comprehension

2.5.1 Skills Related to Listening

Listening comprehension is a multidimensional process (Buck, 2001) and a highly complex problem-solving activity that can be broken down into a set of distinct sub-skills (Byrnes, 1984). Researchers have tried to present a detailed taxonomy of listening sub-skills. One example is Richards (1983). He asserts that different kinds of sub-skills are required, depending on the listening purposes: listening as a component of social action (e.g. conversational listening), or listening for information (e.g. academic listening such as listening to lectures).

Rost (1991) proposes three main skills necessary for listening comprehension: perceptual skills, analysis skills, and synthesis skills. Perceptual skills involve the ones related to distinction of phonemes and sound perception as well as recognition of words. Analysis skills are the ones related to syntactical and discourse analyses in utterance. Synthesis skills include those skills with which the listener can refer to her background knowledge, context and overall situations surrounding the utterance. These three sets of skills seemingly correspond to the three phases of listening, perception, parsing and utilization, proposed by Vandergrift and Goh (2012).

Nishino (1992) lists six related factors that will influence listening comprehension: speech perception, recognition vocabulary, grammatical knowledge, background knowledge, short-term memory, and logical inference, and conducted experiments to confirm which of the six factors are most influential in listening. He concludes that the listener's size of recognizable vocabulary, perceptive skills such as distinction of phonemes, and background knowledge are among the most influential factors in listening comprehension.

Takanashi (1982) conducted several experiments in order to elucidate factors contributing to successful listening comprehension. He suggested four main factors thought to be relevant in listening. They are abilities to understand words and grammar, perceptive skills to recognize phonemes, weak forms and prosodies, memories and abilities to think logically and take notes, and inferential skills such as those which will enable the listener to fill in missed-out information gaps. He concluded that skills to recognize weak forms, abilities to take notes, logical thinking, memories, and inferential skills seemed to be most contributing.

Takashima (1998) investigated correlation between various sub-skills and listening comprehension. He administered four types of test: phoneme identification, word recognition, listening comprehension and reading comprehension. He found that word recognition test accounted for 57% of the variance of listening comprehension test scores, and concludes that word recognition plays a basic and important role in listening comprehension and is a variable to estimate learners' listening proficiency to some degree.

Carrier (1999) lists some of the variables that affect listening comprehension. Among them are speech rate, pausing, stress, rhythmic

patterns, sandhi variations, morphological and syntactic modifications, discourse markers, elaborative detail, memory, text type, and prior knowledge.

Vandergrift and Goh (2012) list several types of knowledge as possible factors contributing to successful L2 listening. They are vocabulary, syntactic, discourse, pragmatic, and prior general background knowledge. They suggest that metacognition and L1 listening ability are also relevant.

These studies and literature suggest that abilities to discriminate phonemes which are not discernable in L1 but must be distinguished in the target language (Richards, 1983), skills to segment the sound stream into meaningful chunks of words, vocabulary knowledge, syntactic and grammatical knowledge, and contextual as well as pragmatic and prior background knowledge are all important factors and variables in successful listening.

In addition, unlike in reading, auditory message is only temporarily available and the next moment it is gone. Therefore, ability to parse and then comprehend the signals, after perception, in real time and in the syntactical order of the target language, with all the information stored in the limited working memory, is all the more important.

Furthermore, more often than not, not all the words are successfully identified in listening. Those missed-out words must be compensated with inferences based on various types of prior knowledge or schemata. In addition, transient nature of auditory input forces the listener to 'make adjustments by conjecture and inference once she has made an incorrect segmentation and has lost the sound image' (Rivers, 1971, p. 131). Only by making these inferences through the activation of top-down strategies can listening comprehension proceed smoothly even with some words

completely unrecognized (Rost, 2002). Presumably, in making inferences on unrecognized words, expectancy grammar, proposed by Oller and Streiff (1975), is yet another important component that must be factored in for L2 learners to be successful in listening. These compensatory skills are a significant aspect of listening (Buck, 2001).

2.5.2 Differences between Reading and Listening

Buck (2001) says about the characteristics of L2 listening as follows;

“If we think of language as a window through which we look at what the speaker is saying, ... in the case of second-language listening, the glass is dirty: we can see clearly through some parts, other parts are smudged, and yet other parts are so dirty we cannot see through them at all. ... When second-language learners are listening, there will often be gaps in their understanding. ... the listener may only understand a few isolated words or expressions.” (p.50)

Auditory information is intangible, something that is floating in the ‘air’ for a fleeting moment. This is all different from visual information, which is tangible, since the script is always there, on the paper, on the board or the screen, ready to be accessed at any time the reader would like. Listening requires on-line processing of acoustic input, whereas reading involves processing of graphic input and allows back tracking and review (Buck, 1992). Speech is ephemeral in nature and it exists in time rather than in space (Lund, 1991).

Quite naturally, as Buck (2001) mentioned in the above excerpt, there always exist information gaps between the speaker and the listener, because the listener may miss some parts of the utterance, unable to recognize some words. In foreign language listening, this is more evident. Therefore, listeners are required to compensate for missing words or

information by using whatever cues or knowledge available, linguistic or otherwise (Bond & Garnes, 1980), even though it is not uncommon for L2 listeners to have insufficient knowledge of the linguistic system or to lack knowledge of the socio-cultural content of the message (Underwood, 1989; Buck, 2001).

In terms of decoding incoming linguistic input into meaningful message, reading and listening share a few characteristics (Buck, 2001). However, there are some distinct differences between the two receptive linguistic activities (Lund, 1991; Flowerdew, 1994; Osada, 2004). Lund (1991) claims that differences in the processing of acoustic and printed input are apparent. He says that, for example, 'skimming is a good way to get the main idea in reading, but in listening the complete text is not available for perusal' (p. 196). Buck (1992) holds that listening comprehension trait is 'different from, or independent of, the reading comprehension trait' (p. 352).

Thompson (1995) asserts that the special effect that the aural medium has on listening comprehension should be taken into consideration. He says that listeners, unlike readers, must comprehend the text as they listen to it, retain information in memory, integrate it with what follows, and continually adjust their understanding of what they hear in the light of prior knowledge and incoming information. He says that this processing imposes a heavy cognitive load on listeners.

Osada (2004) asserts that difference in medium, sound versus print, brings about two distinctive features of spoken language as opposed to written or printed language. First, speech is encoded in the form of sound, and second, it is linear and takes place in real time with no chance of review. Her arguments are as follows:

1. Acoustic input is often indistinct. Speakers can modify the sounds considerably and all the phonemes may not be clearly and unambiguously encoded in the message. In addition, function words have two pronunciations: strong and weak forms. These modifications make the individual sounds either indistinct or missing, especially in a fast and often informal speech. Therefore, prosodic features play an important role
2. Because of the real time nature of speech, listeners must process the text at a rate determined by speakers. It is heard only once, which leaves no chance for the listeners to refer back to the text. All that remains is a memory of what was said, which is often imperfect.

There has been plenty of literature that discusses the difference between speech and written texts (Lund, 1991; Goh, 2000; Alderson, 2005). In this study, we list the following seven features that might pose difficulties in L2 listening as compared to reading.

First, as has been referred to in the previous sections, spoken language, unlike its written counterpart, does not have spaces or any other kinds of cues for boundaries between words (Rost, 2002). In addition, reduced forms resulting from assimilation and weak forms are unique to listening as they never occur in reading where writing systems are established and word boundaries are clearly marked by white spaces (Alderson, 2005). These reduced forms in listening make word boundaries in connected speech ambiguous, and word recognition and speech segmentation more difficult for listeners than readers (Goh, 2000). The listener must phonologically recognize word, phrase, sentence or any other boundaries that would be marked visually in a written text, which is not as easy as is recognized by L1 speakers of the language (Oller, 1971; Field, 2003). Failure in lexical

segmentation inevitably leads to failure in comprehension. This is never the case with reading. Readers can see word boundaries and sentence boundaries because they are clearly marked by spaces and periods. Even sentence-internal clause boundaries are sometimes visible by commas. Listeners, themselves, must punctuate a flow of speech by recognizing irregular pausing, false starts, hesitations, stress, and intonation patterns (Osada, 2004).

Second, since the spoken language is not a physical object and is only temporarily available, the listener must retain what she has just perceived in her limited working memory, while the reader can refer to the written material for information whenever possible to compensate for his¹ limited amount of memory. This nature of the spoken language, the previous information disappearing with the new message constantly coming in, not only forces the listener to process the input in real time, at the rate at which it is spoken, which is cognitively more demanding and places greater burden on their working memory (McBride, 2011), but also in the syntactic order of the target language. Listeners must comprehend the message as it is uttered. This causes a still greater problem, if the syntactic order of the learner's L1 is far different from that of the target language, which is exactly the case with Japanese and English.

Third, in listening, the learner does not have control over the method of delivery (Underwood, 1989; Yamauchi, 2014; Kajiura, 2016). The acoustic linguistic signals have some variations that written text do not share: speech rate, pausing, prosodic characteristics, sandhi variations, accents (Carrier, 1999). All these variables are uncontrollable for listeners. The reader can adjust his rate of processing the written language as he wishes, for example, while the listener cannot control or make any adjustments

over speech in any ways. Since listening is not under the timing control of the listener, it involves attention to a continuous stream of speech, whereas readers' considerable control over the texts enables them to slow down and dwell on parts of the text, skip over other parts, backtrack, and reread the passages (Harmer, 1983; Glisan, 1988; Lund, 1991; Buck, 1992; McBride, 2011).

Fourth, learners recall less information from listening than from reading in terms of both quantity and quality (Lund, 1991; Osada, 2004). When asked to recall a text presented in both oral and written modes, listeners produced less accurate recall and more reporting of ideas not found in the passage (Lund, 1991). Lower-proficiency learners benefit less from listening to a passage again than from rereading a text (Lund, 1991). This means that lower-proficiency listeners do not benefit from repeated presentations of the listening passage.

Fifth, heavy cognitive load imposed on listeners causes them to lose concentration rather quickly (Goh, 2000; Osada, 2004). Especially, lower-proficiency listeners find it difficult to maintain full concentration for a longer period. It is also difficult for listeners in general to concentrate in a foreign language. In listening comprehension, even the shortest break in attention can seriously impair comprehension, because it causes the listener to miss some parts of the speech during the break. In reading, this does not happen, because the reader can retrieve the message from the text, even if he loses concentration for a moment. It requires an enormous amount of effort to follow the meaning in listening.

Sixth, learners have a limited vocabulary. When they encountered an unknown word (in listening, the listener may not be sure if it is a word, a part of a word, or a chunk of words), it causes them to stop and think about

the meaning of the word. In reading, even if they stop, they have the following part of the text in front of them, while in listening, this stopping and thinking causes them to miss the next part of the speech (Underwood, 1989; Goh, 2000).

Last, since, in listening, not all the words are always recognizable, especially to lower-proficiency listeners, with a lot of phonetic changes (Goh, 2000; Fujinaga, 2002; Alderson, 2005) and with some words articulated ambiguously and unclearly or sometimes totally missing (Osada, 2004), compensating for words or phrases that the listener missed plays an important role (Rivers, 1971; Goh, 2000; Hasan, 2000; Rost, 2002; Graham, 2006). In reading, there is no need to predict missing words, except for such situations as taking cloze tests, because all the words are spelled out in the text. In order to predict and fill in the missing information, listeners must 'rely more on top-down, schema-based processing than readers' (Lund, 1991, p. 197).

Inference and prediction, regardless of the modalities of delivery, depends mostly on text redundancy, which is one of the most important factors in listening and is determined by two variables: the amount of information available and the listener's ability to process auditory information (Lieberman, 1963; Kohno, 1993).

In a reading cloze test, for example, the reader must predict the word to be filled in. Reduced redundancy that results from the cloze procedure forces the test-takers to rely upon their knowledge of underlying linguistic rules and also to retain the coherence of the passage to fill in the blanks (Saito, 2003). Word recognition in listening comprehension is similar to a reading cloze test in this sense. The ability to utilize redundancy in speech, which many lower-proficiency listeners lack (Kohno, 1993), is vital in

listening.

In conclusion, listening is more demanding than reading in that, not only do listening tasks place high demands on learners' attention, but also place much heavier cognitive burden on the part of learners than reading (Yamaguchi, 2001).

2.5.3 Speech Rate

Underwood (1989) mentions that the greatest difficulty with listening comprehension, as opposed to reading comprehension, is that the listener cannot control how quickly a speaker speaks and many researchers agree that speech rate is among the most influential variables in L2 listening (Richards, 1983; Griffiths, 1990; Goh, 1999; Buck, 2001; Graham, 2006). If the speech rate is too fast or natural for L2 listeners, the first phase of processing will not work and the spoken texts delivered with natural rate make word boundaries in continuous sound streams blurry, and word recognition and segmentation more difficult for the listeners (Yanagawa, 2016).

In the first phase of perception, the listener has only rapid, fleeting impressions at best, 'crudely segmented before the echo of the stimulus has disappeared from the memory' (Rivers, 1971, p.130), which is why slowing down speech rates greatly helps the listener's recognition of speech signals, since it ensures 'the prolongation of the auditory image' (Rivers, 1971, p.130). Osada (2004) suggests that one of the effective ways to relieve the difficulties of listening is to reduce the speech rate. If slowing down speech rates helps listeners bottom-up processing, which in turn affects both perception and parsing phases positively, there may be some improvement in their word recognition as well as comprehension.

On the other hand, it is said that Japanese EFL learners find natural speech rate too fast and that the speech rate is the greatest source of their listening problems, because they have been accustomed to listening to the slower rate of speech (Noro, 2006). How does a faster or slower speech rate cognitively work on the part of listeners? Many researchers have conducted experiments concerning speech rate manipulation and how different speech rates affect listening comprehension.

2.5.3.1 Effects of Rapid Listening

According to Adank and Devlin (2010), when L1 speakers listened to compressed speech up to 45%, about 1.8 times the original rate, their adaptation-related changes were observed in four separate areas both in left and right hemisphere of the brain. The changes in the right hemisphere means that adaptation may have occurred at an acoustic, rather than linguistic, level. In contrast, the changes in the left hemisphere means that adaptation-related changes have occurred at a linguistic level; that is, these changes are related to comprehension of the speech. They showed that, after hearing just 16 sentences, L1 listeners' comprehension became faster and more accurate to time-compressed speech up to 1.8 times the original rate. They argued that the ability to adapt to a time-compressed speech may rely on mapping novel acoustic patterns onto existing articulatory motor plans in the brain.

According to Dupoux and Green (1997), L1 listeners adapted to sentences compressed about 2.6 times the original rate. They claim that the perceptual system alters its criteria for judging such incoming cues in relation to the rate at which the speech was produced. They argue that some of the improvement in the ability to recognize compressed speech is

due to perceptual mechanisms involved in the normalization for speech rate. In other words, some kind of low-level tuning of the perceptual system, or perceptual recalibration or adjustment, to accommodate faster speech rate has occurred (Dupoux & Green, 1997; Golomb, Peelle, & Wingfield, 2007).

Kajiura (2016) showed that Japanese EFL learners made improvements in their listening proficiency after about fourteen hours of practice with faster rate of speech over five days. She gives two reasons why L2 learners might be able to become accustomed to listening at a fast speech rate with steady practice. One is that practice with faster speech makes it possible to understand the faster speech itself. The other is that listening practice at a faster speech rate makes learners capable of processing sounds and meanings more quickly.

As to whether the adjustment to fast rate of speech is maintained for a long term despite the need to adjustment to other slower rates, which means that the adjustment is related to permanent learning, or whether the normalization resets after the presentation of uncompressed materials, Dupoux & Green (1997) argue that presentation of slower rates does not cause a complete resetting of adjustment parameters to baseline, which means that some kind of perceptual learning takes place. This argument is supported by Golomb et al. (2007), who claim that adaptation to time-compressed speech is not hindered by brief pauses or the insertion of uncompressed sentences.

These studies suggest that listeners are able to make perceptual adjustment to highly compressed speech, both acoustically and linguistically, after several sessions of training, that the effect of adaptation is not disrupted by the normal rate of speech, and that the effect lasts for a comparatively longer period of time.

2.5.3.2 Effects of Slowing down the Speech on L2 Learners

According to Blau (1990), mechanically reducing the velocity of speech from faster (170 wpm) to slower (145 wpm) enhanced the listening comprehension of Polish and Puerto Rican learners only at the lowest level of L2 proficiency.

Griffiths (1990) experimented the effects of speech rates on listening comprehension of lower-intermediate level adult L2 learners, using three different rates; moderately fast (around 200 wpm), average (around 150 wpm), and slow (around 100 wpm). The results showed that the moderately fast speech rate resulted in a significant reduction in comprehension. However, scores on passages delivered at the slow rate did not significantly differ from those delivered at the average rate. Similar improvement in listening comprehension by slowing down the speech rate is also confirmed by Chaudron (1988) and Zhao (1997).

As for the effects of slowing down speech rates on word recognition, Kano and Saito (1997) examined the effects of speech rates, which they say consist of two parameters, articulation rates and pauses, on both word recognition and comprehension of junior college students in Japan. The parameters that they controlled were articulation rates, normal (170 wpm) and slower (130 wpm), and three kinds of pauses different in frequency and location. They found that the articulation rate is an influential factor for word recognition, which in turn influences comprehension, while frequent pauses help learners to comprehend the passages, especially when the articulation rate is high.

On the other hand, McBride (2011) experimented the effects over time with different rates of speech, using Chilean EFL learners, whose L1 is Spanish. The training session consisted of ten lessons over an unspecified

span of period. The results of the posttest, which was conducted on the same speech rate as the pretest, showed that the group that was trained on slow-rate materials did better than the one trained on fast-rate materials. She argues that this is apparently because they could perform additional mental operations such as noticing and rehearsal of features of the input. She suggests that noticed features include grammatical structures, lexical items and details about pronunciation, concluding that noticing vocabulary items and grammar structures is less likely to happen if the stream of words rushes by too quickly for L2 learners.

This might lead to the argument for the slower rate training in listening comprehension. However, as far as decoding of the speech into meaningful chunks and words is concerned, whether the slower-rate training is beneficial in listening is doubtful. Vandergrift (2004) and Yanagawa (2016) question the pedagogical usefulness of slower rate of listening, saying that they might not prepare learners to comprehend normal spoken language, because it will not help them develop strategies to cope with authentic and natural speech.

2.5.3.3 Possibility of a Faster Speech Rate as a New Baseline

There is enough experimental evidence that listeners can normalize a faster speech rate as their new baseline. In addition, according to Dupoux and Green (1997), anecdotal evidence is also abundant that, with continuous exposure to a faster speech rate, listeners find it easier to understand it and report that it sounds less fast. That is to say, when listeners make perceptual adjustment to a faster rate of speech, some kind of normalization of a faster rate takes place, which renders the original baseline somewhat slower rate to the listener.

If L2 listeners, especially listeners with lower levels of proficiency, make improvements in listening comprehension at a slower speech rates, as is evidenced by Chaudron (1988), Griffiths (1990), Blau (1990), and Zhao (1997), the normalization processing to faster speech rate in perceptual phase may affect their listening positively, because, after steady practice with faster speech, they now feel that the original rate is slow. This can best be expected in the first phase of listening, perception, which may lead to improvement in word recognition because improvement in perceptual phase may have some positive effect on the second phase, parsing. Through the improvement in bottom-up processing, some positive effect on comprehension may also be expected.

2.6 Significance of Enhancing Spoken Word Recognition for Japanese EFL learners with Lower Levels of Proficiency

L2 learners report that listening is the most difficult skill to master (Graham, 2006), and as has been discussed in the previous sections, its process is quite complicated with three main phases and two kinds of processing intricately interrelated, involves various sub-skills as well as variables, with the rate of delivery one of the biggest, and places greater cognitive load on the part of listeners as compared to reading comprehension. In this process of listening, there is no denying that deconstructing a continuous sound stream and recognizing words in fluent speech is part and parcel of spoken language comprehension (Takashima, 1998; Rost, 2002; Vandergrift, 2007). This is especially true for learners with levels of lower proficiency, because they have a problem in the first stage of listening comprehension where they cannot recognize speech sounds (Lund, 1990; Ito, 1990; Yamaguchi, 2001).

This study, therefore, focuses on lower-proficiency listeners' spoken word recognition and tries to propose effective methods of enhancing their word recognition from both perspectives of the bottom-up and the top-down processing. Close attention to prosodic, linguistic and contextual information at the same time, which is an ability many lower-proficiency learners lack, is essential to spoken word recognition (Cook and Liddicoat, 2002; Vandergrift and Goh, 2012).

As stated in section 2.4, word recognition in this study means whether or not listeners can match unintelligible chunks of language with their written forms (Field, 2008a), that is, whether or not, after lexical segmentation, they can recognize and understand the meaning of the words in a given speech, which they have no difficulty recognizing and understanding with the written text.

The next chapter discusses why spoken word recognition in English is challenging for Japanese EFL learners, from the perspectives of English phonological characteristics and the differences in the phonological features between English and Japanese.

Notes

1. This study uses a pronoun 'he/his/him' to refer to the reader.

Chapter 3

Spoken Word Recognition and Phonological Features of English

This chapter focuses on differences in phonological features between English and Japanese. The chapter discusses why it is challenging for Japanese EFL learners to recognize spoken words in English and how it should be coped with. Especially under discussion will be a relationship between a basic unit of spoken word recognition in English and the rhythmic structure of the language and a role inferences play in recognizing weak forms in a stress-timed language.

3.1 Differences in Phonological Features between English and Japanese

3.1.1 Matching Sound with Orthographic Representation

English words are not pronounced as they are written (Narita, 2013). This is obviously one of the greatest obstacles Japanese EFL learners face when they start learning English, because in their L1, as long as it is written in *kana* characters, words are pronounced as they are written. English do not have a set of rules applied in pronouncing written characters. There is a gap between the spelling of a word and its acoustic representation (Narita, 2013).

English characters, therefore, are not phonograms, which combine visual image with sound, nor are they ideograms, which combine visual image with meaning. On the other hand, Japanese dual writing systems of *kana* and *kanji* characters are the combination of phonograms and ideograms (Okazaki, 1993; Kaiser, 2001). Because of this writing system, Japanese can best be learned visually. There is no inconsistency in *kana*

characters between their written forms and speech, while visual images of *kanji* can evoke their meanings in learners (Okazaki, 1993). Learning written language is effective in accessing its meanings and it is easy to form the acoustic representation in mind only if the language is written in *kana* characters.

For example, *suisatsu-suru* and *oshi-hakaru* share the same meaning *guess*, and if those words are written in *kana* characters, learners have no difficulty articulating them, while, if written in *kanji* characters, those two words, the auditory representations of which are completely different, both have *sui* in *kanji*, which will evoke the meaning *guess* in them. When Japanese L1 speakers hear those words, they think of the visual image of *kanji*, before having access to mental lexicon for meaning. The circuit is not a direct one from sound to meaning, but from sound to visual image and finally to meaning. This is why it is not easy for Japanese L1 speakers to understand a text written only in *kana* characters, which are exactly the graphic representation of the sound (Okazaki, 1993).

This model of how Japanese L1 speakers understand an auditory stimulus in Japanese is very similar to the one, proposed by Yamaguchi (1997, Figure 2.3), for English spoken word recognition by lower-proficiency Japanese EFL learners. They put the auditory stimulus first into its orthographic representation before accessing their mental lexicon for meaning. This is not the case when L1 speakers of English or highly proficient L2 learners listen. The circuit is direct from the auditory stimulus to the phonological lexicon for meaning. Furthermore, even in reading, L1 speakers of English must first activate the acoustic representation of the text before they try to understand it (Geschwind, 1972).

Thus, it seems that, in English, sound or auditory image plays a greater part in understanding the text, written or spoken, while, in Japanese, visual image plays a bigger part since it is a language in which an ideogram called *kanji*, a visual figure with meaning, is indispensable in accessing the meaning (Okazaki, 1993). All in all, visual image is cognitively more familiar to Japanese learners than auditory image. This may partly explain why Japanese EFL learners are oriented toward and tend to rely upon written forms or visual images in learning English words with rather limited exposure to naturally spoken forms.

Furthermore, according to Kadota (2012), many Japanese EFL learners with lower levels of proficiency form an acoustic representation of a word by matching the sound of the alphabet letters with the string of characters. This image is completely different from the one that should be formed by auditorily perceiving the same word. He asserts that this is why many of them cannot recognize such a word as *apple* or *strength* in spoken forms (p. 270). The acoustic image of the word is completely different from the one they formed through mental acoustic rehearsals repeated on the same word.

Thus, a mismatch between the sound and its written representation in English is a big problem for lower-proficiency Japanese EFL learners.

3.1.2 The Phonemic System

English and Japanese do not share the same phonemic system. According to Kubozono (2013), the numbers of both vowels and consonants in English are greater than those in Japanese. Vowels are differentiated, depending on where and how the tongue is placed in the vowel space in one's palate, while consonants are articulated by disrupting the flow of air

through the pharynx to the palate and to the lips. Depending on how and where the flow of air is disrupted, consonants are distinguished. In other words, manner and place of articulation determine the consonants.

In this sense, human beings can create many more different sounds than people normally distinguish in one given language. However, the numbers of consonants and vowels that are distinguished as phonemes are unique to one language and are different from one language to another. English has 20 vowels and 25 consonants, while Japanese has 5 and 15 respectively (Kubozono, 2013).

Naturally, this difference in the number of phonemes has a significant effect on L2 listening. When Japanese EFL learners hear English sounds, more than one different vowel as well as consonant is perceived as the same one (Kubozono, 2013). Since words that have different phonemes have different meanings, it is challenging for Japanese EFL learners to distinguish and understand two different words only with their individual spoken representations, if the different phonemes in these two words are not distinct in the Japanese phonemic system (e.g., *long* and *wrong*).

Accordingly, when Japanese EFL learners listen to English words, they are required to distinguish vowels and consonants that are not distinct in their L1. Presumably, it would be easier to articulate words in a foreign language, if the numbers of vowels and consonants in the target language was smaller than those of the learner's L1, which means it would be easier to listen. Unfortunately, this is not the case with Japanese EFL learners.

Ur (1984) holds that, if a certain sound in English does not exist in the learner's L1, or if it does exist in her native language but only as an allophonic variation of another phoneme, it is very challenging for the learner to recognize the sound and the word that contains the sound. He

states that, for example, it takes a considerable amount of practice before a Hebrew speaker gets used to distinguishing between *ship* and *sheep* or *fit* and *feet* in English (p. 11).

Date (2014) examined whether phonetic instruction given to Japanese preschoolers facilitate development of their phonemic categorization. Target phonemes were /r/ and /l/ (highest level of difficulty), /b/ and /v/ (medium level of difficulty), and /p/ and /b/ (lowest level of difficulty), which were investigated for both perception and articulation. He found that on the perception side differentiation of /p/ and /b/ improved, while on the articulation side differentiation of /r/ and /l/ and that of /b/ and /v/ were proved to get better. This means that it is difficult to improve on differentiation of phonemes that are among the most challenging in listening.

This will lead to an argument that Japanese EFL learners rely more on syntactical, contextual, and other schematic information in distinguishing these phonemes than on acoustic information (Field, 2008a). Field (2008a) insists that perception of L2 never becomes entirely identical to that of L1 and that information in the form of lexical knowledge can more than compensate for uncertainties at the phoneme level. He continues, therefore, that phoneme level processing may not be as critical to successful L2 listening as is sometimes supposed. Ur (1984) also says that, even if the listener cannot distinguish *feet* from *fit*, for example, the context and the syntax help her recognize the phonemes correctly (p. 12).

However, for lower-proficiency learners, who lack enough top-down information because of poor lexical and syntactical knowledge as well as insufficient acoustic information to build contextual knowledge enough to predict indistinguishable phonemes, it is even more challenging to

recognize apart those words that contain phonemes which are most difficult to distinguish (e.g., /l/ and /r/).

3.1.3 The Syllable Structure

Syllable structures are also different between English and Japanese. English is a typical closed-syllable language, in which consonant clusters often appear, whereas Japanese is an open-syllable one, in which most of the syllables consist of CV, a combination of a consonant and a vowel.

It is said that there exists no language in the globe which does not have CV structure, that, when a person acquires L1, the very first syllable pattern to be acquired is CV and that acquisition of consonant clusters is later and difficult (Kubozono, 2013). For example, babies whose L1 is English try to avoid consonant clusters and codas by articulating *fee* instead of *free* and *fi* instead of *fish* (Kubozono, 2013, p. 14). Quite naturally, it is not easy for Japanese EFL learners, whose L1 has open syllables, to accurately master English consonant clusters and various phonetic changes such as those which often occur between codas and the first phonemes of the following word. Therefore, many Japanese EFL learners insert an unnecessary vowel after every consonant, turning a syllable with consonant clusters into multi-syllables.

For example, *mask* is a one-syllable word (CVCC), but many Japanese EFL learners with lower levels of proficiency articulate the word like *masuku*, making it a three-syllable word (CVCVCV). Because of this, a phrase such as *put a mask on*, which should sound like *putta-ma-skon*, is often articulated like *putto-a-masuku-on* by Japanese EFL learners. They place an extra vowel after each consonant and enunciate every word separately and independently, eliminating any possibility of phonetic

changes such as elision and assimilation.

Furthermore, the fact that the basic unit of CV in Japanese corresponds to the mora system and is encoded in *kana* characters in Japanese orthography complicates the matter. As has been referred to in the previous sections, Japanese EFL learners are accustomed to articulating words based on written characters. These characters in Japanese represent morae, which are basically CV, and it is not easy to access phonemes from them unlike the alphabet letters in English. This inaccessibility of 'submoraic structure' (Cutler & Otake, 2002, p.298) from written representation may be one of the reasons why Japanese EFL learners cannot successfully articulate syllables other than the CV structure.

Consequently, Japanese EFL learners pronounce each English word longer and separately, making natural English speech quite distorted, longer and totally different. What they actually hear is quite different both in length and pronunciation from their mental acoustic images of English thus created, and also from their visual images of written English. Accordingly, this distorted articulation of English words and sentences by Japanese EFL learners presumably affects their perception negatively.

It is said that many L2 learners of English have difficulty with the sequences and juxtapositions of sounds typical of English words and that many of them find consonant clusters particularly difficult to cope with (Ur, 1984). This is exactly the case with Japanese EFL learners whose L1 is an open-syllable language and has a CV-mora structure both in spoken and written representations.

3.1.4 Prosodic Features

Prosody consists of variations in pitch, stress and rhythm and serves semantic purposes, helping the speaker convey subtle shades of meaning (Monrad-Krohn, 1957). Different languages have different prosodic features. Naturally, English and Japanese do not share the same features. Especially, differences in stress patterns and rhythms between the two languages cause difficulties in recognizing spoken words (Muraio, 2006).

English is a stress-timed language while Japanese is a syllable-timed one. In English, stressed syllables are articulated long and clearly, while unstressed syllables are short and weak. In Japanese, however, the rhythm is completely different. Each mora is articulated evenly stressed (even though its pitch-accent pattern is uneven), at even intervals and in the same length (Buck, 2001; Takei, 2002). In English, regardless of the length of each word or the number of syllables, the length of an utterance depends on the number of stressed syllables (Kubozono, 2013). Unstressed syllables are pronounced short and quickly, and sometimes even elided. This makes articulation time of English speech much shorter than what might be imagined from the script (Ur, 1984; Kubozono, 2013). Spoken English is quite disproportionate in length of utterance to the number of syllables and to its written counterpart. This hiatus between spoken and written English causes trouble to Japanese EFL listeners. They are embarrassed by a rhythm and rapidness, or more precisely shortness, of an utterance, which is completely different from their mental representation of the spoken English, which is based on the written version.

Ur (1984) suggests that English systems of stress, intonation, and rhythm can interfere with the foreign learner's proper understanding of spoken English. He gave an example for the English stress pattern, saying,

“It takes roughly the same time to say *the CAT is INterested in proTECTing its KITTens* as it does to say *LARGE CARS WASTE GAS*, though the number of syllables each sentence contains is very different; whereas in most other languages, twice as many syllables simply take twice as long to say.” (p. 13)

Thanks to this ‘varied and idiosyncratic, hence unpredictable, English stress and rhythm patterns’ (Ur, 1984, p.13), lots of phonetic changes are brought about. Among major phonetic changes are assimilation, in which sounds influence the pronunciation of adjacent sounds, elision, in which sounds are dropped in rapid speech, and linking, where a new sound is introduced between other sounds like a /r/ sound in *far away* (Buck, 2001, p.33). Also, due to this stressed-timed rhythm in English, function words usually have two pronunciations; a strong form, which is used in isolation or when the word is receiving stress, and a weak form, used in connected speech when it is unstressed (Buck, 2001).

These phonetic changes significantly reduce word recognition (consequently, comprehension as well), for L2 listeners, especially Japanese EFL listeners, whose L1 is syllable-timed and whose phonological expectations might be biased largely by the written text. It is said that the role of alterations between strong and weak syllables in English is more important than phoneme level processing in perceiving speech (Field, 2008a) and that even higher-proficiency L2 listeners have difficulty with this prosodic features of English, such as stress patterns and rhythmic structure (Graham, 2006), sometimes failing to recognize words they actually know very well (Buck, 2001).

3.2 Segmentation of Continuous Speech and Prosodic Features of the Stress-Timed Language

3.2.1 Segmentation of Continuous Speech by L1 Speakers

The speech stream is not broken up into units corresponding to words at all. If we have the impression that speech comprises a discrete series of words, then this is ‘a consequence of the output of the word recognition process rather than a reflection of the nature of the input itself’ (Norris, McQueen, & Cutler, 1995, p. 1209). Since the boundaries between words are not marked in connected speech, listeners, whether in L1 or L2, need to determine for themselves where word boundaries fall. Segmentation of the continuous auditory signal into portions that can be mapped onto meaning units is a major task on the part of listeners (Cutler, Mehler, Norris, & Segui, 1986). How does the word-recognition process operate so effectively in the absence of clear cues to the location of word boundaries? How do L1 speakers of any language segment words in connected speech almost automatically?

Cross-linguistic studies of spoken-language perception have shown that speakers of different languages are sensitive to differing levels of structure in speech (Norris et al., 1995; Cutler & Otake, 2002). The key is the rhythmic structure of the language in question, which means that speech segmentation procedures vary across languages with different rhythmic structures. Native listeners of a language rely on the prosody characteristic of the language, especially on rhythmic units unique to the language in question. Simply put, language rhythm determines the segmentation unit most natural to native listeners (Cutler & Otake, 2002).

English listeners are sensitive to the boundaries between stress units (Cutler & Norris, 1988; Norris et al., 1995). A stress unit consists of one

stressed syllable and several other unstressed syllables and the boundary of the units falls between two adjacent stressed syllables with several unstressed ones in between. According to a stress-based Metrical Segmentation Strategy (MSS), proposed by Cutler and Norris (1988), English speakers use a stress-based segmentation procedure and do not use the syllable-based procedure. In stress-timed languages like English, there is a contrast between strong (S) and weak (W) syllables. Since strong syllables usually contain full vowels and weak ones schwas, this strategy seems quite appropriate for stress-timed languages like English and Dutch (Cutler & Norris, 1988). Naturally, the procedure could not operate for syllable-timed languages such as French and (more precisely, mora-timed) Japanese, in which there is no alternation of strong and weak syllables.

According to MSS, the speech stream is segmented at the start of strong syllables, and a new lexical access attempt is initiated at the start of each strong syllable. In an example like *government of a dominion* (Cutler & Otake, 2002, p. 298), which has a rhythm of SWWWWSW, the phrase has two stress units, *government of a do* (SWWWWW) and *minion* (SW); that is, the boundary falls just before the second S, *mi*.

The stress unit in this model is very similar to ‘a phonological word’ proposed by Grosjean and Gee (1987). The phonological word is a unit which is ‘made up of one stressed syllable and a number of weak syllables that are phonologically linked to it’ (Grosjean & Gee, 1987, p. 142). The difference between a stress unit and a phonological word is where the boundaries of the unit fall. The boundaries of the phonological words do not necessarily come at the start of every strong syllable (in the above example, *of a dominion* is one phonological word).

On the other hand, Japanese is a syllable-timed language. Japanese

prosody is controlled by mora, a sub-syllabic component (Cutler & Otake, 2002). Otake, Hatano, Cutler, and Mehler (1993) and Cutler and Otake (1994) propose that segmentation in Japanese is mora based and that Japanese native listeners are sensitive to the boundaries between morae.

Table 3.1 shows a comparison of rhythmic categories and the units that work as cues for speech segmentation and word recognition in English, French and Japanese.

Table 3.1.
Rhythmic Categories and the Units that Work as Cues to Speech Segmentation in Each Language (Norris et al., 1995; Cutler & Otake, 2002)

Language	Rhythm	The Units for Lexical Segmentation
English	Stress-Timed	(Boundaries of) Stress Units
French	Syllable-Timed	(Boundaries of) Syllables
Japanese	Sub-Syllable or Mora-Timed	(Boundaries of) Morae

These bodies of work suggest that each rhythmic unit unique to a language plays a role in the way listeners of the language in question segment spoken input in order to find the words in a continuous speech stream as rapidly and efficiently as possible. In other words, the boundaries of the rhythmic units function as cues or alignment points for lexical segmentation (Cutler & Otake, 2002).

Each unit listed also corresponds to the unit most relevant for describing speech rhythms in the language in question, which means that the prosodic, especially rhythmic, feature of a given language is important in the processing of spoken language. Therefore, it can be concluded that language-specific rhythmic categories and their boundaries play an important role in spoken word recognition in the target language (Cutler &

Otake, 2002).

3.2.2 Utilization of Prosodic Information in Segmenting Speech in a Stress-Timed Language

In the last section, we reviewed studies on how L1 speakers segment speech of their native language and it seems that, of the two kinds of information available for the listener, prosodic and linguistic, L1 speakers of English rely very much on stress-timed prosodic cues to locate word boundaries and segment speech into phrases and words, thereby comprehending its meaning (Peters, 1983; Oakeshott-Taylor, 1984; Grosjean & Gee, 1987; Cutler & Norris, 1988; Norris et al., 1995; Herron & Bates, 1997; Cutler & Otake, 2002; Murao, 2006).

According to Bond and Garnes (1980), listeners segment the spoken sentence into phrases which can be at least partially identified on the basis of suprasegmental patterns. Even though the distinct features of isolated utterances of syllables and words will be lost when the same utterances are delivered in continuous speech in real situations, with segments and syllables omitted and vowel colors significantly changed by consonantal environment, words can still be recognized. They claim that this is because L1 speakers of English usually understand the prosodic features of words, even when they do not catch the actual sounds.

However, it seems that this is not necessarily the case with L2 listeners. According to Cutler and Otake (2002), the language specificity of the role of prosodic or rhythmic units working as a cue in spoken word recognition is underlined by the fact that listeners are not sensitive to units relevant for other languages, or units irrelevant for their own. In other words, listeners are sensitive to their own native rhythmic categories even

in foreign language input.

For example, French listeners syllabify speech even when they listen to English words (Cutler et al., 1992), which in turn means that French learners of English find it easier to understand English spoken in a French syllable-timed manner than normal English (Vanderplank, 1993), and Japanese L1 speakers are sensitive to sub-syllabic structure even when they listen to English spoken input (Cutler & Otake, 1994), which also means that, for Japanese EFL learners, English spoken in a mora-timed manner is easier to understand than natural English. Accordingly, just as English listeners are insensitive to syllabic units or morae (Cutler et al., 1986; Otake et al., 1993), Japanese listeners are insensitive to stress units (Cutler & Norris, 1988) or phonological units (Grosjean & Gee, 1987).

Furthermore, according to Cutler and Otake (2002), morae seem to play a more central role in perception of Japanese than other rhythmic categories that are relevant for lexical segmentation in other languages. They say that morae are relevant for poetic forms, for orthography, and for many language games such as *shiritori*. This in turn illustrates that, quite arguably, this sub-syllabic system is more deeply ingrained in Japanese native speakers in perception of the language than is generally believed. This might be part of a reason why Japanese EFL learners have hard time in getting accustomed to other rhythmic categories in foreign languages, which presumably leads to the challenging nature of spoken word recognition in English.

Studies show, in fact, that Japanese EFL learners, especially learners with lower levels of proficiency, do not take advantage of prosodic features and rely only on linguistic information when they listen (Murao, 2006; Goto, 2016).

Murao (2006) examined how recognition of unstressed syllables preceded by stressed ones is varied across L1 listeners and different levels of Japanese EFL listeners in proficiency, depending on whether there is prosodic information or not and whether the expression is a formulaic one or not. The results were that native listeners rely more on prosodic features than how formulaic the sequence is, while L2 listeners, advanced or intermediate (the TOEIC scores of the participants are all 600 or more), rely more on 'formulaicness' than prosody.

This shows that native listeners of English make more use of prosodic than linguistic information in recognizing, or sometimes predicting, weak syllables, whereas Japanese EFL learners depend more on formulaicness of the sequence, that is, linguistic information.

3.2.3 A Role of Formulaic Sequences in Speech Perception

Given that not a word-by-word unit but a chunk of words that consists of a stressed syllable and several other unstressed ones plays an important role in perceiving stress-timed English speech, then the unit with its unique prosody should be a key in listening to English speech, especially in word recognition.

Peters (1983) proposed a model of language acquisition and use, in which she argued that the unit children take in is not always a word. Her main points are as follows:

1. The first units acquired by children do not necessarily correspond to minimal units or morphemes described by conventional linguists. They frequently consist of a chunk of more than one word. In storage and use, however, there are no difference between chunks and minimal ones. They are all stored in the mental lexicon and retrieved

as such

2. All the chunks in the learner's lexicon are then broken down into smaller words or morphemes. This segmentation process may be applied to material in ongoing conversations or to units already stored in the lexicon. The original chunks as well as the smaller units that result from segmentation coexist in the learner's mental lexicon
3. The learner's mental lexicon grows as the learner collects not only chunks perceived in conversations and the results of their segmentation, but also the results of a process called fusion. This process of fusion, in which already-stored chunks and units are combined, happens as often-used combinations are stored as preassembled units, without being aware what is in the units, for quick and easy retrieval. This process of fusion continues into adulthood.

In English, these chunks of words that children perceive and store in their mental lexicon are most likely to correspond to a stress unit or phonological word that helps L1 speakers of English segment speech, because the stressed syllable surrounded by several weak ones is the easiest to segment. This is why prosodic features help a baby find word boundaries and extract words (Peters, 1983).

Kadota (2012) says that English native speakers segment and store speech in their mental lexicon by a unit of chunk accompanied by its prosodic features without dividing it into words. Those units include frequently used formulaic ones.

According to Kadota (2012), a formulaic sequence is a prefabricated sequence made up of more than one word which has been memorized and is stored in the mental lexicon as an unseparated chunk and which is

searched and utilized as a whole when needed in communication. It is a sequence of words that occur together beyond the chance level such as idioms, collocations, and sentence stems (Wray, 2002). They also include syntactic relationships such as *want to* and *more and more* (Yanagawa, 2016).

Kadota (2012) holds that Japanese EFL learners, including those of higher proficiency, are less fluent than L1 speakers of English in processing formulaic sequences and that, in order to acquire formulaic sequences, learners should turn not so much to explicit learning than to implicit learning such as extensive reading and listening. Japanese EFL learners, especially those with lower levels of proficiency, supposedly lack this implicit learning in school, which will lead to insufficient amount of knowledge of English chunks that plays an important role in segmenting speech.

Studies show that opportunities of repeatedly perceiving a particular lexical sequence and its unique prosodic features such as stress and rhythm will give English listeners a cue to segment an utterance into chunks that have some meanings, even if they cannot recognize the words in the sequence (McDonough & Trofimovich, 2011; Kadota, 2012). It seems, however, that Japanese EFL learners, especially those with lower levels of proficiency, lack both utilization of prosodic information and sufficient implicit knowledge about formulaic sequences.

3.3 Acoustic Challenges of Syllable-Timed L1 Speakers in Perceiving Stress-Timed Language

3.3.1 Stress-Timed Rhythm and Speech Rate in English

As has been shown in the studies stated in the previous sections, word

recognition in stress-timed English is closely related to its prosody, especially its rhythm. What is most challenging for Japanese EFL learners is that this rhythm makes a huge gap between spoken and written versions of English.

English rhythm has a distinct feature of stressed syllables appearing at the same intervals (Kubozono, 2013). Unstressed or weak syllables, regardless of their number between the adjacent strong ones, are articulated very quickly in order to keep the rhythm from getting broken. This means that the more unstressed or weak syllables there are in-between, the more quickly each one of them are articulated and the shorter the time for its articulation becomes.

Let us compare the following four sentences. In the examples, the uppercase letters represent stressed syllables and the lowercase unstressed ones.

1. *BETH COOKS LUNCH.*
2. *BETH will COOK a MEAL.*
3. *BET⁺ty will be COOK-ing po-TA-toes.*
4. *e-LI⁻za-beth would have been COOK-ing some as-PA-ra-gus.*

(Also see sections 3.1.4 & 3.2.1)

The stress patterns of the four sentences are SSS (3 syllables), SWSWS (5 syllables), SWWWSWWSW (9 syllables), and WSWWWWWSWWSW (14 syllables) respectively in the order of No. 1 to 4. However, these sentences each contain three strong syllables, so that the rhythm and the articulation time for each sentence are no different, even though the total numbers of syllables are all different. In principle, English stress-timed nature dictates that the three stressed syllables be produced at roughly the same intervals (Martin, 1972), making the orthographically long sequences

of weak syllables very short or in some cases non-existent. From the above examples, literally at a glance, one can see the orthographic representation is quite disproportionate to its spoken counterpart.

Naturally, this rhythm is not unrelated to speech rate. In the above example, the time it takes the L1 speaker of English to articulate each of the sentences is the same, even though the total numbers of syllables are different. Accordingly, in terms of words per minute (wpm) or syllables per minute (spm), the speech rate becomes higher as the sentence contains more weak syllables, with No. 1 sentence slowest and No.4 fastest. Longer word length or longer sentence length does not necessarily translate into longer articulation time in stress-timed English, but occasionally into higher articulation rate, depending on the ratio of stressed syllables to unstressed ones (Vanderplank, 1993).

Vanderplank (1993) says that stress and rhythm unique to English are elusive and tricky phenomena for syllable-timed language speakers. In his study, the participants, all advanced-level learners of English (mostly English non-native but European language speakers), were asked to transcribe Margaret Thatcher, former British prime minister, being interviewed.

He argues that her speech, the combination of slow tempo with a high number of unstressed words, is hard to listen for L2 listeners, especially for those whose native tongue is syllable-timed. He then proposed that the best indicators of difficulty in listening are the 'pacing,' which is the tempo at which stressed words or syllables are spoken and the 'spacing,' which is the proportion of stressed words or syllables to the total. In a stress-timed language like English, he says, the influence of stress and rhythmic patterning should not be ignored in determining difficulty of understanding

speech. He concludes that the difficulties facing speakers of syllable-timed languages learning English are indeed formidable.

Thus, given that the English stress-timed rhythm is essential in its articulation, seemingly fast speech rate with a lot of unstressed syllables between each stressed one, as well as the gap in length between spoken and written English, is a hurdle that should be overcome by syllable-timed language speakers, including Japanese EFL learners whose L1 is mora-timed.

3.3.2 Difference Between Actual Auditory Stimulus and Acoustic Image Created by Japanese EFL learners

As has many times been repeated, word recognition and prosody are closely related to each other and this is none the less true in stressed-timed English, especially in its rhythmic structure of many unstressed syllables sandwiched between two adjacent stressed ones. All these characteristics of spoken English is also relevant with its speech rate, or articulation time, which in turn causes a huge gap in length between written and spoken version of English.

It is said that English native speakers have enormous amount of individual words as well as formulaic sequences, with their prosodic features attached, stored in their mental lexicon (Kadota, 2003; Murao, 2006; Kadota, 2012). In addition, it is believed that they match what they perceive in the auditory input with each metrical and phonological representations stored in their mental lexicon in segmenting the speech (Murao, 2006).

However, in the case of Japanese EFL learners, phonological representations they have in their mental lexicon, which have frequently

been transformed from written representations, are different from those that are formed through repeatedly perceiving natural speech (Kadota, 2012). Kadota (2012) holds that they have tried to remember the pronunciation of words, formulaic or other sequences through grapheme-phoneme correspondence based on *romaji* GPC rules, which seems to imply that words they have in their phonological lexicon are not always ready to be retrieved when they actually hear them. This is especially true with lower-proficiency learners (Kadota, 2012).

Fujimoto (2014) also says that Japanese EFL learners' phonetic perception of English is assisted by alphabetical information through phoneme-grapheme correspondence. This implies that Japanese EFL learners try to perceive speech by first matching the perceived sound with corresponding graphic representation.

Moreover, what they have in their phonological lexicon is the pronunciation of individual words, not the phonological representations of chunks, formulaic sequences, or stress units, made up of several words, which are combinations of one stressed syllable accompanied by a few unstressed ones, the units indispensable for word recognition in English.

Furthermore, articulation of these units (and continuous speech as well), due partly to the English stress-timed nature and partly to the closed-syllable structure, brings about a lot of phonetic changes. This is why weak syllables in connected speech are reduced or eliminated through these modifications, making its acoustic representations quite different from those of combined individual words pronounced separately. Words that can be recognized when articulated individually are not necessarily recognizable when articulated in continuous speech (Ur, 1984).

Rost (2002) refers to allophonic variations (e.g., *gonna*), alternate

pronunciations of citation forms (e.g., *going to*, p.39). These variations are brought about by co-articulation processes such as assimilations, reductions, and elisions and these simplifications make shorter not only production time for the speaker but also reception time for the listener (Rost, 2002). These efficiency principles in production hold true only for L1 speakers of the language and L2 listeners often find the simplifications more difficult to process, particularly if they have first learned the written forms of the language and ‘the citation forms of the pronunciation of words before they have begun to engage in natural spoken discourse’ (Rost, 2002, p. 40).

Another reason why Japanese EFL learners are accustomed to creating phonological representations from written characters is that morae in Japanese, which is relevant in finding lexical boundaries, are very accessible in articulating words based on *kana* characters, since the *kana* orthographies explicitly encode mora structure (Cutler & Otake, 2002). On the other hand, in English, stress units, which play an important role in spoken word recognition, are not readily available from the written text. They are not at all explicit, since ‘there are no stress marks in the orthography’ (Cutler & Otake, 2002, p. 298). Consequently, phonological representations of these units crucial in word recognition are rarely to be found in the mental lexicon of lower-proficiency Japanese EFL learners.

Thus, Japanese EFL learners’ acoustic image of English speech, combinations of individual words articulated separately based on phoneme-grapheme matching, truly reflects its visual image. The gap between their acoustic image and the natural speech, caused mainly by the disagreement between acoustic and written representation of English as well as the differences in phonemic systems, syllable structures, and rhythms of the

two languages, is huge. Japanese EFL learners generally assume that English speech is proportionate to its written form, hence is much longer than it really is. Learners with lower levels of proficiency are not generally aware that English speech is quite disproportionate in length to its written counterpart unlike in Japanese.

In addition, listening materials predominantly used in Japanese educational environment is far from natural both in terms of rhythm and phonetic changes (Osada, 2004; Yanagawa, 2016). Japanese EFL learners are accustomed to listening to a speech without reduced forms, a speech in which syllable-timed rather than stress-timed feature appears. (Yanagawa, 2016). Listening to such a speech is naturally less challenging to Japanese EFL listeners, because it is more similar to its written version and has less prosodic features unique to English speech. Naturally, hence, it might be also easier for them to recognize words in such a quasi-syllable-timed speech, without a rhythm typically found in English. Consequently, their phonological expectations might be biased largely by the written text or the quasi-syllable-timed version of the speech.

What would be expected from these discussions is that function words, which frequently consist only of unstressed syllables are more difficult for syllable-timed-language speakers, such as Japanese EFL learners, to recognize than content words, which often contain strong syllables. They expect to hear syllable-timed rhythm with its written version in mind and assume that the more syllables or words there are, the longer the articulation time is, just as it is in mora-timed Japanese. Especially for Japanese EFL learners with lower levels of proficiency, it might be beyond imagination that it takes almost the same to articulate *will* and *would have been* in many contexts.

3.3.3 Recognition of Unstressed Syllables and Function Words in a Stress-Timed Language

Studies suggest that function words, or weak syllables, are hard to recognize for many L2 listeners than content words or strong syllables (Fujinaga, 2002; Field, 2008b). In stress units, the stressed syllable contained in the content word is predominant in strength of articulation (Kubozono, 2013).

Concerning lexical segmentation and word recognition, some studies suggest that, in stress-timed English, words are not recognized in a sequential manner from left to right, but search for words begins with recognition of stressed syllables in stress units, followed by recognition, or prediction, of surrounding weak syllables based on prosodic and linguistic information (Grosjean, 1985; Luce, 1986; Grosjean & Gee, 1987; Norris et al., 1995).

Luce (1986) suggests that, in fluent speech, many of the most frequent words will not be recognized until some portion of the word-initial acoustic-phonetic information of the following word is processed, given, of course, minimal word boundary cues and contextual information relevant to the recognition of the target word.

Bard, Shillcock, and Altmann (1988) also suggest that earlier words were often belatedly recognized as subsequent words were added and that, if word recognition has failed to occur by word offset, processing must continue through the input corresponding to the next word with function words recognized late more often than content words.

Grosjean and Gee (1987) conducted gating¹ experiments. The results and the conclusions they drew can be summarized in the following four arguments.

1. Lexical search in the speech stream does not follow the process of words getting recognized sequentially from left to right. It is based on a tightly bound phonological unit, or a phonological word that is made up of one stressed syllable and a number of weak syllables that are phonologically linked to it.
2. The weak syllables in the phonological word may be the unstressed syllables of a content word and reduced function words lexically attached to or phonologically linked to content words. In segmenting the speech stream, a content word, which contains a strong syllable, is searched first and then a number of function words on either side of the content word are recognized.
3. Lexical access is done through two types of analyses: a search for stressed syllables and a pattern-recognition-like analysis to identify the weak syllables. These two types of analyses constantly interact with each other and the speech stream is segmented into a string of words with constant help from other sources of information and listeners' linguistic and situational knowledge.
4. In search of function words, the system often refers directly to a separate lexicon specifically stored for such function phrases as *might have been* and *out of the*, which is located apart from general lexicon, independent of the other lexical search for content words.

Eastman (1993) further claims that the two-way lexical search model based on prosodic structure, which was presented by Grosjean and Gee (1987), shows difficulties that syllable-timed language speakers have in listening to English and also pedagogical clues. His arguments are as follows.

1. Of the two systems stress-timed language speakers use in parsing

the speech stream into a string of words, the one shared by syllable-timed language speakers is the system in which lexical access is initiated by a search for a content word that contains a stressed syllable. The other pattern-recognition-like search system for weak syllables does not exist. Therefore, L2 listeners whose L1 is syllable-timed depend more on content words in parsing the speech stream. In order for them to recognize function words, the pattern-recognition-like search system must be developed.

2. In a stress-timed language, function words are reduced to weak forms and often with phonetic changes. Vowels are reduced to schwas or occasionally totally eliminated, which is a difference not only between speech and written language in English but also between speech in English and spoken language in a syllable-timed language. L2 learners of English whose L1 is syllable-timed pronounce every word literally, reproducing every phoneme and syllable, and stressing all syllables or avoiding distressing them while speaking. This in turn illustrates how these L2 learners listen, attempting to reconstitute unstressed syllables to their full salient form. They attempt to listen to unstressed syllables and weak forms just the way they do to content words which contains a stressed syllable.

If word recognition waits for syllables of particular clarity and is not set off at the start of every actual or potential word, if function words are more often than not belatedly recognized only after referring to some relevant linguistic and prosodic information and if there is a two-way system going on and syllable-timed language speakers lack one of them, how should weak syllables and function words be recognized by Japanese EFL learners?

Eastman (1993) suggests that it is important to teach L2 learners to pronounce content words with the weak syllables reduced and function words without stress and to explain explicitly the importance of differential weak-strong syllable rhythm. He, therefore, says that repeating frequent patterns of weak syllables, such as *out of the*, *into an*, and *to the*, both isolated and in context should help establish a growing library for L2 learners' pattern analytical system.

This draws parallels with the importance of learning and storing in the mental lexicon a lot of chunks (not individual words) and frequently used sequences or formulaic sequences, which should correspond to stress units or phonological words, with their prosodic features attached: the significance of articulating them, destressing weak syllables and appreciating their rhythm, in order to be able to listen to them. The learners should acquire not only linguistic but also prosodic information of the language

Vanderplank (1993) emphasizes the links between articulation and perception and insists on the psychological as well as linguistic benefits of training syllable-timed language speakers in the perception and production of good native speaker stress-timed speech. McDonough and Trofimovich (2009) also say that repeatedly perceiving and articulating particular prosodic patterns enables the listeners to segment the utterance into meaningful units and formulaic sequences, even if they cannot recognize each individual word.

The problem is that, as has been discussed, prosodic, especially rhythmic, cues are less likely to be taken advantage of than linguistic information by syllable-timed language speakers, especially if their proficiency is lower (Murao, 2006; Nakamura, 2012). Lower-proficiency

EFL learners cannot but rely on linguistic information to segment the utterance, which inevitably calls for activation of some kind of predictive skills in order to make up for elusive weak syllables and missing information about function words.

3.4 Significance of Top-Down Strategies Adopted by Lower-Proficiency Japanese EFL Learners

3.4.1 Information from the Bottom-Up and Activation of Top-Down Strategies

In the previous sections, it has been discussed that very little prosodic information is taken advantage of by Japanese EFL learners and they rely greatly on linguistic knowledge in segmenting speech. Naturally, linguistic knowledge without prosody means knowledge in written forms or that of spoken forms articulated in a syllable-timed manner. This is something that a large number of Japanese EFL learners, whose learning style is largely limited to that of written forms, share, regardless of their proficiency.

However, the amount of linguistic knowledge possessed by lower-proficiency learners is presumably smaller than that enjoyed by higher-proficiency learners. Considering that a certain amount of linguistic knowledge, especially that related to formulaic sequences, is necessary in making up for missing information in listening, increasing the amount of linguistic knowledge, especially grammatical and phrasal knowledge, as well as teaching them how to activate top-down strategies, especially those related to prediction, is important.

In addition, it goes without saying that there should be a minimum amount of information through the bottom up process, words picked up

from acoustic signals, in order for the top-down predictive strategies to be applied. Without a threshold level of information through the bottom up process, enough to apply top-down strategies, prediction does not function (O'Malley, Chamot, & Kupper, 1989; Rost, 2002; Vandergrift & Goh, 2012). That is why both types of approach are taken into consideration in enhancing spoken word recognition by Japanese EFL learners with lower levels of proficiency.

3.4.2 Prediction and Expectancy Grammar

Thus, even though both linguistic and prosodic information and both bottom-up and top-down processing play significant parts in spoken word recognition, it seems that learners cannot be successful without some form of predictive skills based on grammatical and phrasal knowledge. Among top-down strategies, making up for missing information, especially recognition or prediction of elusive weak syllables, is very important in stress-timed English listening.

In addition, it is said that children, in acquiring their native language, do not recognize words out of chunks segmented from continuous speech, based merely on information from acoustic signals, because they do not regard function words as words at all (Peters, 1983). They first segment speech into chunks which include stressed syllables, referring to salient forms in the speech, and when they sub-segment those chunks into words, they use top-down strategies, relying on syntactical knowledge and knowledge of the chunks (Peters, 1983). Since spoken word recognition begins with search for stressed syllables in the stress units (Grosjean, 1985; Luce, 1986 ; Grosjean & Gee, 1987; Norris et al., 1995), further lexical segmentation, or recognition of weak syllables in function as well as

content words do not proceed without reference to syntactical and phrasal knowledge (Murao, 2006). Application of top-down strategies is a must.

Oller and Streiff (1975) claim that listeners formulate some forms of expectancies or hypotheses concerning the sound stream based on their 'internalized grammar of the language' (Oller & Streiff, 1975, p.33). In the word *grammar*, they include semantic and pragmatic facts and they called this predictive skill 'a grammar of expectancy' (p. 33) or expectancy grammar. This means that listeners usually make predictions about what they expect to hear in the continuous speech, based on their grammatical, semantic, and pragmatic knowledge of the language. Oller (1979) argues that this expectancy grammar underlies language performance and is the same kind which test takers would use for completing a cloze test.

Bond and Garnes (1980) also hold that active hypothesizing concerning the speech is clearly a part of the speech perception process. They claim that acoustic information must be supplemented by non-acoustic sources for word recognition during the perception of speech and that, if phonetic signal is unclear, listeners actively employ grammatical and semantic knowledge on phonological, lexical, and sentence levels.

In the case of native English speakers, the rhythm also plays a part in anticipating what comes next in the stream of speech because there will be a stressed syllable following the last one by a roughly constant interval (Martin 1972). However, this is hardly the case with syllable-timed language speakers. Consequently, all they can rely on in prediction is their grammatical, semantic, pragmatic, contextual, and other related knowledge.

According to Lieberman (1963), one usually makes predictions through the utilization of linguistic redundancy in dealing with spoken texts and

the skill of prediction plays a greater part in listening. He says that the auditory perception of a given word in a sentence depends on the listener's knowledge of the semantic and grammatical information contained in the entire sentence. Especially, he claims, when the speaker is well aware that the listener knows the semantic and grammatical environment of a word, the speaker may utter a word with less care, because he² knows that the listener can identify the word from its context. Consequently, he says, the speaker may modify his production of a word in the light of the subsequent context of the sentence. This is more often the case with function words than with content words.

For example, in the sentence 'A stitch in time saves nine,' (Kadota, 2012, p. 276), the speaker may neglect to articulate clearly the word *nine*. He expects the listeners to understand the sentence even though they do not hear the word. However, in a sentence like 'The number that you will hear is nine,' the speaker usually articulates very carefully the word *nine*, since he knows his listeners will not be able to understand the sentence unless they recognize the word *nine*.

The same is true in formulaic sequences or idioms that involve function words. In a sentence like 'He's been under the weather lately' or 'You'd better take advantage of this,' the word *under* or *of* will never be stressed or articulated clearly. If the listener misses these words, she has to make up for herself, where activation of expectancy grammar (Oller & Streiff, 1975) is required. There is little doubt that L1 speakers or highly proficient L2 listeners do this almost automatically and has no problem comprehending the utterance even if they miss these words. However, this is not always the case with lower-proficiency L2 listeners and the inability to catch even one function word may lead to comprehension problems.

It is said that the ability to activate pragmatic knowledge during listening and to take advantage of linguistic redundancy to make predictions on the text depends on listeners' language proficiency (Kohno, 1993) and that lower proficiency listeners have greater difficulty processing both contextual and linguistic information, and, therefore, are less able to activate their pragmatic knowledge (Vandergrift & Goh, 2012).

Finally, as for recognition of function words in listening, Bard et al. (1988) make the following statements:

1. Function words used in a given context depend on the unit or the sequence to which it belongs and may be constrained by subsequent as well as by preceding context
2. In instances in which words are not immediately recognized, the word token may first activate the correct word hypothesis during its acoustic lifetime, albeit weakly and with many competitors
3. A late recognized word must be unintelligible if heard in isolation, because the acoustic evidence should have otherwise yielded immediate recognition of the word.

These bodies of literature suggest that those words which can easily be predicted in reference to contextual and pragmatic knowledge, especially many function words often found in formulaic sequences or idioms, are frequently pronounced with reduced forms and with its acoustic evidence in speech being totally different from that produced in complete isolation and that the recognition of those words which consist of unstressed syllables largely depends on the context in which they appear as well as on the listener's syntactic, semantic, and pragmatic knowledge.

3.5 What is Necessary in Enhancing Spoken Word Recognition by Japanese EFL Learners with Lower Levels of proficiency

In this section, based on variables and necessary skills related to spoken word recognition reviewed above, we itemize some possible ways of enhancing spoken word recognition by lower-proficiency Japanese EFL learners.

First, top-down strategies must be utilized. In order for that to happen, a sufficient amount of grammatical and phrasal knowledge is a must (See 3.4.2). In addition, to have them realize that in spoken word recognition they need not only acoustic information picked up through the bottom-up process but also by top-down strategies, in which they turn to grammatical, contextual and semantic knowledge, is no less important (See 3.2.3 & 3.4).

Second, from the perspective of brushing up bottom-up skills, getting accustomed to natural speech rate as well as English stress-timed rhythm is seemingly essential (See 2.5.3 & 3.3). One option may be manipulation of speech rate (See 2.5.3.3). Another should be to give the learners some sessions focusing on the differences in phonemes, syllable structures, and the rhythms between English and Japanese as well as the difference between acoustic signals and its written version of stress-timed English (See 3.1, 3.3.1, & 3.3.2). English teachers need not only to explain these differences explicitly but also to give learners a lot of perception as well as articulation practice in a correct English phonemic, syllabic, and prosodic manner as much as possible (See 3.3.3). Especially, to have them practice and remember formulaic sequences and frequently used idioms with its prosodic features attached is of prime importance (See 3.2.2, 3.2.3, 3.3.2, & 3.3.3), since they are the basic units for spoken word recognition in stress-timed English (See 3.2.1).

In the next chapter, we discuss the results of an experiment in which learners with lower proficiency levels (the TOEIC scores of the participants are mostly 500 or lower) took a spoken word recognition test in two different speech rates. Recognition of content and function words is separately analyzed, since function words, which are often pronounced weak, are presumably harder to recognize in English speech.

Notes

1. Gating is a research method in which participants hear successively longer pieces of a word in increments of 0.03 to 0.05 seconds, where a syllable lasts 0.2 seconds on average, and the participants are asked to say, after each presentation of accumulated increments, what they believe that the word is. According to Grosjean and Gee (1987), using gating, they carried out an experiment on sentences of the sort 'I saw the bun in the store', gated from the beginning of the word *bun* and measured how accurately participants recognized the word *bun* after each presentation. The results showed that 45% of the participants accurately recognized the word before the syllable *bun* ended, but that the remaining 55% did not recognize *bun* until *in* or *the* ended or some even until *store* ended. This demonstrates that words are not necessarily recognized sequentially from left to right and that the beginning of a word is not necessarily crucial to its recognition or to initiation of lexical access.
2. This study uses a pronoun 'he/his/him' to refer to the speaker.

Chapter 4

Experiment 1

Examining Lower-Proficiency Japanese EFL Learners' Spoken Word Recognition Gap Between Content and Function Words

This chapter empirically examines the effects of differences in speech rates and learners' proficiency on recognition gaps between content and function words. Based on the results of the experiment, we discuss how listeners adopt bottom-up and top-down strategies in segmenting the speech stream and recognize words. Following this, the chapter provides several implications.

4.1 Introduction

In Chapter 3, it has theoretically been confirmed that spoken word recognition requires both bottom-up and top-down processing and that without higher than threshold level of information through the bottom-up processing, no top-down strategies can be applied. In addition, even though L1 speakers of English relies greatly on prosodic information for segmentation of speech, L2 learners of English, especially those with lower levels of proficiency, tend to pay attention solely to linguistic information.

Thus, it can be presumed that, when lower-proficiency Japanese EFL learners listen to speech and try to segment it into meaningful words, they first turn to what little linguistic information they can get from the acoustic signals through the bottom-up processing, before applying assumingly meager top-down strategies and trying to make inferences on missing information, most of which would be unstressed syllables. Hence, they will

rely mostly on their hypotheses for recognition of function words.

Due to insufficient amount of information picked up through the bottom-up process, it can be assumed that recognition of content words is less than satisfactory. Accordingly, recognition of function words (and sometimes content words as well) would be even harder, because they must be hypothesized from some of the content words the learners successfully recognized. Furthermore, making inferences based on their recognition, which may often be inadequate, is no easier, because they lack sufficient grammatical and phrasal knowledge. To verify this, an experiment was conducted.

Field (2008b) empirically verified that function words are harder to recognize than content words for L2 listeners of English. In his study, however, the participants are native speakers of various languages, with L1 speakers of European languages most dominant, and not focused on mora-timed Japanese speakers. In addition, the participants were not necessarily learners with lower levels of proficiency.

In the present experiment, two different speech rates were set. The participants were Japanese EFL learners with lower levels of proficiency, whose TOEIC scores are mostly 500 or lower. In order to investigate if the difference in English proficiency works as a variable and makes a difference in recognition between content and function words, they were divided into three groups according to their proficiency levels.

4.2 Experiment

4.2.1 Purpose

The purposes of this experiment are, first, to verify whether function words are harder to recognize than content words for lower-proficiency

Japanese EFL learners, second, to investigate whether the recognition gap between the two word categories depends on their proficiency if there is a gap at all, and third, to examine how difference in speech rate affects their recognition.

4.2.2 Participants

Participants were 142 third-year and fourth-year technical college students in Japan, majoring in engineering. All of them speak Japanese as their first language. Their levels of proficiency in English were estimated using the reported Test of English for International Communication (TOEIC) score, $M = 395.71$, $SD = 119.97$.

4.2.3 Materials

For the preliminary listening comprehension test to divide the participants into three groups according to their listening proficiency, the second and pre-second grade STEP (Society for Testing English Proficiency, 2004) listening tests which consisted of 60 questions, 30 questions for each grade, were adopted.

For the listening material of a transcription test to measure the participants' word recognition, one dialogue and one monologue each recorded by English native speakers, adopted from a listening textbook¹, the texts of which are written, using the 1,000 most commonly used words in the graded vocabulary list of Standard Vocabulary List 12000 (SVL 12000), were used. The speech rate of the material was 157 wpm on average, 178 wpm for the dialogue and 141 wpm for the monologue. The texts were all read in a natural stressed-timed manner of English.

On adopting the above-mentioned materials, participants' levels of

English proficiency were taken into consideration. Especially, as for the transcription test, which was made up only of the most basic 1,000 words on the SVL list of 12000, was used to ensure that the participants of the above-mentioned level could recognize every word in the texts if given the written version of the test. All the words in the texts were so commonly used that the students in Japan should learn them while in junior high school or in the first year of senior high school. Therefore, if given the written scripts of the test and asked to recognize the words, the participants would have had no trouble recognizing and understanding them.

4.2.4 Method

The method adopted for the tests was paused transcription. The participants were told beforehand that there would be pauses inserted at irregular intervals. They listened to the recordings only once and were asked to write down the last four to five words they thought they heard before each pause. Each pause lasted about 10 seconds. The general specification of four to five words was used so as not to create unnecessary cognitive demands by encouraging participants to count the number of words to be transcribed.

Quirk, Greenbaum, Leech, and Svartvik (1985) was referred to in distinguishing content words from function words. Content words include nouns, adjectives, adverbs, and verbs, and function words are prepositions, pronouns, determiners, conjunctions, modal auxiliary verbs, and primary verbs (such as *be*, *have*, *do*).

4.2.5 Procedure

The 142 participants were grouped into two, according to the speech rate they listened at, the standard-rate group ($n = 73$) and the slower-rate group ($n = 69$), who listened to the recording made mechanically slower, at 0.7 times the rate of the standard. Regarding the reason why the slower rate was set at 70% the standard, Griffiths (1992) states that, compared with the normal rate of 188 wpm, a slower rate of around 127 wpm, about 0.68 times the normal rate, significantly facilitates the understanding of oral texts. Based on this data, the rate was reduced by 30 % for the slower-rate group.

Then, the preliminary listening comprehension test was given and the participants were divided, on the basis of the deviation values (DV) of the test, further into three groups each, six in total, depending on their respective proficiency in listening comprehension. As a result, 42 participants belonged to the high-proficiency group ($DV \geq 55$), 56 to the medium-proficiency group ($55 > DV \geq 45$), and 44 to the low-proficiency group ($45 > DV$)². Table 4.1 shows the respective number of participants in each proficiency group and the speech rate they listened at.

Table 4.1.
The Respective Number of Participants in Each Proficiency Group and the Speech Rate They Listened at ($n = 142$)

	Average Speech Rate (wpm)	The Number of Participants		
		high	medium	low
Standard-Rate Group ($n = 73$)	157	21	26	26
Slower-Rate Group ($n = 69$)	110	21	30	18

Pauses were inserted in the same place in the same text across all these six groups and participants transcribed the last four to five words

they heard before each pause. They were also asked not to use *katakana* when they were unsure of the spellings, but to use alphabet letters they thought they had heard. Pauses were inserted 16 times, 8 each for the dialogue and the monologue. Judgment of whether the participants' handwritten responses were accurate or otherwise was limited to the last four words before each pause. Therefore, 64 items, 4 each for every pause (16 times), were the maximum accurate responses possible. Of the 64, 28 were content words and the remaining 36 were function words. Sections of recordings targeted for transcription are shown in Table 4.2.

Table 4.2.

Sections of Recording Targeted for Transcription

1	<i>is in the</i> hospital	9	good <i>at</i> making baskets
2	<i>will be</i> all right	10	down <i>to the</i> river
3	<i>to eat after</i> dinner	11	go <i>across the</i> river
4	want <i>to</i> kill <i>him</i>	12	<i>were</i> tents <i>for</i> camping
5	<i>them to</i> come here	13	get <i>there by</i> car
6	asking <i>them to</i> come	14	<i>who</i> know <i>the</i> place
7	<i>who would</i> marry <i>her</i>	15	<i>but none of</i> them
8	<i>up at</i> school now	16	<i>and</i> beautiful <i>it is</i>

Function words are in italics.

Nos. 1 to 8 are the sections from the dialogue and 9 to 16 from the monologue.

Content and function words were classified in reference to Quirk et al. (1985). Due to the contextual and syntactical functions, however, *all* in No. 2 and *here* in No. 5 had meaningful content and was stressed, so that they were classified as content words, whereas *there* in No. 13 and *none* in No. 15 was functional and were articulated unstressed, so that they were classified as function words.

In grading the participants' transcription,

1. If the word boundaries were breached, all the items involved were judged to be incorrect

2. If the sound was recognized correctly with phonemes accurately distinguished (e.g. l/r, b/v), that target item was judged to be correct, even if it was misspelled
3. Of the four words which should be graded in a particular targeted section, when a blank in the first item, in the last item, or in the middle of the section was found, the remaining transcribed item or items, if they were accurately recognized, were regarded as correct.

All the data were computed into the percentage of correct word recognition, with the number of items (content and function words) correctly recognized being the numerator and the total number of items in all the sections targeted for transcription (28 content and 36 function words) the denominator.

4.3 Results

4.3.1 Results of the Listening Comprehension Test

Table 4.3 is a descriptive statistics of the preliminary listening comprehension test (Cronbach's alpha = .820).

Table 4.3.

Descriptive Statistics of the Preliminary Listening Comprehension Test (n = 142)

Groups		<i>n</i>	<i>M</i>	<i>SD</i>
Proficiency	Speech Rate			
High	Standard	21	40.38	4.84
	Slower	21	42.38	5.44
Medium	Standard	26	29.62	2.38
	Slower	30	29.63	2.24
Low	Standard	26	22.23	2.80
	Slower	18	22.11	3.61

A two-way (proficiency / speech rate) between-subjects-design ANOVA³

was conducted and no significant interaction was found ($F(2, 136) = 1.184$, $p = .309$, partial $\eta^2 = .003$), nor was the main effect of speech rate significant ($F(1, 136) = 1.062$, $p = .305$, partial $\eta^2 = .001$). However, the main effect of proficiency was significant ($F(2, 136) = 305.722$, $p = .000$, partial $\eta^2 = .806$). The results of multiple comparisons in proficiency (Tukey-Kramer Method) showed that there were significant differences between all three proficiency groups ($p < .001$).

Thus, no significant differences in listening comprehension was found across the two different speech-rate groups on all the proficiency levels.

4.3.2 Results of the Paused Transcription Test

Table 4.4 shows the descriptive statistics of the paused transcription test (Cronbach's alpha = .795). Means in percentage of content as well as function word recognition for each group (three different proficiency levels and two different speech rates) are shown in Figure 4.1.

Table 4.4.

Descriptive Statistics of the Paused Transcription Test (n = 142)

Groups		n	Correct Word Recognition in Percentage					
Proficiency	Speech Rate		A: Content Words		B: Function Words		Gap (A - B)	
			M	SD	M	SD	M	SD
High	Standard	21	63.27	11.07	27.12	13.32	36.15	11.04
	Slower	21	67.86	9.24	39.55	14.22	28.31	10.67
Medium	Standard	26	50.14	13.34	20.19	7.10	29.95	10.24
	Slower	30	60.71	9.38	25.83	11.04	34.88	10.49
Low	Standard	26	48.90	14.24	15.60	9.30	33.30	10.68
	Slower	18	51.79	15.26	19.60	8.11	32.19	11.36

A three-way mixed ANOVA³ (proficiency: high/medium/low, speech rate: standard/slower, word category: content/function) was conducted and the results are shown in Table 4.5.

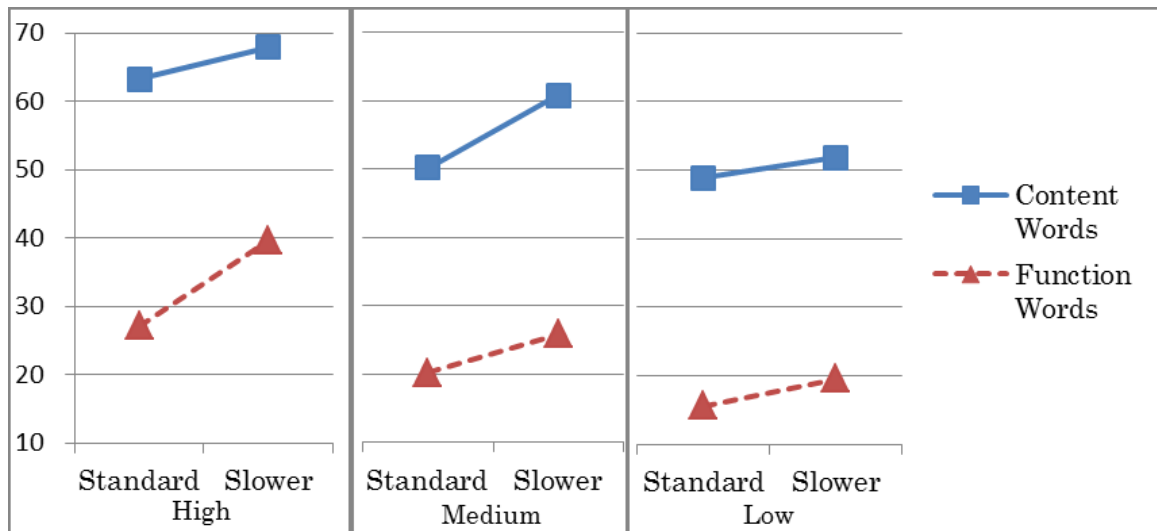


Figure 4.1. Comparisons of the means in percentage of correct content and function word recognition for six groups (three different proficiency levels and two different speech rates) in the paused transcription test.

Table 4.5.

The Results of the Three-Way Mixed ANOVA for the Paused Transcription Test ($n = 142$)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Proficiency	11390.08	2	5695.04	27.587	.000 ***	.090
B: Speech Rate	3084.30	1	3084.30	14.940	.000 ***	.024
Interaction (AB)	365.04	2	182.52	0.884	.415	.003
S: Error (AB)	28075.87	136	206.44			
C: Word Category	72660.31	1	72660.31	1269.550	.000 ***	.576
Interaction (AC)	3.15	2	1.58	0.028	.973	.000
Interaction (BC)	30.98	1	30.98	0.541	.463	.000
Second-Order Interaction (ABC)	469.53	2	234.76	4.102	.019 *	.004
Error (CS)	7783.70	136	57.23			
Total	126197.76	283				

***: $p < .001$, *: $p < .05$

The results of the ANOVA demonstrated that a significant second-order interaction was found ($F(2, 136) = 4.102$, $p = .019$, partial $\eta^2 = .004$), so that simple interactions at each level of the three factors between all the combinations of the other two factors were examined (Table 4.6). The only simple interaction that was significant was the one between the speech rate and the word category in the high proficiency group ($F(1, 136) = 6.174$, p

= .014, partial $\eta^2 = .043$).

Table 4.6.

Simple Interactions at Each Level of the Three Factors (n = 142)

	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Proficiency	BC (High)	353.37	1	353.37	6.174	.014 *	.043
	BC (Medium)	139.99	1	139.99	2.446	.120	.017
	BC (Low)	7.15	1	7.15	0.125	.724	.001
	Error	7783.70	136	57.23			
B: Speech Rate	AC (Standard)	221.65	2	110.83	1.936	.148	.027
	AC (Slower)	251.03	2	125.51	2.193	.116	.030
	Error	7783.70	136	57.23			
C: Word Category	AB (Content)	375.05	2	187.53	1.422	.243	.010
	AB (Function)	459.51	2	229.76	1.743	.177	.013
	Error	35859.58	272	131.84			

*: $p < .05$

The fact that no other simple interaction but the one between the speech rate and the word category in the high proficiency group was significant implies that only in the high proficiency group did the gap between content and function word recognition become closer because of the slower speech rate, which can also be seen from the graph in Figure 4.1. This was not the case with the other two proficiency groups, medium and low.

For further analyses, examinations of simple main effects and multiple comparisons (Ryan's method) were carried out. The results are shown in Tables 4.7 and 4.8.

Table 4.7.

Simple Main Effects of Three Factors at Each Level of the Combinations of the Other Two Factors (n = 142)

	Source	SS	df	MS	F	p	η_p^2
A: Proficiency B: Speech Rate	C (High & Standard)	15017.16	1	15017.16	262.386	.000 ***	.186
	C (High & Slower)	9208.34	1	9208.34	160.892	.000 ***	.114
	C (Medium & Standard)	10304.99	1	10304.99	180.053	.000 ***	.127
	C (Medium & Slower)	13982.16	1	13982.16	244.302	.000 ***	.173
	C (Low & Standard)	12745.57	1	12745.57	222.696	.000 ***	.157
	C (Low & Slower)	11905.76	1	11905.76	208.022	.000 ***	.147
	Error	7783.70	136	57.23			
A: Proficiency C: Word Category	B (High & Content)	242.31	1	242.31	1.838	.176	.006
	B (High & Function)	1776.68	1	1776.68	13.476	.000 ***	.045
	B (Medium & Content)	1285.63	1	1285.63	9.752	.002 **	.032
	B (Medium & Function)	365.69	1	365.69	2.774	.097	.009
	B (Low & Content)	95.63	1	95.63	0.725	.395	.002
	B (Low & Function)	183.92	1	183.92	1.395	.239	.005
	Error	35859.58	272	131.84			
B: Speech Rate C: Word Category	A (Standard & Content)	2912.86	2	1456.43	11.047	.000 ***	.061
	A (Standard & Function)	1545.41	2	772.71	5.861	.003 **	.032
	A (Slower & Content)	2980.51	2	1490.25	11.304	.000 ***	.062
	A (Slower & Function)	4789.02	2	2394.51	18.163	.000 ***	.100
	Error	35859.58	272	131.84			

***: $p < .001$, **: $p < .01$

Table 4.8.

The Results of Multiple Comparisons of Proficiency at Each Level of the Speech Rate and the Word Category (Ryan's method) (n = 142)

Source	Group 1 (I)	Group 2 (J)	Difference (I - J)	t	p
A: Proficiency (Standard & Content)	High	Medium	13.13	3.897	.000 ***
	High	Low	14.36	4.264	.000 ***
	Medium	Low	1.24	0.388	.698
A: Proficiency (Standard & Function)	High	Medium	6.92	2.055	.041 *
	High	Low	11.52	3.419	.001 ***
	Medium	Low	4.59	1.443	.150
A: Proficiency (Slower & Content)	High	Medium	7.14	2.186	.030 *
	High	Low	16.07	4.358	.000 ***
	Medium	Low	8.93	2.608	.010 **
A: Proficiency (Slower & Function)	High	Medium	13.72	4.199	.000 ***
	High	Low	19.95	5.410	.000 ***
	Medium	Low	6.23	1.821	.070

***: $p < .001$, **: $p < .01$, *: $p < .05$

First, as for the percentage in correct word recognition, recognition of

content words is significantly better than that of function words across all the proficiency levels and speech rates ($p < .001$). Next, as for the difference in word recognition by different proficiency groups, there was a significant difference across both the speech rates and both the word categories between high and medium proficiency groups (standard rate & content words $t = 3.897$, $p < .001$, standard rate & function words $t = 2.055$, $p < .05$, slower rate & content words $t = 2.186$, $p < .05$, slower rate & function words $t = 4.199$, $p < .001$). However, differences in standard rate and function words and in slower rate and content words were smaller than those in standard rate and content words and in slower rate and function words. On the other hand, between medium and low proficiency groups, difference was significant only in the combination of slower speech rate and recognition of content words (standard rate & content words $t = 0.388$, $p = .698$, standard rate & function words $t = 1.443$, $p = .150$, slower rate & content words $t = 2.608$, $p < .01$, slower rate & function words $t = 1.821$, $p = .070$).

Finally, as to an influence of different speech rates on word recognition, recognition of function words by high proficiency groups ($F(1, 272) = 13.476$, $p < .001$) and that of content words by medium proficiency groups ($F(1, 272) = 9.752$, $p < .01$) were significantly better at the slower speech rate than at the standard speech rate. However, no significant influence was found for all the other combinations except that recognition of function words by medium proficiency groups was only marginally significantly better at the slower speech rate than at the standard speech rate (high proficiency & content words $F(1, 272) = 1.838$, $p = .176$, medium proficiency & function words $F(1, 272) = 2.774$, $p = .097$, low proficiency & content words $F(1, 272) = 0.725$, $p = .395$, low proficiency & function words $F(1, 272) = 1.395$, $p = .239$).

4.3.3 Analyses of the Gap in Word Recognition Between the Two Word Categories

The results of the three-way mixed ANOVA showed that there was a significant second-order interaction and that, of all the simple interactions, only the one between speech rate and word category in the high proficiency group was significant. This means, as has been mentioned above, that only in the high proficiency group did the recognition gap between content and function words become closer at the slower speech rate. In this section, this was re-examined through analyses of the recognition gap between content and function words. The following are results of a two-way (three proficiency levels / two speech rates) between-subjects-design ANOVA³ conducted on the recognition gap between content and function words (Table 4.9). The means of word recognition gaps at different speech rates for three levels of proficiency groups are shown in Figure 4.2.

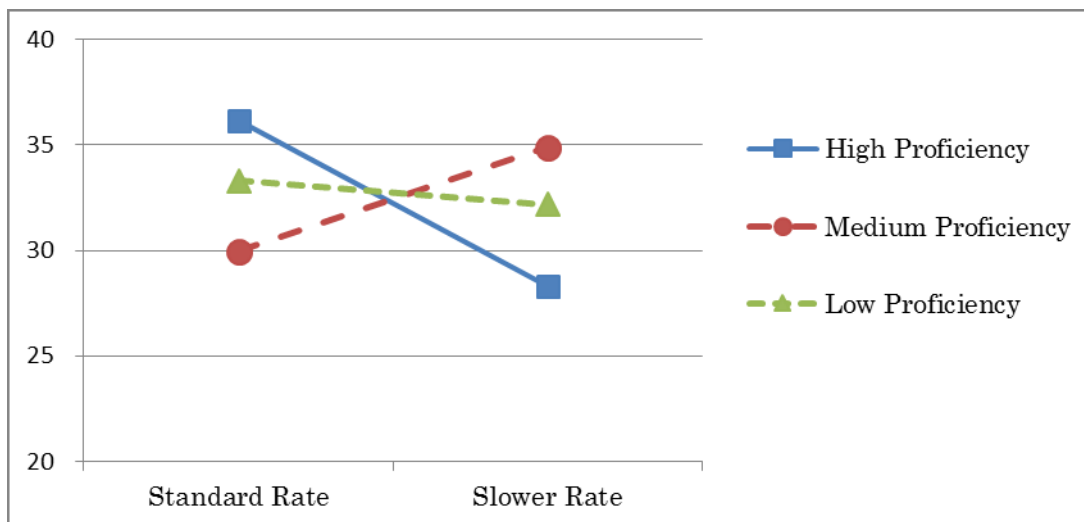


Figure 4.2. Word recognition gap in percentage between content and function words at different speech rates for the three different listening proficiency groups.

Table 4.9.

The Results of the Two-Way ANOVA for the Word Recognition Gap Between the Two Word Categories (n = 142)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Proficiency	5.84	2	2.92	0.026	.975	.001
B: Speech Rate	61.97	1	61.97	0.541	.463	.004
Interaction (AB)	979.82	2	489.91	4.280	.016 *	.059
Error	15567.41	136	114.47			
Total	16574.00	141				

*: $p < .05$

Table 4.10.

Simple Main Effects of Proficiency and Speech Rate on Word Recognition Gap Between the Two Word Categories (n = 142)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2	
B: Speech Rate	A (Standard)	452.86	2	226.43	1.978	.142	.027
	A (Slower)	533.88	2	266.94	2.332	.101	.032
	Error	15567.41	136	114.47			
A: Proficiency	B (High)	645.72	1	645.72	5.641	.019 *	.039
	B (Medium)	339.34	1	339.34	2.965	.087	.020
	B (Low)	13.24	1	13.24	0.116	.734	.001
	Error	15567.41	136	114.47			

*: $p < .05$

A significant interaction was found ($F(2, 136) = 4.280, p = .016$, partial $\eta^2 = .059$), so that simple main effects of proficiency and of speech rate at each level of the other factor were examined (Table 4.10). The results of ANOVA demonstrated that no other simple main effect than that of speech rate in the high proficiency group was significant ($F(1, 136) = 5.641, p = .019$, partial $\eta^2 = .039$). This signifies that the slower speech rate enabled only the high proficiency group listeners to close the recognition gap between content and function words significantly. This was not the case with the other two proficiency groups.

4.3.4 Correlation Between Listening Proficiency and Word Recognition

Finally, correlations between listening proficiency and word recognition were analyzed both for the standard and slower speech rates, using Pearson product-moment correlation coefficients³ (Table 4.11).

Table 4.11.

Correlations Between Listening Proficiency and Word Recognition for Different Speech Rates (n =142)

	Standard Rate		Slower Rate		Total				
	Content	Function Gap	Content	Function Gap	Content	Function Gap			
Listening Proficiency	.433 ***	.548 ***	.024	.496 ***	.664 ***	-.265 *	.466 ***	.607 ***	-.126
Content Word		.667 ***	.659 ***		.656 ***	.321 **		.676 ***	.477 ***
Function Word			-.122			-.504 ***			-.325 ***

***: $p < .001$, **: $p < .01$, *: $p < .05$

First, Pearson product-moment correlation coefficients suggested moderate positive relationships between listening proficiency and both content and function word recognition across both the speech rates. Relationships between listening proficiency and recognition of function words were slightly stronger than those between listening proficiency and recognition of content words and, as for a comparison between two different speech rates, the relationships were slightly stronger for the slower rate group.

Second, positive correlations were found between content word recognition and the recognition gap across both the speech rates. This means that, as percentage in correct recognition of content words becomes higher, the recognition gap between content and function words grows wider; that is, recognition of function words does not improve as that of content words does. On the other hand, as for correlations between

recognition of function words and the recognition gap, no significant relationships were found for the standard rate group, though there were moderate negative relationships for the slower rate group. This implies that, if the speech rate is fast enough, the recognition gap between content and function words stays near constant with the growth in correct recognition of function words. However, at the slower rate, the recognition gap between the two word categories narrows as the percentage in correct recognition of function words rises; that is, recognition of function words becomes even better than that of content words.

Third, as for the correlations between listening proficiency and the gap in word recognition, no significant relationships were found for the standard rate group. This means that even listeners who belong to the upper bracket among the participants in the present study cannot close the recognition gap between content and function words. However, for the slower rate group, weak but significant negative relationships were found between listening proficiency and the recognition gap, which implies that the recognition gap between the two word categories closes as listening proficiency rises. This demonstrates that the slower speech rate had an influence on recognition gap between content and function words, especially when learners' listening proficiency is moderately higher.

4.4 Discussion

The results of the experiment demonstrated that the percentage of correct recognition of function words was significantly lower than that of content words across all the levels of listening proficiency and speech rates. This means that function words are more challenging to recognize than content words for lower-proficiency Japanese EFL learners, whose TOEIC

scores are mostly 500 or lower. This is consistent with the hypothetical discussion we had before this experiment.

Second, as far as the Japanese EFL learners with lower levels of proficiency are concerned, their listening proficiency has positive correlations with recognition of content and function words. However, the gap in recognition between content and function words at the standard speech rate was 36.15% for the high proficiency group, 29.95% for the medium proficiency group, and 33.30% for the low proficiency group respectively. As the results of ANOVA for the word recognition gap showed, the simple main effect of proficiency at standard speech rate was not significant ($F(2, 136) = 1.978, p = .142, \text{partial } \eta^2 = .027$), and the Pearson product-moment correlation coefficient for the standard rate group suggested no relationship between listening proficiency and the word recognition gap. These results demonstrate that the gap in word recognition between content and function words is constant across all the listening proficiency levels, which means even those in the higher bracket of lower-proficiency Japanese EFL learners find recognition of function words challenging, when the speaker's speech rate is around 150 to 160 wpm. This is also consistent with the discussion before the experiment, implying that, in recognition of function words, there are some other factors involved than in recognition of content words.

It takes some inferences based on the information gathered through the bottom-up strategies to recognize function words, because they are articulated weak and different from the pronunciation in isolation, or sometimes totally disappear. However, to make inferences on the fragments of information successfully picked up, one needs to have sufficient amount of grammatical, semantic, and pragmatic knowledge, something which the

participants of the present experiment supposedly lack.

Finally, intriguing results were obtained as to the way a difference in speech rate affects spoken word recognition and the recognition gap between content and function words. The data obtained in the experiment reveal that different speech rates have different influences on word recognition by lower-proficiency Japanese EFL listeners as their listening proficiency varies. The results indicate, when the participants listened to the recordings at a slower rate (about 110 wpm), that percentage of their correct recognition of function words for the high proficiency group and that of content words for the medium proficiency group were significantly higher. In addition, only for the high proficiency group did the recognition gap between content and function words become significantly smaller.

As to the high proficiency groups, the participants in the slower rate group recognized function words significantly more correctly than those in the standard rate group, even though no significant difference was found across both the speech rates in the recognition of content words. As a result, the recognition gap between the two word categories became significantly smaller due to the difference in speech rates.

One possibility is that the slower speech rate enabled the participants to identify word boundaries and words themselves more easily so that the recognition of both content and function words slightly improved through bottom-up processing. This increased amount of information retrieved from bottom-up processing was large enough to trigger, in turn, top-down processing. In other words, the amount of information that listeners gathered through bottom-up processing went past the threshold (O'Malley, Chamot, & Kupper, 1989; Eastman 1993; Vandergrift & Goh, 2012), which enabled them to make use of top-down strategies and the participants were

able to make inferences on missing function words after turning to top-down processing, taking advantage of schema and grammatical knowledge. Consequently, their recognition of function words improved significantly better than that of content words.

Another possibility is that the slowness of the rate itself had some influence on participants' word recognition. Many function words are not spoken independently or isolated from other words in a connected speech, but are usually spoken in a group of content and other function words, thus forming a stress unit or a phonological word (Grosjean & Gee, 1987), which is assumed to be a basic unit for speech segmentation in stress-timed English. These groups of words, or phonological and lexical sequences, in the speech stream were supposedly recognized over a slightly longer time span, by a split second, which affected recognizing process positively and helped the participants identify more accurately the words they heard, even the unstressed syllables. This was presumably impossible when they were listening to the recordings at the standard rate. Quite possibly, there may also have been some cues from top-down processing. On the other hand, content words can be searched from stressed syllables so that they were recognized easily enough at a faster rate. Therefore, recognition of content words was not affected even if the speech rate was slower.

For the medium proficiency groups, however, the participants' recognition of content words was significantly better at the slower speech rate than at the standard rate, whereas their recognition of function words was only marginally better at the slower rate. These results imply that top-down processing, which presumably became available to the participants in the high proficiency groups, was still not in use by them, or that the threshold itself to initiate top-down processing may be higher than in the

case of the participants in the high proficiency groups.

One possibility is that the participants in the medium proficiency groups needed more information retrieved from bottom-up processing in order to activate higher-order knowledge. Another is that, even with the increased amount of information through the bottom-up process, they were still unable to successfully recognize function words because of insufficient grammatical and phrasal knowledge.

However, the participants in the slower rate group recognized content words significantly better than those in the standard rate group. This implies that the threshold for recognition of spoken content words, which divides recognition of content words in spoken English and that in written English, falls somewhere between these speech rate ranges for the participants in the medium proficiency groups, a threshold that did not affect in any way those in the high proficiency groups due to the difference in speech rates. Simply put, the participants of this proficiency groups recognized at the slower rate what they could have recognized in the transcribed version of the recordings. Due to the slower speech rate, the participants were able to recognize content words, made up of stressed syllables, which would have been impossible to recognize at the standard rate.

Nevertheless, this difference in speech rate never affected their recognition of function words, which in the speech stream can be heard and recognized only in a stress unit, a sequence of words including a content word. This is probably because the slower rate was still too fast for the participants in the medium proficiency groups to recognize function words, made up only of unstressed syllables, exclusively from the information they gathered from bottom-up processing. Furthermore, the increased

information on content words through the bottom up processing due to the slower rate did not trigger top-down processing, which would otherwise have led to successful recognition of more function words. Consequently, the recognition gap between content and function words, though not significantly, widened. Insufficient grammatical knowledge may have played a part as well.

Lastly, for the participants in the low proficiency groups, neither influence was observed over the speech rate ranges in the experiment. Possibly, there might have been some changes measured, if the rate had been made further down. Another possibility is that their vocabulary size may be too small to make a difference or recognize words in the speech stream.

One possibility which can be drawn from the discussion above is that the threshold level, at which listeners can make use of top-down strategies, may fall somewhere between the percentages of correct word recognition at the standard and the slower speech rate for the high proficiency groups; that is, around 65% for content words and a little above 30% for function words. Granted that only the participants in the high proficiency groups improved their recognition of function words significantly better than that of content words due to top-down processing triggered by the slower rate, the participants in the medium proficiency groups did not reach this probable threshold, with 61% for content and 26% for function words even at the slower speech rate, leading to the failure to close the recognition gap between content and function words.

4.5 Further Implications

In parsing the speech stream into a string of words, a stress unit,

which consists of one stressed syllable and many unstressed ones, or a sequence of words that are phonologically and semantically linked, plays an essential part in stress-timed English, which has been already discussed in the previous chapter.

Therefore, in order to successfully recognize unstressed syllables, many of which are function words, the process must begin with the recognition of sufficient amount of stressed syllables through the bottom-up process, which eventually leads to recognition of content words. In addition, the amount of content words should be enough to trigger and activate top-down strategies. Furthermore, in order to make the most of top-down strategies, one must possess sufficient amount of grammatical, semantic, and pragmatic knowledge. Naturally, formulaic sequences also play a bigger part, even without the prosodic information, because they become a basis for inferences of function words or utilization of expectancy grammar.

Hence, to give learners a lot of phrasal as well as semantic and other background knowledge is important, especially when the learners' native language is a syllable-timed one, because the learners, especially lower-proficiency learners tend to rely only on linguistic information of the speech, rendering inferences through the top-down processing a very important strategy to be taken advantage of. These will be empirically examined in Chapters 5 and 6.

Second, this chapter showed that the speech rate is an important variable that should be taken into consideration in syllable-timed language speakers' recognition of spoken words in stress-timed English, which has a prosodic structure featured by the alternation of weak and strong syllables.

If difference in speech rate affects spoken word recognition, it may be

effective to use mechanically compressed recordings in listening class in order to improve Japanese EFL learners' spoken word recognition. By having the learners constantly exposed to higher-rate listening, and thereby having them get used to it, it may be plausible for the learners to improve word recognition at the baseline rate. This will be examined in Chapter 7.

Finally, it is also important for the listeners to get accustomed to English stress-timed rhythm and to make at least some use of prosodic features in order to enhance spoken word recognition. After all, one of the main reasons why function words are more challenging to recognize than content words can be attributed to English stress-timed prosodic features.

As Eastman (1993) claimed, if English stress-timed prosodic features cause the gap that divides written and spoken English or a spoken syllable-timed language and spoken English, and if Japanese EFL learners attempt to recognize every syllable, stressed or unstressed, as a stressed one, then there certainly is an excessive cognitive load on them by listening to every single unstressed syllable as if they were stressed just the way they do when they deal with written English or a syllable-timed language. It is also highly probable that they read aloud or speak English without distressing any syllable. If so, as Rost (2002) claims, Japanese EFL learners should not learn and practice citation forms of the pronunciation from written English, but engage in natural spoken discourse and continue to practice pronouncing each syllable and word just as they hear in the speech stream, stressing and distressing each syllable distinctly.

In short, when Japanese EFL learners read aloud English text, their articulation should fully reflect phonological features unique to English. This may be an effective way to hone the learners' bottom-up skills in

perceiving speech in a stress-timed manner, taking advantage of prosodic cues, which will help further recognize function as well as content words.

Quite possibly, this may be an effective method to develop in Japanese EFL learners' mind a system of recognizing strings of unstressed function words accurately in the fleeting speech stream, searched from the stressed syllable contained in the content words; that is, segmenting speech by stress unit just as native English speakers do. This is what Grosjean and Gee (1987) and Eastman (1993) call the high-rate and pattern-recognition-like search system for function words. In addition, if learners can remember formulaic sequences or idiomatic phrases with their prosodic features attached, it may help develop a separate lexicon specifically stored for direct search of function words, claimed by Grosjean and Gee (1987) and Eastman (1993).

For this purpose, it may be also important not to let learners' pronunciation habits formed by grapheme-phoneme correspondence based on *romaji* GPC rules or isolated pronunciation of each word get into English listening practices. In addition, it is necessary to stick to the same measures in oral reading and shadowing practices so that the gap between Japanese EFL learners' oral reading and English L1 speakers' natural speech will be closed. This will be examined in Chapter 8

Notes

1. The title of the listening textbook used in the experiment was *Kyukyoku-no-eigo-listening* (Ultimate English listening) series level 1, published by ALC Press.
2. TOEIC scores of each proficiency group were as follows: high, $M = 489.56$ ($SD = 94.65$), medium, $M = 373.22$ ($SD = 62.06$), and low, $M =$

300.30 ($SD = 65.16$).

3. As for analyses of the data, an online software, ANOVA 4, was used for the three-way ANOVA and a Microsoft add-in software for Excel was used for the two-way ANOVAs and for the analysis of Pearson product-moment correlation coefficients.

Chapter 5

Experiment 2

Examining Whether Handing out Japanese Translation Beforehand Can Activate Top-Down Strategies

This chapter empirically examines whether giving meanings before dictation practices activates top-down strategies and has positive effects on word recognition. In the experiment, only one of the experimental groups was given Japanese translations of the scripts beforehand. Whether the treatment was effective in activating top-down strategies will be discussed.

5.1 Introduction

In Chapter 4, it has been suggested that in order to enhance spoken word recognition, especially recognition of weak syllables or phonologically modified and weakened versions of function words, application of top-down strategies is necessary.

However, lower-proficiency learners tend to process connected speech only on a word-by-word basis, desperately trying to match the incoming sound with words familiar to them, and use top-down strategies insufficiently (Berne, 1998; Osada, 2001; Field, 2003). Nevertheless, in order for top-down strategies to be applied, sufficient amount (above threshold level) of words, especially content words, recognized from strong syllables through bottom-up processing, is essential in triggering and activating such strategies.

In this chapter, it will empirically be examined if treatment of giving out Japanese translations before dictation practices would help activate

top-down strategies. This is because giving learners certain amount of semantic and other background knowledge would not only be useful in guessing the content words that will appear in the speech, but also help learners make inferences on elusive weak syllables, which we assume is possible by taking advantage of increased amount of information from the content words successfully recognized and also by turning to learners' internalized grammatical knowledge, such as expectancy grammar (Oller & Streiff,1975)¹, which is supposedly quite limited in the case of lower-proficiency learners.

5.2 Experiment

5.2.1 Purpose

The purposes of this experiment are to investigate whether Japanese EFL learners with lower levels of proficiency can enhance their spoken word recognition by paying more attention to semantic, contextual, and syntactical elements, thereby activating and effectively applying top-down strategies.

In the experiment, only the experimental group was given Japanese translation of the text they would soon hear and was also instructed to make inferences on English sentences that may appear in the upcoming speech before each dictation practice during the treatment period. The effects of this treatment will be examined from two perspectives.

First, it is examined whether or not the learners' spoken word recognition will improve, if they are given semantic information through Japanese translation and instructed to guess about English sentences they will hear.

Second, it is examined whether the improvement in spoken word

recognition is due not only to the application of such top-down strategies as semantic and contextual inferences but also to the activation of internalized grammatical and phrasal knowledge, which will presumably enhance recognition of function words.

5.2.2 Participants

Participants were 56 third-year students at a private high school in Japan. All of them speak Japanese as their first language. All the participants took a two-credit elective subject 'English Practice' and the experiment was conducted in this class. Fifty-six participants were divided into three groups, two control groups (Control Groups 1 and 2), and one experimental group (Experimental Group), so that the participants' English proficiency in each respective group would become even. Consequently, 20 students belonged to Control Group 1, 16 to Control Group 2, and 20 to Experimental Group.

Besides 'English Practice' class, the third-year students of the school were supposed to take six credits of required English class. It can be assumed, however, that there was little difference among the participants of the three groups in the time they spent on English study including the time they spared for English at home, except for the treatment stated below, during the three-month experiment. This is because all the participants had already decided to proceed to the university affiliated to the school, hence no need for preparation for entrance examinations to other universities in February and March, even though the experiment was conducted from the beginning of November until mid-January. Nevertheless, the participants were fairly motivated to study English, because they were required to reach the goal of 400 on TOEFL ITP Level 1

Test, whose score range is the same as that of TOEFL PBT Test, which spans from 310 to 677, after entering the affiliated university.

As for the participants' English proficiency, they could be assumed to be learners at beginner level, because most of them had yet to reach the goal of 400 in TOEFL ITP Test. Their mean score of a listening comprehension test using TOEFL ITP, which was conducted in class, was 34.68 ($SD = 11.31$) in percentage, well below 40%.

The mean scores for each group of this test were as follows: 35.10 for Control Group 1 ($n = 20$, $SD = 14.59$), 32.31 for Control Group 2 ($n = 16$, $SD = 6.87$), and 35.88 for Experimental Group ($n = 20$, $SD = 9.99$). The result of one-way between-subjects-design ANOVA² showed that there was no significant difference between the three groups in terms of listening proficiency ($F(2, 53) = 0.383$, $p = .684$, partial $\eta^2 = .001$).

5.2.3 Pretest and Posttest

Different texts were used for the pretest and the posttest (Appendix 1) to avoid the learning effects. They were both adopted from a listening textbook³, the texts of which are written, using the 2,000 most commonly used words in SVL 12000. The words used in the texts would have been easy enough for the participants to recognize, if they had been given the written script of the texts. Both the pretest and the posttest consisted of one dialogue and one monologue.

The materials in the textbook were graded into three levels, depending on the vocabulary used in the text, the number of words, and the speech rate, and the dialogues and monologues used in the tests were all from the most difficult level. The numbers of words in the dialogues and the monologues for both the pretest and the posttest were around 170 and 330,

respectively. The speech rate was all around 170 wpm. The texts were all read in a natural stressed-timed manner of English. Even though the speech rate of 170 wpm was rather fast for high school students, compared with that of CDs they usually listen to, it was adopted because the rate was considered to be close to the standard and naturalness of English was thought to be important.

The tests were transcription tests in which the participants were asked to spell out one word each in the blanks. The blanks were located every several words. One hundred words in total were blanked out in about 500-word texts for both the pretest and the posttest. Pauses were inserted at the end of each sentence so that the participants could have enough time to write. Each pause lasted about five to ten seconds, depending on the number of blanks they were supposed to fill in. However, when a sentence was considered to be too long for them to retain what they heard, additional pauses were inserted where major syntactic and/or semantic boundaries were located. The participants listened to the recordings only once and Japanese translations of the listening texts were not given beforehand in both the tests.

Finally, in grading the transcription, if the sound was recognized correctly with phonemes accurately distinguished (e.g. l/r, b/v), that target item was judged to be correct, even if it was misspelled

5.2.4 Treatment

Between the pretest and the posttest, the participants of different groups were given different treatments. Those in the Control Group 1 were given only the normal class during the period. No additional listening activities were provided. To Control Group 2 and Experimental Group, on

the other hand, dictation practices were given once or twice every week during the period, in addition to the normal class. There were 11 dictation practice sessions altogether and each session lasted about 20 minutes.

The same materials were used in the dictation sessions for Control Group 2 and Experimental Group, and they were from the same series of the listening textbook³ used for the pretest and the posttest. Materials used for the pretest and the posttest were excluded. Dialogues and monologues were alternately adopted. In the sessions dictation practices were given just in the way that the pretest and the posttest were conducted. The participants were asked to transcribe the missing word in each blank in the text (Appendix 2).

The speech rate of the materials used in the sessions was around 170 wpm. The participants were instructed to fill in the blanks in the dictation sheets while listening to the recordings. Pauses were inserted in the same way as they were in the pretest and the posttest and also the recordings were played only once. Each text used in the sessions was about 250 words long on average, where roughly 50 words were blanked out. In light of the purpose of this experiment, missing words included almost the same number of content and function words.

After listening to the recordings once for dictation, the participants were given an opportunity to listen to the same text again with pauses inserted, while at the same time given some explanatory comments on the cues and hints in perceiving spoken English. Following all these procedures, the scripts of the recordings were finally distributed and the participants corrected their mistakes on the dictation sheets. After that, the recordings were played one last time in order that the participants would be able to review their wrong guesses and incorrect recognitions on the sheet. The

participants were also instructed to pay close attention to the words they had missed or incorrectly recognized during this reviewing sessions.

The only difference between Control Group 2 and Experimental Group, however, was whether the Japanese translations were given to the participants before listening to the recordings or otherwise. The participants in Control Group 2 were given the translations as well as the English scripts after listening to the recordings, just to make sure what they heard and what it meant. No hints or background knowledge about the listening text they were going to hear was provided before the dictation.

To the participants in Experimental Group, on the other hand, Japanese translations were handed out before the dictation. They were asked to read them and understand the content that they were going to listen about. In addition, they were asked to guess about the English sentences they might hear in the recordings. Furthermore, every time a pause was inserted during the dictation session, they were asked to look at and read the next part of the translation and to make inferences on the next sentence or part of the sentence they would hear. However, while listening to the recordings, they were asked not to consult the translation but to pay close attention to the sound. This cycle was repeated until up to the end of the dictation session. Japanese translation was all that was given to the participants of Experimental Group before the session and they were given no hints or cues about English words and phrases they would hear beforehand.

5.2.5 Method of Analyses

Since the purpose of the experiment was to examine how different treatment between the groups affected their word recognition, the results

were first analyzed in terms of the total number of correct answers on the dictation sheets, followed by the separate analyses of content and function words. The data were analyzed using two-way mixed-design ANOVAs² (between-subjects-factor of group: Control 1/Control 2/Experimental, and within-subjects factor of time: pre/post). After the ANOVAs, chi-square tests² were also conducted in order to examine the differences in recognition of each word in the posttest between the groups.

As in Experiment 1, Quirk et al. (1985) was referred to in distinguishing content words from function words. In categorizing the words targeted for transcription, whether a word in question would be articulated with stress in the context or not was also taken into consideration.⁴ As a result, there were 55 content words and 45 function words for the target sections in the pretest, while, in the posttest, there were 56 content and 44 function words. All the data were computed into the respective percentage of correct word recognition.

5.3 Results

5.3.1 Results of ANOVA for Word Recognition in Total

Table 5.1 shows the descriptive statistics of correct word recognition in total (content and function words combined) in the pretest (Cronbach's alpha = .832) and in the posttest (Cronbach's alpha = .878). The results of a two-way mixed ANOVA for the total percentage of correct word recognition are shown in Table 5.2 and its graph in Figure 5.1.

Table 5.1.

Descriptive Statistics for the Total Percentage of Correct Word Recognition in the Pretest and the Posttest (n = 56)

Groups	n	Percentage for Correct Word Recognition in Total			
		Pretest		Posttest	
		M	SD	M	SD
Control Group 1	20	52.90	9.88	49.65	12.60
Control Group 2	16	48.50	11.15	45.81	12.88
Experimental Group	20	52.00	12.51	57.40	12.40

Table 5.2.

The Results of the Two-Way Mixed ANOVA for the Word Recognition in Total (n = 56)

Source	SS	df	MS	F	p	η_p^2
A: Group	1215.24	2	607.62	2.582	.085	.073
S: Error (A)	12472.75	53	235.33			
B: Time	2.49	1	2.49	0.053	.819	.000
Interaction (AB)	405.00	2	202.50	4.321	.018 *	.024
Error (BS)	2483.78	53	46.86			
Total	16581.49	111				

*: $p < .05$

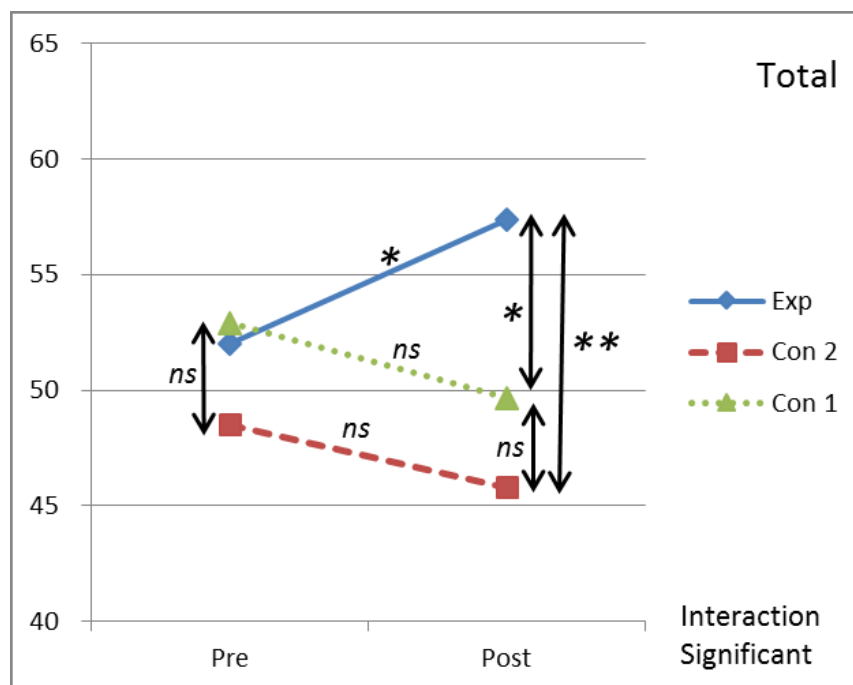


Figure 5.1. Means of total word recognition in percentage for the three groups in the pretest and the posttest (**: $p < .01$, *: $p < .05$).

The results of the ANOVA demonstrated that there was a significant interaction between the group and the time ($F(2, 53) = 4.321, p = .018$, partial $\eta^2 = .024$). Therefore, simple main effects of the group in the pretest and the posttest, and those of the time for three different groups were examined (Table 5.3; Figure 5.1). Since the simple main effect of the group in the posttest was found significant ($F(2, 73) = 4.537, p = .014$, partial $\eta^2 = .107$), a multiple comparison procedure using Tukey's method was performed to assess the differences between the three groups (Table 5.4).

Table 5.3.

Simple Main Effects of the Group in Both the Tests and Those of the Time for the Three Groups ($n = 56$)

	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
B: Time	A (Pretest)	339.82	2	169.91	1.204	.306	.029
	A (Posttest)	1280.43	2	640.21	4.537	.014 *	.107
	Error	10300.25	73	141.10			
A: Group	B (Control 1)	105.63	1	105.63	2.254	.139	.037
	B (Control 2)	12.50	1	12.50	0.267	.608	.004
	B (Experimental)	291.60	1	291.60	6.222	.016 *	.101
	Error	2483.78	53	46.86			

*: $p < .05$

Table 5.4.

The Results of Multiple Comparison Between Three Groups in the Posttest ($n = 56$)

Source	Group 1 (I)	Group 2 (J)	Difference (J - I)	<i>t</i>	<i>p</i>
A: Group (Posttest)	Control 1	Control 2	-3.84	1.362	.365
	Control 1	Experimental	7.75	2.918	.013 *
	Control 2	Experimental	11.59	4.113	.000 ***

***: $p < .001$, *: $p < .05$

The results showed that the simple main effect of the group in the pretest was not significant ($F(2, 73) = 1.204, p = .306$, partial $\eta^2 = .029$),

while in the posttest it was. In the posttest, the difference between Experimental Group and Control Group 2 ($t = 4.113, p = .000$) and between Experimental Group and Control Group 1 ($t = 2.918, p = .013$) were both significant, while there was no significant difference between the two control groups ($t = 1.362, p = .365$).

In addition, only the simple main effect of the time for Experimental Group was found significant ($F(1, 53) = 6.222, p = .016, \text{partial } \eta^2 = .101$), while the ones for the other two groups were nonsignificant (Control Group 1: $F(1, 53) = 2.254, p = .139, \text{partial } \eta^2 = .037$, Control Group 2: $F(1, 53) = 0.267, p = .608, \text{partial } \eta^2 = .004$).

These results show that the treatment the participants were given during the period between the two tests had positive effects only on those in Experimental Group.

5.3.2 Results of ANOVA for the Recognition of Content and Function Words

Table 5.5 shows the descriptive statistics of correct word recognition in percentage for the content and function words in the pretest and in the posttest. The results of two-way mixed ANOVAs for the percentage of correct word recognition for respective word categories are shown in Tables 5.6 and 5.7, and their graphs in Figure 5.2.

Table 5.5.
Descriptive Statistics for the Content and Function Word Recognition in the Pretest and the Posttest (n = 56)

Groups	n	Content Word Recognition (%)				Function Word Recognition (%)			
		Pretest		Posttest		Pretest		Posttest	
		M	SD	M	SD	M	SD	M	SD
Control Group 1	20	52.45	10.69	55.09	11.98	53.44	11.13	42.61	16.10
Control Group 2	16	48.75	11.03	49.67	13.03	48.19	13.25	40.63	14.39
Experimental Group	20	52.82	12.84	61.43	10.58	51.44	13.53	52.05	17.31

Table 5.6.

The Results of the Two-Way Mixed ANOVA for the Content Word Recognition (n = 56)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Group	1114.66	2	557.33	2.489	.093	.068
S: Error (A)	11867.98	53	223.92			
B: Time	454.99	1	454.99	9.078	.004 **	.028
Interaction (AB)	305.17	2	152.58	3.044	.056	.019
Error (BS)	2656.37	53	50.12			
Total	16456.52	111				

***p* < .01

Table 5.7.

The Results of the Two-Way Mixed ANOVA for the Function Word Recognition (n = 56)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Group	961.89	2	480.94	1.488	.235	.039
S: Error (A)	17127.20	53	323.15			
B: Time	974.81	1	974.81	10.196	.002 **	.039
Interaction (AB)	687.86	2	343.93	3.597	.034 *	.028
Error (BS)	5067.35	53	95.61			
Total	24791.48	111				

***p* < .01, **p* < .05

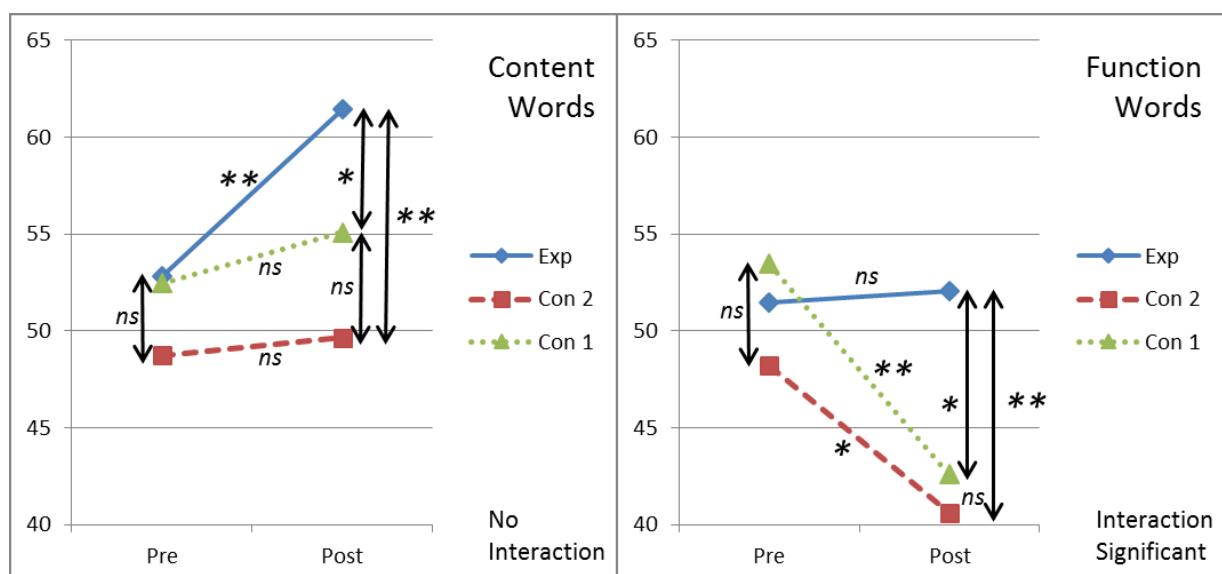


Figure 5.2. Means of content and function word recognition in percentage for the three groups in the pretest and the posttest (***p* < .01, **p* < .05).

The results of ANOVAs demonstrated that there was no significant interaction between the group and the time for content word recognition ($F(2, 53) = 3.044, p = .056, \text{partial } \eta^2 = .019$), while a significant interaction was found for recognition of function words ($F(2, 53) = 3.597, p = .034, \text{partial } \eta^2 = .028$).

First, the results of ANOVA for content words showed that only the main effect of the time was significant ($F(1, 53) = 9.078, p = .004, \text{partial } \eta^2 = .028$) and the main effect of the group was not ($F(2, 53) = 2.489, p = .093, \text{partial } \eta^2 = .068$). This implicates that the recognition of content words improved during the treatment period across all three groups and not specifically for Experimental Group and that the difference among the three groups was not significant. Even though there was no significant interaction found, the p value was around the borderline of 0.05, so that, the simple main effect of the group in the pre-test and the post-test and that of the time in the experimental and the control groups were computed, coupled with multiple comparison procedures using Tukey's method to assess the differences between the three groups in the posttest, and the results are added in the graph (Figure 5.2).

As can be seen in the graph, the simple main effect of the time was significant only for Experimental Group ($F(1, 53) = 14.792, p = .000, \text{partial } \eta^2 = .213$). In addition, the simple main effect of the group was significant in the posttest ($F(2, 76) = 4.546, p = .014, \text{partial } \eta^2 = .105$) between Control Group 1 and Experimental Group ($t = 2.422, p = .046$), as well as between Control Group 2 and Experimental Group ($t = 4.237, p = .000$), even though in the pretest it was not significant ($F(2, 76) = 0.635, p = .533, \text{partial } \eta^2 = .015$).

These results indicate that the treatment given to Experimental group

played a role in bettering the participants' recognition of content words and that this was not the case with the other two groups.

On the other hand, since the results of ANOVA for function words showed a significant interaction between the group and the time, simple main effects were computed. The simple main effect of the time for Control Group 1 ($F(1, 53) = 12.269, p = .001, \text{partial } \eta^2 = .175$) and Control Group 2 ($F(1, 53) = 4.794, p = .033, \text{partial } \eta^2 = .068$) was significant, while that for Experimental Group was not ($F(1, 53) = 0.038, p = .847, \text{partial } \eta^2 = .001$). This means that the recognition of function words by the two control groups significantly deteriorated from the pretest to the posttest, while the participants in Experimental Group recognized function words as correctly in the posttest as in the pretest.

Further, the simple main effect of the group in the posttest was significant ($F(2, 82) = 3.351, p = .040, \text{partial } \eta^2 = .075$), even though no significant difference was found in the pretest ($F(2, 82) = 0.589, p = .558, \text{partial } \eta^2 = .013$). The results of multiple comparison procedures (Tukey's method) showed that, in the posttest, the differences between Control Group 1 and Experimental Group ($t = 2.915, p = .013$) and between Control Group 2 and Experimental Group ($t = 3.328, p = .004$) were both significant, while no significant difference was found between the two control groups ($t = 0.580, p = .831$).

These results implicate that the difference in the treatment given to the three groups had some effects on the recognition of function words and only Experimental Group fared significantly better in recognizing function words correctly in the posttest than the other two groups.

5.3.3 Results of Chi-Square Tests for Recognition of Each Targeted Word in the Posttest

In order to thoroughly investigate what kind of effects the different treatments for the three groups had on the word recognition in the posttest, the total numbers of right (R) and wrong (W) transcription of each targeted word in the posttest (100 in total) for each group were computed and statistically analyzed.

Table 5.8.
Words in Which There Was Significant Difference in Recognition Between the Two Control and One Experimental Groups in the Posttest (n = 56)

Group	13 <i>you</i>		24 <i>else</i>		26 <i>machine</i>		27 <i>cycle</i>					
	R	W	R	W	R	W	R	W				
Control 1 (<i>n</i> = 20)	6	14	12	8	18	2	1	19				
Control 2 (<i>n</i> = 16)	6	10	13	3	8	8	5	11				
Experimental (<i>n</i> = 20)	15	5	20	0	19	1	11	9				
Standardised Residuals	Con 1	-2.033	2.033	-2.858	2.858	**	1.354	-1.354	-3.076	3.076	**	
	Con 2	-1.015	1.015	0.106	-0.106		-3.616	3.616	***	0.092	-0.092	
	Exp	2.990	-2.990	**	2.758	-2.758	**	2.056	-2.056	*	2.989	-2.989
χ^2 (<i>df</i> = 2)	9.140		10.148		13.236		11.833					
<i>p</i>	.010		*	.006	**	.001	**	.003	**			
Cramer's V	.404		.426		.486		.460					
Group	29 <i>that</i>		36 <i>for</i>		54 <i>its</i>		55 <i>central</i>					
	R	W	R	W	R	W	R	W				
Control 1 (<i>n</i> = 20)	8	12	4	16	1	19	3	17				
Control 2 (<i>n</i> = 16)	5	11	5	11	0	16	3	13				
Experimental (<i>n</i> = 20)	13	7	11	9	6	14	11	9				
Standardised Residuals	Con 1	-0.719	0.719	-1.829	1.829	-1.265	1.265	-1.863	1.863			
	Con 2	-1.440	1.440	-0.441	0.441	-1.789	1.789	-1.195	1.195			
	Exp	2.077	-2.077	*	2.245	-2.245	*	2.951	-2.951	**	2.989	-2.989
χ^2 (<i>df</i> = 2)	4.588		5.530		8.914		8.995					
<i>p</i>	.101		.063		.012		*	.011	*			
Cramer's V	.286		.314		.399		.401					
Group	83 <i>both</i>		94 <i>children's</i>		95 <i>There'll</i>		97 <i>Years</i>					
	R	W	R	W	R	W	R	W				
Control 1 (<i>n</i> = 20)	2	18	1	19	0	20	8	12				
Control 2 (<i>n</i> = 16)	2	14	1	15	1	15	5	11				
Experimental (<i>n</i> = 20)	7	13	5	15	4	16	14	6				
Standardised Residuals	Con 1	-1.354	1.354	-1.265	1.265	-1.746	1.746	-0.917	0.917			
	Con 2	-0.851	0.851	-0.894	0.894	-0.445	0.445	-1.607	1.607			
	Exp	2.156	-2.156	*	2.108	-2.108	*	2.166	-2.166	*	2.432	-2.432
χ^2 (<i>df</i> = 2)	4.684		4.457		5.117		6.186					
<i>p</i>	.096		.108		.077		.045		*			
Cramer's V	.289		.282		.302		.332					

Function words in italics.

***: $p < .001$, **: $p < .01$, *: $p < .05$

In examining in which targeted words on the posttest differences in recognition were found between the three groups, chi-square tests were conducted. Table 5.8 shows the results of the chi-square tests. The table lists those targeted words in which significant difference in recognition between the groups was found in terms of standardized residuals for Experimental Group. Experimental Group fared significantly better in twelve targeted words (six content and function words each) in the posttest than the other two groups.

5.4 Discussion

First, the results of the ANOVA for total word recognition showed that Experimental Group improved significantly more on their word recognition than the two control groups. This means that the treatment, in which Japanese translations were given before dictation practices and instructions were provided to make inferences from the translation on the text they would hear, had some effects in enhancing the learners' word recognition. In addition, the fact that significant differences were found not only between Control Group 1 and Experimental Group but also between Control Group 2 and Experimental Group, coupled with the fact that there was no significant difference between the two control groups, implies that treatment of simple dictation practices does not have any positive effects on learners' word recognition. The data obtained implicate that this significant difference observed resulted not from the dictation practices themselves but from the fact that the learners were informed of the content they would hear beforehand and instructed to guess about the sentences they would soon perceive.

In the dictation practices during the treatment, the participants of

Experimental Group had previous knowledge about the content they would hear, which might have worked more like a scaffolding in word recognition. In the posttest, however, they had no such information beforehand on the content. Nevertheless, they fared significantly better in recognizing words than the participants in the other two groups. These implicate that instructions to help learners pay more attention to the meanings and forms of what they would perceive activated some sorts of top-down strategies and enabled them to listen, utilizing such strategies, to the speech, even without a scaffolding of previous knowledge about the content.

Second, the results of the ANOVAs for content and function words indicated that these positive effects of the treatment on Experimental Group should hold true for recognition of both content and function words. However, a significant interaction between the groups was found only for recognition of function words and not for that of content words, which implicates that positive effects of the treatment specifically aimed at Experimental Group were even more pronounced in recognition of function words.

Further, given that the difference in the treatment between Experimental Group and the two control groups did not involve strategies related to bottom-up processing, the gaps found in the posttest between the groups in recognizing content and function words could presumably be due to some form of activation of top-down strategies.

It can be deduced, therefore, that Experimental Group's enhancement in recognition of content words resulted from their application of such top-down strategies as semantic and contextual inferences, which, during the treatment period, could have been fortified enough for them to make such inferences even without a scaffolding, since the participants of Control

Group 2, who were given simple dictation practices during the treatment period, were shown to make little progress on content word recognition.

It can also be assumed, therefore, that this increased amount of content words recognized by the participants of Experimental Group may have resulted in their increased reference to internalized linguistic knowledge of some sort, but not limited to grammatical and phrasal sort, which led to their fairly successful recognition of function words in the posttest, which was significantly better, compared with the other two groups.

Nevertheless, the results of chi-square tests showed that Experimental Group fared significantly better than the other two groups in as few as 12 words out of 100 targeted words in the posttest. Among them, No. 24 *else* in 'Is there anything else?' No. 29 *that* in 'in the machine that dries clothes,' No. 36 *for* in 'is famous for,' No. 54 *its* in 'burn its way through,' and No. 83 *both* in 'both business and industry' are examples in which significant differences could have been caused by activation of internalized grammatical and phrasal knowledge, which could have been possible only for the participants of Experimental Group.

However, only from these results, it cannot simply be deduced that the treatment given to Experimental Group resulted in the activation of internalized grammatical and phrasal knowledge, which eventually enhanced recognition of function words. They only implicate that the difference in the treatment caused significant differences in recognition of function as well as content words and that this difference is presumably due to whether the participants effectively utilized some form of top-down strategies or not.

Notes

1. Oller and Streiff (1975) state that expectancy grammar is a form of hypotheses that the listener will build about what is going to be articulated in the incoming speech, based on her grammatical, semantic, and pragmatic knowledge of the language. They say that dictation is a device which measures the efficiency of grammar-based expectancies and also that, if the listener's grammar of expectancy is incomplete, the kinds of hypotheses she will accept will 'deviate substantially from the actual sequences of elements in the dictation' (p. 34). They further claim that dictation activates the learner's internalized grammar of expectancy, which they assume is the central component of her language competence.
2. As to the analyses of the data, a Microsoft add-in software for Excel was used for the ANOVAs and for the chi-square tests.
3. The title of the listening textbook used in the experiment was *Kyukyoku-no-eigo-listening* (Ultimate English listening) series level 2 and 3, published by ALC Press. The materials for the pretest and the posttest were all from the level 2 book and those used in the treatment were adopted from both level 2 and 3 books.
4. For example, in the posttest, No. 11 *on* in 'You put the clothes in and turn it *on*' was judged to be articulated with stress. Therefore, it was categorized as a content word. On the other hand, No. 95 *There'll* in '*There'll* be a hot time in the old town tonight' was judged to be articulated without stress, so that it was categorized as a function word. See Appendices.

Chapter 6

Experiment 3

Examining Whether Providing Grammatical and Phrasal Knowledge Can Enhance Word Recognition

This chapter empirically examines whether it would be effective on word recognition to provide listeners with short-term grammatical and phrasal knowledge. The experiment was conducted at two different speech rates. The chapter discusses the effectiveness of fortified top-down processing on word recognition.

6.1 Introduction

In Chapter 5, it has been empirically demonstrated that listeners turn to top-down strategies after the treatment in which Japanese translations are given before listening to a speech and instructions are provided to make inferences from the meanings about the English they will hear in dictation sessions.

In order to make full use of top-down strategies, it is essential to possess sufficient amount of grammatical and phrasal knowledge, which underlies many of the formulaic sequences. In the case of learners with lower levels of proficiency, however, there is possibility that lack of grammatical and phrasal knowledge necessary to make inferences about missing information prevents them from using top-down strategies. If so, they might be able to improve word recognition by referring to strengthened top-down information, if they are given the necessary knowledge.

In this chapter, it will empirically be examined whether learners' word recognition will be enhanced, if they are given such grammatical and phrasal knowledge beforehand. For this purpose, an experiment was conducted.

6.2 Experiment

6.2.1 Purpose

The purpose of this experiment is to investigate whether Japanese EFL learners with lower levels of proficiency can enhance their spoken word recognition, if grammatical and phrasal knowledge is provided. In the experiment, only experimental groups were provided with treatment in which grammatical and phrasal knowledge was given.

The experiment was conducted at two different speech rates to examine how the difference in speech rate affects the influence of the treatment as a variable. In addition, to investigate if recognition of content words will be more, or less, enhanced than that of function words, which are usually pronounced weak and phonologically modified, those two word categories were separately analyzed.

6.2.2 Participants

The participants were 121 third-year and fourth-year students at a technical college in Japan who majored in engineering. Their L1 is Japanese. They ranged in their levels of English proficiency from an elementary to a lower-intermediate level; their mean score of the TOEIC (IP) was 393.30 ($SD = 105.34$).

6.2.3 Materials

A listening comprehension test was preliminarily conducted to assess the participants' level of proficiency in listening comprehension. For the test, the second and pre-second grade STEP listening tests consisting of 60 questions, 30 questions for each grade, were adopted.

For materials of the pretest and the posttest (Appendix 3), one dialogue and one monologue each for the pretest and the posttest, four different texts in total, were used. The texts were adopted from a listening textbook¹ in which all the texts consisted of the most basic 1,000 words. The speech rate of the pretest material (standard rate) was 157 wpm on average, 178 for the dialogue and 141 for the monologue, and for the posttest, 185 wpm on average, 181 for the dialogue and 190 for the monologue.² As to the difficulty level of the transcription tests, the texts were easy enough for the participants to understand if given the written scripts. All the vocabulary and syntactical structures used were at a junior- or basic senior-high school level in Japan.

As for handouts given to the participants to provide them with grammatical and phrasal knowledge related to the tests, they were custom-made for the present experiment, based on the pretest and the posttest.

6.2.4 Method

The method adopted for the pretest and the posttest was paused transcription, which was exactly the same as in Experiment 1. The pauses, which lasted about 10 seconds each, were inserted at irregular intervals in the spoken text and the participants were asked to transcribe the last four to five words they thought they heard before each pause. They listened to the recordings only once.

6.2.5 Procedure

First, the 121 participants were grouped into two, according to the speech rate, the standard-rate group ($n = 57$) and the slower-rate group ($n = 64$), who listened to the recordings made mechanically slower at 0.7 times the rate of the standard rate. Then, each group was divided into two, four in total, with one group experimental and the other control.³ The experimental groups for both speech rates were provided with the same treatment between the pretest and the posttest besides the normal classes. Table 6.1 shows the respective number of participants in each group and the speech rates in both tests.

Table 6.1.

Number of Participants in Each Group and the Speech Rates in the Pretest and the Posttest ($n = 121$)

	Average Speech Rate (wpm)		Number of Participants	
	Pretest	Posttest	Experimental	Control
Standard-Rate Group ($n = 57$)	157	185	24	33
Slower-Rate Group ($n = 64$)	110	130	33	31

Pauses were inserted in the same place in the same text across all these four groups in both the pretest and the posttest. They were discouraged from using *katakana* when they were unsure of the spellings, and asked to use alphabet letters that they thought they had heard. Pauses were inserted 16 times, 8 each for the dialogue and the monologue. Judgment of whether the participants' handwritten responses were accurate or otherwise was limited to the last four words before each pause. Therefore, 64 items, 4 each for every pause, were the maximum accurate responses possible for both the pretest and the posttest. Sections of recordings targeted for transcription in the posttest are shown in Table 6.2.

Table 6.2.
Sections of Recording Targeted for Transcription in the Posttest

1	<i>are no</i> more tickets	9	<i>As you all</i> know
2	telling <i>me about it</i>	10	ordered <i>from</i> ABC company
3	giving <i>me</i> great help	11	excuse <i>for</i> arriving late
4	<i>you'll</i> enjoy <i>the</i> show	12	business plan <i>is</i> off
5	<i>For the</i> best seats	13	<i>will all</i> miss <i>her</i>
6	close <i>to the</i> musicians	14	taking <i>over her</i> job
7	<i>if you are</i> late	15	<i>from our</i> parent company
8	<i>them to</i> someone else	16	<i>to eat and</i> drink

Function words are in italics.

Nos. 1 to 8 are the sections from the dialogue and 9 to 16 from the monologue.

Each of the 64 (28 content and 36 function) words in the pretest was graded either correct or incorrect, while, as for the posttest, the word *ABC* of No. 10 was excluded from grading because it was a proper noun. That left 63 (31 content and 32 function) words for the posttest.⁴ In grading, if the word boundaries were breached, all the items involved were judged to be incorrect. However, if the sound was recognized correctly, the item was judged to be correct, even if it was misspelled. All the data were computed into the percentage of accurate word recognition, with the number of items correctly recognized being the numerator and the total number of items targeted for transcription the denominator.⁵

6.2.6 Treatment

The experimental groups for both rates were provided with treatment besides normal classes, starting one week after the pretest. The treatment lasted two weeks. After the treatment, another two weeks were set before the posttest. During the period, the control groups for both rates were given only the normal classes so that the following treatment was the only difference between the experimental and control groups.

Grammatical and phrasal knowledge related to the posttest as well as

the pretest was put into 93 English short sentences (Appendix 4). Some of them shared the contexts or the phrases in the pretest and the posttest and some of the words used in the sentences were identical with those used in the tests. Others contained only the key expressions, idioms, or grammatical items used in a different context.⁶ These were randomly mixed into a single handout.

Grammatical and phrasal knowledge contained in the sentences covered the whole script of the tests, not limited to the sections targeted for transcription. Accordingly, the participants in the experimental groups had not been aware of which of the sentences in the handout were related to the sections targeted in the posttest. In addition, half of the sentences in the handout were related to the pretest so that the items contained in them naturally must have had no direct significance in the posttest. All the sentences were preceded by Japanese translation.

The handout was given and explicitly explained. The participants were asked to repeat after the model reading and also to read the sentences aloud many times by themselves. They were also told to read them aloud at home and memorize them.

6.3 Results

6.3.1 Listening Comprehension Test

Table 6.3 shows the descriptive statistics of the preliminary listening comprehension test (Cronbach's alpha = .810). A one-way between-subjects-design ANOVA⁷ was conducted and no significant difference was found between the four groups ($F(3, 120) = 0.684$, $p = .564$, $\eta_p^2 = .001$). It has been confirmed that the four groups have the same level of proficiency in listening comprehension.

Table 6.3.

Descriptive Statistics of the Preliminary Listening Comprehension Test

Groups		<i>n</i>	<i>M</i>	<i>SD</i>
Standard Rate	Experimental	24	29.88	7.18
	Control	33	30.21	8.31
Slower Rate	Experimental	33	32.52	9.09
	Control	31	30.48	7.04

6.3.2 The Paused Transcription Test

Table 6.4 shows the descriptive statistics of the pretest (Cronbach's alpha = .829) and the posttest (Cronbach's alpha = .885).

Table 6.4.

Descriptive Statistics of the Pretest and the Posttest (Paused Transcription)

Groups	<i>n</i>	Pretest (Words Recognized, %)						Posttest (Words Recognized, %)						
		Content		Function		Total		Content		Function		Total		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Standard	Experimental	24	54.61	16.99	19.44	11.65	34.83	12.94	43.95	14.89	18.93	9.72	31.22	11.44
	Control	33	50.87	11.33	21.38	8.22	34.28	8.64	39.20	11.46	19.04	10.73	28.96	10.64
Slower	Experimental	33	61.91	9.50	30.30	10.56	44.13	9.25	56.79	14.98	34.38	12.99	45.41	12.85
	Control	31	60.02	13.44	25.99	13.08	43.00	11.71	46.20	12.05	25.91	12.73	35.89	11.99

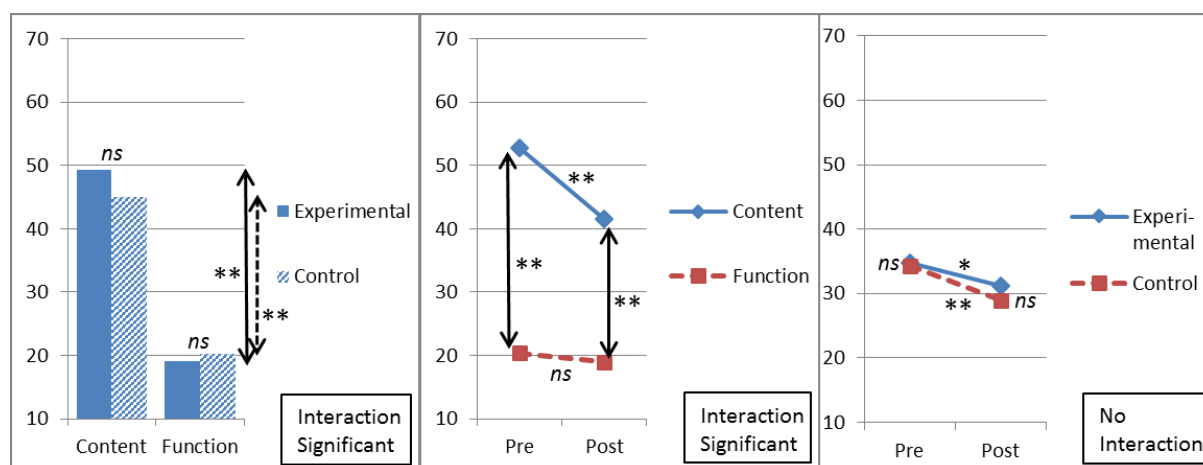


Figure 6.1. Words successfully recognized in percentage at the standard speech rate (pre: 157 wpm, post: 185 wpm, $**p < .01$, $*p < .05$).

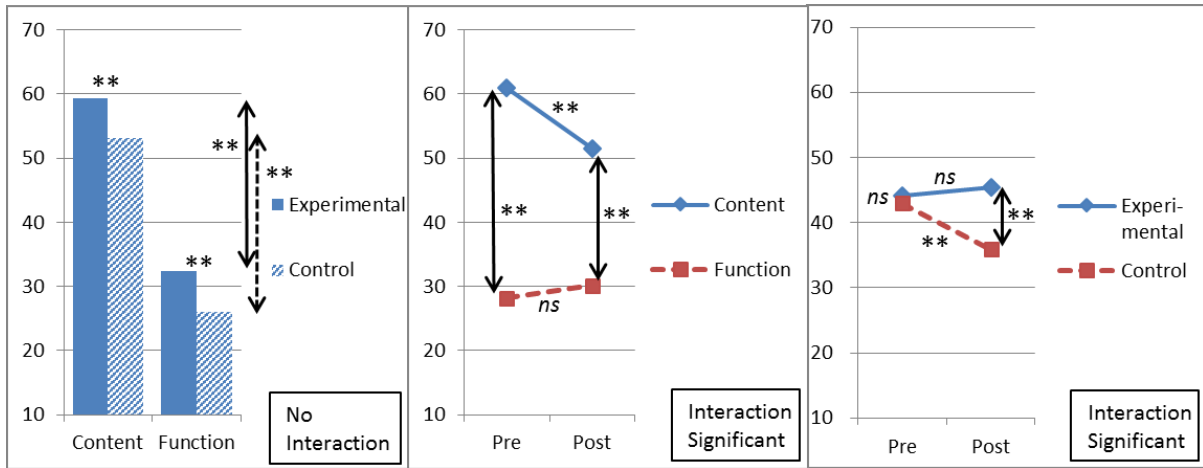


Figure 6.2. Words successfully recognized in percentage at the slower speech rate (pre: 110 wpm, post: 130 wpm, **: $p < .01$).

A three-way mixed ANOVA⁷ (groups: experimental/control, word categories: content/function, time: pre/post) was conducted for each speech rate. The results are shown in Tables 6.5 and 6.6. Figures 6.1 and 6.2 show respectively the means of correct word recognition in percentage for each two of the three factors at the standard and the slower speech rate.

Table 6.5.

The Results of the Three-Way ANOVA (Standard Speech Rate, $n = 57$)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	144.58	1	144.58	0.342	.561	.002
S: Error (A)	23225.89	55	422.29			
B: Word Categories	41911.82	1	41911.82	801.590	.000 ***	.535
Interaction (AB)	386.40	1	386.40	7.390	.009 **	.005
Error (BS)	2875.72	55	52.29			
C: Time	2204.64	1	2204.64	28.730	.000 ***	.028
Interaction (AC)	27.84	1	27.84	0.363	.549	.000
Error (CS)	4220.57	55	76.74			
Interaction (BC)	1316.33	1	1316.33	35.400	.000 ***	.017
Second-Order Interaction (ABC)	2.33	1	2.33	0.063	.803	.000
Error (BCS)	2045.16	55	37.18			
Total	78361.27	227				

***: $p < .001$, **: $p < .01$

Table 6.6.

The Results of the Three-Way ANOVA (Slower Speech Rate, n = 64)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	2549.98	1	2549.98	5.644	.021 *	.027
S: Error (A)	28011.43	62	451.80			
B: Word Categories	46910.95	1	46910.95	758.197	.000 ***	.503
Interaction (AB)	0.39	1	0.39	0.006	.937	.000
Error (BS)	3836.05	62	61.87			
C: Time	891.24	1	891.24	9.439	.003 **	.010
Interaction (AC)	660.95	1	660.95	7.000	.010 *	.007
Error (CS)	5853.97	62	94.42			
Interaction (BC)	2101.49	1	2101.49	55.392	.000 ***	.023
Second-Order Interaction (ABC)	83.12	1	83.12	2.191	.144	.001
Error (BCS)	2352.18	62	37.94			
Total	93251.74	255				

***: $p < .001$, **: $p < .01$, *: $p < .05$

As for the standard speech rate groups, significant interactions were found between groups and word categories ($F(1, 55) = 7.390$, $p = .009$, $\eta_p^2 = .005$) as well as word categories and time ($F(1, 55) = 35.400$, $p = .000$, $\eta_p^2 = .017$), so that respective simple main effects were examined. As for groups and word categories, the effects of word categories in both the experimental ($F(1, 55) = 481.456$, $p = .000$, $\eta_p^2 = .557$) and the control ($F(1, 55) = 327.524$, $p = .000$, $\eta_p^2 = .379$) group were significant, while neither difference between the two groups in content ($F(1, 110) = 2.115$, $p = .149$, $\eta_p^2 = .019$) nor in function ($F(1, 110) = 0.123$, $p = .727$, $\eta_p^2 = .001$) words was significant.

As for the interaction between word categories and time, simple main effects of words in both the pretest ($F(1, 110) = 649.190$, $p = .000$, $\eta_p^2 = .603$) and the posttest ($F(1, 110) = 317.120$, $p = .000$, $\eta_p^2 = .295$) were significant. Additionally, recognition percentage for content words significantly decreased from the pretest to the posttest ($F(1, 110) = 60.814$, $p = .000$, $\eta_p^2 = .354$). However, recognition of function words remained unchanged ($F(1, 110) = 1.000$, $p = .320$, $\eta_p^2 = .006$). This means that the gap in successful recognition between content and function words became smaller in the

posttest than in the pretest.

On the other hand, no significant interaction was found between groups and time ($F(1, 55) = 0.363, p = .549, \eta_p^2 = .000$). The fact that the main effect of groups was not significant ($F(1, 55) = 0.342, p = .561, \eta_p^2 = .002$) and that the one of time was significant ($F(1, 55) = 28.730, p = .000, \eta_p^2 = .028$) means that, as is evident from the graph, word recognition percentage significantly decreased for both experimental and control groups from the pretest to the posttest and that the difference in recognition between the two groups was unchanged.

As for the slower speech rate groups, there was a significant interaction between groups and time ($F(1, 62) = 7.000, p = .010, \eta_p^2 = .007$), which is evident from the graph. Simple main effects were computed and the following results were obtained. No significant difference was found between the experimental and control groups in the pretest ($F(1, 124) = 1.125, p = .291, \eta_p^2 = .008$). In the posttest, however, the difference was significant ($F(1, 124) = 10.632, p = .001, \eta_p^2 = .078$). In addition, even though the effect of time was not significant in the experimental group ($F(1, 62) = 0.091, p = .764, \eta_p^2 = .001$), it was significant in the control group ($F(1, 62) = 16.348, p = .000, \eta_p^2 = .208$). These results imply some positive effects of the treatment on the word recognition by the experimental group.

A significant interaction between word categories and time was found ($F(1, 62) = 55.392, p = .000, \eta_p^2 = .023$), also for the slower speech rate groups. Simple main effects were examined. Effects of word categories in the pretest ($F(1, 124) = 690.012, p = .000, \eta_p^2 = .624$) and the posttest ($F(1, 124) = 292.101, p = .000, \eta_p^2 = .264$) were both significant, and recognition percentage for content words significantly decreased from the pretest to the posttest ($F(1, 124) = 43.291, p = .000, \eta_p^2 = .256$). However, the

percentage for function words remained unchanged ($F(1, 124) = 1.931$, $p = .167$, $\eta_p^2 = .011$). This means that, as is the case with the standard rate groups, the gap in successful recognition between content and function words for the slower rate group also became smaller in the posttest than in the pretest.

6.3.3 Two-Way ANOVAs for Content and Function Words

In order to more thoroughly analyze how the treatment made differences between the experimental and control groups in the posttest, successful recognition of content words and that of function words were separately examined, using two-way ANOVAs⁷. The results of mixed two-way ANOVAs (between-subjects factor of groups and within-subjects factor of time) are shown in graphs (Figure 6.3) and tables (Table 6.7).

Table 6.7.
The Results of the Two-Way ANOVAs

Content-Word Recognition (Standard Speech Rate, $n = 57$)							Function-Word Recognition (Standard Speech Rate, $n = 57$)								
Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2		
A: Groups	502.00	1	502.00	1.621	.208	.020	A: Groups	30.34	1	30.34	0.184	.670	.003		
S: Error (A)	17037.27	55	309.77				S: Error (A)	9067.65	55	164.87					
C: Time	3463.80	1	3463.80	50.414	.000	***	.139	C: Time	58.86	1	58.86	1.300	.259	.005	
Interaction	7.03	1	7.03	0.102	.750	.000	Interaction	22.07	1	22.07	0.488	.488	.002		
Error (CS)	3778.91	55	68.71				Error (CS)	2489.37	55	45.26					
Total	24928.30	113					Total	11682.03	113						
***: $p < .001$															
Content-Word Recognition (Slower Speech Rate, $n = 64$)							Function-Word Recognition (Slower Speech Rate, $n = 64$)								
Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2		
A: Groups	1243.51	1	1243.51	4.982	.029	*	.050	A: Groups	1306.40	1	1306.40	4.947	.030	*	.062
S: Error (A)	15475.85	62	249.61					S: Error (A)	16371.78	62	264.06				
C: Time	2864.62	1	2864.62	35.459	.000	***	.114	C: Time	127.46	1	127.46	2.471	.121	.006	
Interaction	606.34	1	606.34	7.505	.008	**	.024	Interaction	137.67	1	137.67	2.669	.107	.007	
Error (CS)	5008.83	62	80.79					Error (CS)	3198.06	62	51.58				
Total	25120.10	127						Total	21149.92	127					
***: $p < .001$, **: $p < .01$, *: $p < .05$							*: $p < .05$								

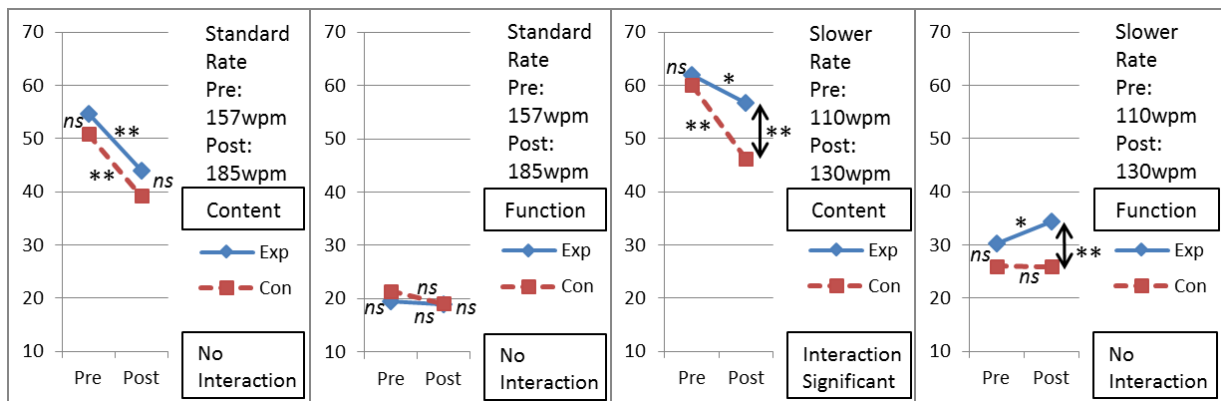


Figure 6.3. Content and function words successfully recognized in percentage for both standard and slower speech rate groups (**: $p < .01$, *: $p < .05$).

First, as for recognition of content words by the standard rate groups, the interaction ($F(1, 55) = 0.102, p = .750, \eta_p^2 = .001$) and the main effect of groups ($F(1, 55) = 1.621, p = .208, \eta_p^2 = .065$) were not significant. However, the main effect of time was significant ($F(1, 55) = 50.414, p = .000, \eta_p^2 = .447$). Second, as for recognition of function words by the standard rate groups, none of the interaction ($F(1, 55) = 0.488, p = .488, \eta_p^2 = .008$), the main effect of groups ($F(1, 55) = 0.184, p = .670, \eta_p^2 = .012$), and the main effect of time ($F(1, 55) = 1.300, p = .259, \eta_p^2 = .023$) were significant. These results signify that, as for the standard speech rate groups, no effect of treatment on word recognition, content or function, was found.

As for recognition of content words by the slower rate groups, however, the interaction was significant ($F(1, 62) = 7.505, p = .008, \eta_p^2 = .062$). Simple main effects were computed and the following results were obtained. No significant difference was found between the experimental and control groups in the pretest ($F(1, 98) = 0.343, p = .560, \eta_p^2 = .003$). In the posttest, however, the experimental group's recognition percentage was significantly higher than that of the control group ($F(1, 98) = 10.855, p = .001, \eta_p^2 = .099$). Additionally, the simple main effects of time in the experimental group ($F(1,$

62) = 5.335, $p = .024$, $\eta_p^2 = .051$), and in the control group ($F(1, 62) = 36.650$, $p = .000$, $\eta_p^2 = .352$) were both significant.

Finally, as for recognition of function words by the slower rate groups, the interaction ($F(1, 62) = 2.669$, $p = .107$, $\eta_p^2 = .029$) and the main effect of time ($F(1, 62) = 2.471$, $p = .121$, $\eta_p^2 = .027$) were not significant, and only the main effect of groups was significant ($F(1, 62) = 4.947$, $p = .030$, $\eta_p^2 = .274$). For reference, simple main effects of groups in the pretest and the posttest and those of time in the experimental and control groups are shown in Figure 6.3.⁸

These results implicate that, as for the slower speech rate groups, there were some effects of treatment on word recognition, especially recognition of content words.

6.3.4 Fisher's Exact Tests

In order to examine the effects of the treatment in detail, Fisher's exact tests⁷ were conducted. The numbers of right (R) and wrong (W) transcriptions by the experimental and control groups of each rate for all the 63 words in the posttest were computed and which of the 63 words in the posttest made a significant difference in recognition between the experimental and control groups was examined.

Results are shown in Tables 6.8 and 6.9. Significant differences were found in 7 words for the standard rate groups and 11 words for the slower rate groups. Out of 7 words for the standard rate groups, recognition of the word *off* in No. 12 was significantly better in the control group. However, all the other words listed have significantly higher percentage in recognition by the experimental group.

Table 6.8.

Words in Which There Was Significant Difference in Recognition Between the Experimental and the Control Groups (Standard Speech Rate, n = 57)

Group	<i>7 late</i>		<i>9 as</i>		<i>11 excuse</i>		<i>11 for</i>		<i>11 late</i>		<i>12 off</i>		<i>16 drink</i>	
	R	W	R	W	R	W	R	W	R	W	R	W	R	W
Experimental (<i>n</i> = 24)	10	14	16	8	8	16	5	19	6	18	2	22	17	7
Control (<i>n</i> = 33)	5	28	10	23	2	31	0	33	0	33	11	22	8	25
<i>p</i>	.035 *		.008 **		.012 *		.010 *		.004 **		.030 *		.001 **	
Cramer's V	.297		.360		.354		.364		.402		.294		.464	

Function words in italics.

***p* < .01, **p* < .05

Table 6.9.

Words in Which There Was Significant Difference in Recognition Between the Experimental and the Control Groups (Slower Speech Rate, n = 64)

Group	<i>1 are</i>		<i>1 no</i>		<i>6 close</i>		<i>7 late</i>		<i>8 someone</i>		<i>8 else</i>		<i>9 you</i>	
	R	W	R	W	R	W	R	W	R	W	R	W	R	W
Experimental (<i>n</i> = 33)	21	12	33	0	10	23	18	15	20	13	16	17	24	9
Control (<i>n</i> = 31)	7	24	27	4	2	29	8	23	5	26	5	26	13	18
<i>p</i>	.001 **		.050 *		.023 *		.024 *		.000 **		.008 **		.022 *	
Cramer's V	.414		.266		.305		.292		.456		.344		.312	

Group	<i>14 taking</i>		<i>16 eat</i>		<i>16 and</i>		<i>16 drink</i>	
	R	W	R	W	R	W	R	W
Experimental (<i>n</i> = 33)	27	6	12	21	14	19	30	3
Control (<i>n</i> = 31)	17	14	1	30	2	29	20	11
<i>p</i>	.030 *		.001 **		.001 **		.015 *	
Cramer's V	.291		.412		.415		.319	

Function words in italics.

***p* < .01, **p* < .05

6.4 Discussion

First, as for the standard speech rate groups, the results of the three-way ANOVA showed that there was no interaction between groups and time, that recognition percentage for both groups had been significantly lower in the posttest than in the pretest, and that the recognition gap between the experimental and control groups had not been changed. These results indicate that there was no effect of treatment. As for the difference in recognition between content and function words, the gap closed both for the experimental and control groups in the posttest. However, the difference was still significant.

Second, as for the slower speech rate groups, there was a significant interaction between groups and time, and even though there was no significant difference in word recognition between the experimental and control groups in the pretest, the difference was significant in the posttest. This signifies that there were some effects of the treatment on the experimental group. Additionally, the gap in recognition between content and function words closed both for the experimental and control groups in the posttest. However, the recognition of function words was still more difficult than that of content words for the slower speech rate groups as well.

These results indicate that, at the average speech rate of 185 wpm, no effect of treatment was found, while at the average rate of 130 wpm, there were some effects. However, it can be said that the closing of the gap in recognition by both speech rate groups between content and function words in the posttest has little to do with the treatment, because this is not only the case with the experimental groups, but also with the control groups. This may be attributable to the content of the posttest listening materials and their phonetic properties.

In addition, word recognition became poorer across all the groups except for the slight gain logged by the experimental group of the slower speech rate. This is presumably due mainly to the rate gap between the pretest and the posttest, which in turn implicates that speech rates play a very important role in spoken word recognition.

Next, the results of the two-way ANOVAs indicate that there was no effect of the treatment either on content or function words for the standard rate groups. For the slower rate groups, however, there was a significant interaction only for content words and the recognition gap between the two

groups, which had not existed in the pretest, emerged in the posttest. There was a similar tendency for function words as well, even though no interaction was found, which means that effects of the treatment was less profound on the recognition of function words.

Furthermore, the results of Fisher's exact tests show that recognition by the experimental groups was significantly better than by the control groups in 6 (4 content and 2 function) words at the standard speech rate and 11 (7 content and 4 function) words at the slower speech rate. Only *late* in No.7 and *drink* in No.16 were shared across both rates.

What was distinctive for the standard rate groups was that there was significant difference in three of the four words in No.11 *excuse for arriving late; excuse, for, and late*. The sentence 'You have no excuse for being late' was in the handout and this knowledge might have intervened in the transcription. There was no *arriving* in the handout, which made no difference in the recognition of the word. The difference in the other three words presumably resulted from a successful matching only by the experimental group of some additional intervention from top-down processing and phonetic information through bottom-up processing.

These results implicate that, as far as lower-proficiency listeners are concerned, perfect matching might be a necessity for a difference to be made in recognition of the words at least at the average speech rate of 185 wpm. The sight of *arriving* in the written script would have certainly resulted in successful recognition. However, with this speech rate, the mere perception of the sound would not have been enough even with the help of additional knowledge. Additionally, the reason why no significant difference was made in these words for the slower speech rate groups would presumably be that, at the average rate of 130 wpm, bottom-up processing

combined with top-down processing without the help of additional knowledge would have been enough in recognizing these words.

As for the slower speech rate groups, on the other hand, what was distinctive was significant differences found in the words *are* and *no* in No.1 *are no more tickets, someone* and *else* in No.8 *them to someone else*, and *eat, and, and drink* in No.16 *to eat and drink*. There were sentences in the handout ‘There are no more vacant seats available for the musical,’ ‘Don’t give this information to anyone else,’ and ‘It’s time to eat and drink.’ There were words *are, no, else, eat, and* and *drink* in the handout. However, no *someone* appeared. Nevertheless, at the average rate of 130 wpm, a difference was made because only the participants in the experimental group have enough information from top-down processing to lead them to successful recognition of the word *someone*, given the level of information obtained through the bottom-up processing possible at this rate.

Finally, we discuss the probable reasons why the effects of the treatment were found more evidently on the recognition of content words than on that of function words even at the slower speech rate. Given the additional grammatical and phrasal knowledge, the experimental group had more information to turn to, which might have triggered more activated top-down intervention. However, their ability to recognize and process phonetic information of function words through the bottom-up process could barely reach the extended hand from the top-down processing, even at the slower rate of 130 wpm. In other words, their bottom-up processing scarcely reached the threshold conditioned by the fortified top-down processing. Therefore, the interaction between the top-down and bottom-up processing could have been only marginally successful for the slower-rate experimental group.

On the other hand, in the case of content words, which mostly contain stressed syllables, phonetic information obtained by the experimental group reached the threshold required by the fortified top-down processing. The control group did not have such fortification and they must have had higher level of threshold for the bottom-up processing to reach. Consequently, the difference in recognition between the groups emerged.

The same logic can be applied to the standard speech rate groups as well. At the average rate of 185 wpm, phonetic information not only of the function words but also of the content words obtained through the bottom-up processing did not reach the lowered threshold of the experimental group. Therefore, there was no effect of the treatment on recognition.

These results indicate that, for learners with lower levels of proficiency, recognition of function words is challenging, even with the fortified top-down processing (additional grammatical and phrasal knowledge) and with the fortified bottom-up processing (the slower speech rate of 130 wpm). In addition, if the speech rate is around 185 wpm, even the content words are difficult to recognize, even with the help of fortified top-down processing, because phonetic information obtained through the bottom-up processing is not enough.

If the speech rate were to be lowered further down, or if the participants were to be given grammatical and phrasal knowledge repeatedly for a longer period, there might appear some effects of treatment on recognition.

6.5 Implications

Thus, when the speech rate was fast enough (around 185 wpm), no effects of additional grammatical and phrasal knowledge on word

recognition were found. However, when the speech rate was lowered down to around 130 wpm, there were effects. The present study did not clarify at which rate between 130 and 185 wpm the threshold falls. Nevertheless, it has been confirmed that the difference in speech rates brings about different effects on word recognition when grammatical and phrasal knowledge is given.

From these results of the experiment, the following implications can be drawn.

1. It is necessary to develop a pedagogical method in which perception of phonetic information can effectively be processed through the bottom-up processing.
2. In teaching how to listen, speech rate is an important variable.

The results of this experiment indicates that, if learners do not have skills to deal with a moderately-high speech rate in which a speaker articulates words in a natural stress-timed manner, they have difficulty recognizing words even with extended helping hand from the top-down information armed with additional grammatical and phrasal knowledge. If learners cannot recognize sufficient content words, they presumably have very little recourse left available and find themselves unable to take advantage of the knowledge they already have.

The next two chapters focus on reinforcement of the bottom-up processing in spoken word recognition: first on speech rate and second on English phonological features including phonemes, syllable structures, and the stress-timed rhythm.

Notes

1. The title of the listening textbook used in the experiment was

Kyukyoku-no-eigo-listening (Ultimate English listening) series level 1, published by ALC Press.

2. Materials used in the pretest were a dialogue of 342 words with the recording time of 115 seconds and a monologue of 382 words in 162 seconds, which make 724 words in 277 seconds in total. In the posttest, a dialogue of 292 words in 97 seconds and a monologue of 266 words in 84 seconds, which make 558 words in 181 seconds in total, were used. In terms of wpm, there was a difference in rates between the two tests, which presumably led to the lower recognition percentage in the posttest. However, all the materials used were from the most difficult level of the three levels available in the listening textbook, which meant the fastest. The materials were used to test participants' accuracy in spoken word recognition so that the ones with simple vocabulary but high speech rates were adopted.
3. Grouping of the participants was based on the classes that they belong to. Consequently, the number of the participants in each group was not equal.
4. As for the distinction between content and function words, Quirk et al. (1985) was referred to.
5. The way the transcription of each targeted word was judged to be correct or otherwise was in conformity with the method used in Experiment 1.
6. Example sentences in the handout are 'For good seats, the tickets are 30 euros each.' for 'For best seats, they're \$30 each.' in the posttest, 'The convenience store is close to the station.' for 'I like to be close to the musicians.' in the posttest, and 'If you are late, they'll be angry.' and 'Don't give this information to anyone else.' for '..., because if you are

late, they may sell them to someone else.' in the posttest.

7. In analyzing the data, an online software of ANOVA 4 was used for the three-way ANOVAs and a Microsoft add-in software for Excel was used for the two-way and one-way ANOVAs and for the Fisher's exact tests.
8. The results of the simple main effects' analyses on function words for the slower speech rate groups were as follows. The simple main effect of groups in the pretest was $F(1, 85) = 1.888$, $p = .173$, $\eta_p^2 = .020$. The simple main effect of groups in the posttest was $F(1, 85) = 7.262$, $p = .009$, $\eta_p^2 = .077$. The simple main effect of time in the experimental group was $F(1, 62) = 5.304$, $p = .025$, $\eta_p^2 = .079$. The simple main effect of time in the control group was $F(1, 62) = 0.002$, $p = .966$, $\eta_p^2 = .000$.

Chapter 7

Experiment 4

Examining the Effects of Compressed Speech Rates on Spoken Word Recognition

This chapter empirically examines whether it would be effective on word recognition to use mechanically compressed recordings in listening class. The participants of the experiment listened to conversations and sentences in a textbook at four different speech rates for half a year. The effects of this treatment on word recognition will be discussed.

7.1 Introduction

In Chapters 4 and 6, it has been empirically demonstrated that speech-rate-related stimulus packages of some sorts may be necessary to fortify bottom-up processing, which will presumably lead to enhanced spoken word recognition, and it is a tempting leap to suggest that repeated sessions of high-speech-rate listening may affect the learners' spoken word recognition positively.

This is because, as has been discussed in Chapter 2, there is enough experimental as well as anecdotal evidence that listeners can normalize a faster speech rate as their new baseline, if they are exposed to it continuously. In other words, there is a high possibility that, with continuous exposure to a faster speech rate, listeners find it easier to recognize words in a speech delivered at the original baseline rate because it sounds less fast.

In addition, there is also some empirical evidence, as was discussed in

Chapter 2, that slowing down speech rates has some positive effects on L2 listeners' comprehension. Since speech rate is most likely to have some effects on the first phase of listening, perception, through bottom-up processing, improvement in comprehension must very likely result from enhanced word recognition.

Hence, if listeners can recalibrate their baseline speech rate to a faster one after steady practices using a higher rate of speech, then this cognitive recalibration presumably works positively on their word recognition, because it enables listeners to perceive the sound at the original baseline rate as slower. However, empirical evidence concerning the effects of speech rate manipulation on L2 listeners is scarce.

In this chapter, it is examined whether Japanese EFL learners with lower levels of proficiency can adapt to faster speech rate in the perceptual phase and how long-term training sessions in which they continue to listen to various rates of mechanically compressed speech affect their spoken word recognition.

7.2 Experiment

7.2.1 Purpose

The purpose of this experiment is to investigate how repeated sessions of listening practices in which mechanically compressed speech is used affect spoken word recognition at the baseline speech rate by Japanese EFL learners with lower levels of proficiency. In the experiment, four different speech rates were used for half a year between the pretest and the posttest.

The aims of the experiment are twofold. One is to investigate if steady exposure to a faster speech rate is effective in improving word recognition by Japanese EFL listeners with lower levels of proficiency and the other is

to examine if there is any specific speech rate at which lower-proficiency listeners fare best in word recognition after the exposure to that rate of speech.

7.2.2 Participants

The participants were 206 first-year technical college students in Japan who majored in engineering. Their L1 is Japanese. The participants ranged in their levels of English proficiency from a beginner to an elementary level; their mean score of the TOEIC bridge test was 108.88 ($SD = 10.72$).

7.2.3 Materials

The experiment had a pretest-treatment-posttest design. Prior to the experiment, however, a listening comprehension test was preliminarily conducted to assess the participants' level of proficiency in listening comprehension. For the test, a pre-second grade STEP listening test consisting of 30 questions was adopted.

For the pretest and the posttest (Appendix 5), word recognition tests, which had a cloze-test format, were conducted. For these tests as well, pre-second grade STEP listening tests, which were different from the one used for the listening comprehension test, were adopted. Materials for the posttests were different from the ones for the pretests. However, they shared the same format.

The word recognition tests were transcription tests and participants were required to write down one word in each blank, which they thought they had heard. For both the pretest and posttest, five sections each for the dialogue and the monologue part of the STEP test, ten in total, were

extracted. In these ten sections, 25 words each in the dialogue and the monologue, 50 in total, were blanked out. Of the 50 words which were blanked out, 25 were content words with the rest being function words¹.

All the 50 blanked-out words in both the pretest and the posttest were at a junior-high school level in Japan. The participants would have had little difficulty in recognizing them, if those words had been in written scripts or articulated individually.

As to the material used in the treatment, an English textbook² authorized by MEXT was adopted, because this was the only and main textbook that participants used in class (Appendix 6).

7.2.4 Procedure

First, 206 participants were divided into four groups: one control group (Control Group) and three experimental groups (Experimental Groups 1, 2, and 3), who listened to the treatment materials at different speech rates. Participants in Control Group listened at the original speech rate, while those in the experimental groups at the mechanically compressed speech rates. Experimental Group 1 listened at 1.2 times the original rate, Experimental Group 2 at 1.5 times, and Experimental Group 3 at 2.0 times. The 206 participants were made up of five classes and the groups were divided according to the classes that they belonged to. Consequently, two of the five classes belonged to Experimental Group 2, while the other groups were each made up of one class.

The results of the listening comprehension test, which was conducted to assess the participants' listening proficiency preceding the experiment, were as follows: Control Group ($n = 42$, $M = 10.24$, $SD = 3.71$), Experimental Group 1 ($n = 41$, $M = 11.02$, $SD = 4.96$), Experimental Group 2 ($n = 83$, $M =$

11.11, $SD = 3.39$), and Experimental Group 3 ($n = 40$, $M = 11.78$, $SD = 3.93$). The result of one-way between-subjects-design ANOVA³ showed that there was no significant difference between the four groups in terms of listening proficiency ($F(3, 202) = 1.068$, $p = .364$, $\eta_p^2 = .001$).

In compressing the speech rates, a speed-changing software⁴ was used. Table 7.1 shows the respective number of participants in each group and the average speech rates⁵ of the pretest, the posttest, and treatment materials, listened by four groups, in wpm. The participants listened to all the dialogues, example sentences in the grammar sections, and exercises in the textbook at the manipulated speech rate for each group.

Table 7.1.
Average Speech Rates in WPM of the Pretest, the Posttest, and Treatment Materials (n = 206)

Groups	<i>n</i>	Word Recognition Tests		Treatment
		Pretest	Posttest	
Control	42			126
Experimental 1	41	131	115	151
Experimental 2	83			189
Experimental 3	40			252

The treatment lasted about half a year until the posttest was conducted. During the treatment period, the participants were given normal classes, one 90-minute session a week, in which they learned English, using the textbook. Accordingly, the only difference among the groups was the rates at which they listened to the CD attached to the textbook. The activities using the CD at the manipulated speech rate involved listening practices with the textbook closed, listening with the textbook open and with the participants following the written scripts while listening, a couple of sets of sentence-by-sentence repeating and shadowing

practices, reading the Japanese translation while listening, and shadowing practices with the participants seeing the Japanese translation. All these activities involved were given both before and after explicit syntactical and phrasal explanations concerning the scripts.

As far as word recognition tests are concerned, each of the 50 (25 content and 25 function) words in the pretest and the posttest were graded either correct or incorrect. The participants were discouraged from using *katakana* when they were unsure of the spellings, and asked to use alphabet letters that they thought they had heard. Therefore, all the *katakana* answers were judged to be incorrect.

In grading, if the sound was recognized correctly, the item was judged to be correct, even if it was misspelled. However, if, for example, *here* was misspelled as *hear* or *buy* as *by*, they were judged to be incorrect. These misspellings may well have been caused by a failure to recognize a word at the parsing phase and correct word recognition must involve both perception and parsing phases while the listener segments the incoming speech, through phonological analysis and word retrieval from the listener's mental lexicon, into meaningful units, with reference to syntactic and semantic cues (Vandergrift & Goh, 2012).

All the data were computed into the percentage of correct word recognition, with the number of items correctly recognized being the numerator and the total number of blanked-out items the denominator. Content words and function words were separately examined.

7.3 Results

7.3.1 Results of Three-Way ANOVA

Table 7.2 shows the descriptive statistics of the words correctly

recognized in percentage in the pretest (Cronbach's alpha = .789) and in the posttest (Cronbach's alpha = .828).

Table 7.2.

Descriptive Statistics of the Pretest and the Posttest (Words Correctly Recognized, %, n = 206)

Group	n	Pretest						Posttest					
		Content		Function		Total		Content		Function		Total	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	42	39.43	13.31	39.05	11.51	39.24	11.14	53.62	12.47	34.19	11.65	43.90	11.52
Experimental 1	41	34.83	14.63	37.76	13.11	36.29	12.85	50.73	14.66	32.49	13.23	41.61	13.04
Experimental 2	83	39.08	13.79	37.83	14.43	38.46	13.05	60.10	13.24	39.13	13.07	49.61	12.24
Experimental 3	40	35.80	11.45	36.00	11.76	35.90	9.58	54.20	13.56	34.20	12.13	44.20	12.14

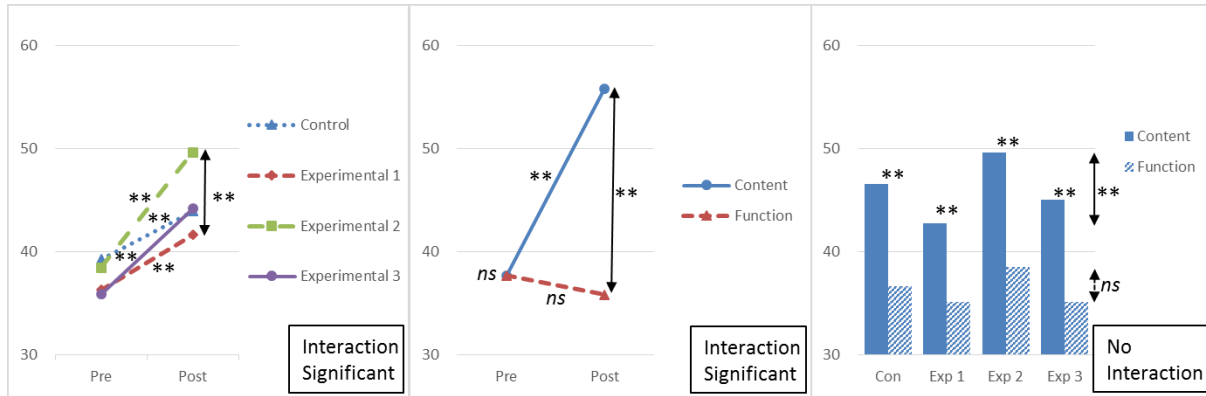


Figure 7.1. Words successfully recognized in percentage for each two of the three factors. (**: $p < .01$).

Table 7.3.

The Results of the Three-Way ANOVA (n = 206)

Source	SS	df	MS	F	p	η_p^2
A: Groups	2731.24	3	910.41	1.723	.163	.014
S: Error (A)	106707.92	202	528.26			
B: Time	10167.22	1	10167.22	166.709	.000 ***	.053
Interaction (AB)	1254.02	3	418.01	6.854	.000 ***	.006
Error (BS)	12319.58	202	60.99			
C: Word Categories	17452.37	1	17452.37	276.274	.000 ***	.090
Interaction (AC)	291.86	3	97.29	1.540	.205	.002
Error (CS)	12760.46	202	63.17			
Interaction (BC)	18829.45	1	18829.45	342.838	.000 ***	.097
Second-Order Interaction (ABC)	28.12	3	9.37	0.171	.916	.000
Error (BCS)	11094.29	202	54.92			
Total	193636.52	823				

***: $p < .001$

A three-way mixed ANOVA³ (A: groups: control/experimental 1-3, B: time: pre/post, C: word categories: content/function) was conducted and the results are shown in Table 7.3. Figure 7.1 shows the means of correct word recognition in percentage for each two of the three factors.

The second-order interaction was not significant ($F(3, 202) = 0.171, p = .916, \eta_p^2 = .000$). Neither was there significant interaction between groups and word categories ($F(3, 202) = 1.540, p = .205, \eta_p^2 = .002$). However, significant interactions were found between groups and time ($F(3, 202) = 6.854, p = .000, \eta_p^2 = .006$) as well as time and word categories ($F(1, 202) = 342.838, p = .000, \eta_p^2 = .097$), so that respective simple main effects were examined (Tables 7.4 & 7.5). Differences in mean values at the posttest between four groups were assessed with multiple comparison procedure using Tukey-Kramer method (Table 7.6).

Table 7.4.
Simple Main Effects in AB Interaction (n = 206)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups (Pretest)	746.33	3	248.78	0.844	.470	.006
A: Groups (Posttest)	3238.92	3	1079.64	3.664	.013 *	.026
Error	119027.49	404	294.62			
B: Time (Control)	1021.86	1	1021.86	16.755	.000 ***	.043
B: Time (Experimental 1)	1326.54	1	1326.54	21.751	.000 ***	.056
B: Time (Experimental 2)	5840.39	1	5840.39	95.763	.000 ***	.246
B: Time (Experimental 3)	3232.45	1	3232.45	53.001	.000 ***	.136
Error	12319.58	202	60.99			

***: $p < .001$, *: $p < .05$

Table 7.5.
Simple Main Effects in BC Interaction (n = 206)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
B: Time (Content)	28334.64	1	28334.64	488.907	.000 ***	.544
B: Time (Function)	357.90	1	357.90	1.979	.160	.007
Error	23413.87	404	57.96			
C: Words Categories (Pre)	13.07	1	13.07	0.221	.638	.000
C: Words Categories (Post)	36268.74	1	36268.74	614.241	.000 ***	.603
Error	23854.76	404	59.05			

***: $p < .001$

Table 7.6.

The Results of Multiple Comparison Between Four Groups at the Posttest (n = 206)

Source	Group 1 (I)	Group 2 (J)	Difference (J - I)	<i>t</i>	<i>p</i>
A: Groups (Posttest)	Control	Experimental 1	-2.30	1.218	.610
	Control	Experimental 2	5.71	3.513	.003 **
	Control	Experimental 3	0.30	0.156	.999
	Experimental 1	Experimental 2	8.00	4.886	.000 ***
	Experimental 1	Experimental 3	2.59	1.358	.520
	Experimental 2	Experimental 3	-5.41	3.278	.006 **

***: $p < .001$, **: $p < .01$

As for the interaction between groups and time, although the simple main effects of time in all the four groups were significant (Control Group: $F(1, 202) = 16.755$, $p = .000$, $\eta_p^2 = .043$, Experimental Group 1: $F(1, 202) = 21.751$, $p = .000$, $\eta_p^2 = .056$, Experimental Group 2: $F(1, 202) = 95.763$, $p = .000$, $\eta_p^2 = .246$, Experimental Group 3: $F(1, 202) = 53.001$, $p = .000$, $\eta_p^2 = .136$), those of groups were not significant in the pretest ($F(3, 404) = 0.844$, $p = .470$, $\eta_p^2 = .006$), while in the posttest they were significant ($F(3, 404) = 3.664$, $p = .013$, $\eta_p^2 = .026$). This means that, even though the participants in all the four groups fared better in the posttest than in the pretest, difference in word recognition, which did not exist between the four groups in the pretest, emerged in the posttest after the treatment.

In addition, the results of multiple comparison procedure indicated that the participants in Experimental Group 2, who constantly listened to the textbook's CD at 1.5 times the original rate, fared significantly better in the posttest than the other three groups (Control Group & Experimental Group 2: $t = 3.513$, $p = .003$, Experimental Groups 1 & 2: $t = 4.886$, $p = .000$, Experimental Groups 2 & 3: $t = 3.278$, $p = .006$). Although, as the graph shows, Experimental Group 3's word recognition also improved better than those of Control Group and Experimental Group 1, the difference in the

posttest was not significant (Control Group & Experimental Group 3: $t = 0.156$, $p = .999$, Experimental Groups 1 & 3: $t = 1.358$, $p = .520$).

As for the interaction between time and word categories, although the simple main effects of time for content words were significant ($F(1, 404) = 488.907$, $p = .000$, $\eta_p^2 = .544$), those for function words were not significant ($F(1, 404) = 1.979$, $p = .160$, $\eta_p^2 = .007$) and, although the simple main effects of word categories in the pretest was not significant ($F(1, 404) = 0.221$, $p = .638$, $\eta_p^2 = .000$), those in the posttest was significant ($F(1, 404) = 614.241$, $p = .000$, $\eta_p^2 = .603$). These results mean that recognition of content words improved across the groups, while that of function words did not. As a result, difference in recognition between content and function words, which did not exist in the pretest, emerged in the posttest.

7.3.2 Results of Two-Way ANOVAs

In order to more thoroughly analyze the effects of the treatment on word recognition, a two-way mixed ANOVAs³ (A: groups: control/experimental 1-3, B: time: pre/post) were conducted separately for content and function words.

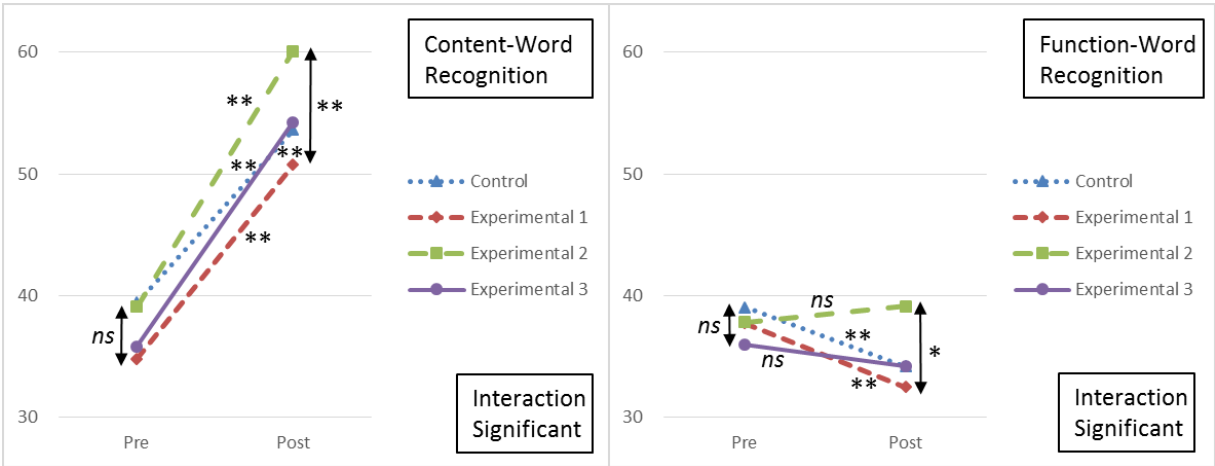


Figure 7.2. Content and function words successfully recognized in percentage at the pretest and the posttest. (**: $p < .01$, *: $p < .05$)

As a result, significant interactions were found for both content ($F(3, 202) = 4.636, p = .004, \eta_p^2 = .007$) and function ($F(3, 202) = 4.654, p = .000, \eta_p^2 = .012$) words, so that simple main effects were examined. Figure 7.2 shows the means of correct content and function word recognition in percentage by each group at the pretest and the posttest.

First, recognition of content words improved significantly across all the four groups between the pretest and the posttest (Control Group: $F(1, 202) = 76.015, p = .000, \eta_p^2 = .092$, Experimental Group 1: $F(1, 202) = 93.190, p = .000, \eta_p^2 = .113$, Experimental Group 2: $F(1, 202) = 329.362, p = .000, \eta_p^2 = .401$, Experimental Group 3: $F(1, 202) = 121.718, p = .000, \eta_p^2 = .148$). That of function words, however, deteriorated significantly for Control Group ($F(1, 202) = 8.219, p = .005, \eta_p^2 = .037$) and Experimental Group 1 ($F(1, 202) = 9.439, p = .002, \eta_p^2 = .043$), while that of the other two groups, Experimental Group 2 ($F(1, 202) = 1.166, p = .282, \eta_p^2 = .005$) and 3 ($F(1, 202) = 1.075, p = .301, \eta_p^2 = .005$), remained unchanged.

Second, difference between the groups, which was nonexistent at the pretest (content: $F(3, 404) = 1.386, p = .247, \eta_p^2 = .014$, function: $F(3, 404) = 0.381, p = .767, \eta_p^2 = .004$), arose after the treatment at the posttest (content: $F(3, 404) = 5.218, p = .002, \eta_p^2 = .054$, function: $F(3, 404) = 3.117, p = .027, \eta_p^2 = .032$) for both content and function word recognition. Though the difference between the groups at the posttest was greater in recognition of content words than in that of function words, the results of multiple comparison procedure, using Tukey-Kramer method, illustrate that, in both content and function word recognition, Experimental Group 2 fared significantly better than the other three groups, while the differences between the other three groups were not significant (Table 7.7).

Table 7.7.

The Results of Multiple Comparison for Content and Function Word Recognition at the Posttest (n = 206)

Source	Group 1 (I)	Group 2 (J)	Content-Word Recognition			Function-Word Recognition		
			Diff (J - I)	<i>t</i>	<i>p</i>	Diff (J - I)	<i>t</i>	<i>p</i>
A: Groups (Posttest)	Control	Exp 1	-2.89	1.370	.513	-1.70	0.843	.830
	Control	Exp 2	6.48	3.562	.002 **	4.94	2.837	.024 *
	Control	Exp 3	0.58	0.274	.993	0.01	0.005	1.000
	Exp 1	Exp 2	9.36	5.109	.000 ***	6.64	3.784	.001 **
	Exp 1	Exp 3	3.47	1.625	.360	1.71	0.838	.833
	Exp 2	Exp 3	-5.90	3.190	.008 **	-4.93	2.786	.028 *

***: $p < .001$, **: $p < .01$, *: $p < .05$

7.3.3 Results of Chi-Square Tests

In order to examine the effects of the treatment in detail, chi-square tests³ were conducted. The numbers of right (R) and wrong (W) transcriptions by the three experimental and control groups for all the 50 words in the posttest were computed and in which of the 50 words in the posttest there were significant differences in recognition between the four groups was examined.

Results are shown in Table 7.8. In terms of chi-square values, significant differences were found in 8 words. In addition, significant standardized residuals were found in 4 other words for Experimental Group 2, and in still another word for Experimental Group 3, even though the chi-square values were not significant. Among these 13 words, participants in Experimental Group 2 fared significantly better than the other three groups in 11 words, seven content and four function words. The greatest difference was found in the word No. 30 *yet* ($\chi^2(3) = 18.991$, $p = .000$, Cramer's $V = .304$).

Table 7.8.

*Words in Which There Was a Significant Difference in Recognition
Between the Control and Three Experimental Groups (n = 206)*

Group	2 popular		3 <i>such</i>		4 <i>should</i>		6 math						
	R	W	R	W	R	W	R	W					
Control (n = 42)	39	3	18	24	34	8	28	14					
Experimental 1 (n = 41)	35	6	19	22	27	14	17	24					
Experimental 2 (n = 83)	81	2	56	27	75	8	65	18					
Experimental 3 (n = 40)	33	7	22	18	37	3	27	13					
	Con	0.410	-0.410	-1.897	1.897	-0.600	0.600	0.025	-0.025				
Standardised	Exp 1	-1.494	1.494	-1.366	1.366	-3.536	3.536	***	-3.796	3.796	***		
Residuals	Exp 2	2.642	-2.642	**	2.765	-2.765	**	2.051	-2.051	*	2.950	-2.950	**
	Exp 3	-2.186	2.186	*	-0.117	0.117	1.636	-1.636	0.149	-0.149			
χ^2 (df = 3)		9.941		8.934		14.970		16.755					
p		.019	*	.030	*	.002	**	.001	***				
Cramer's V		.220		.208		.270		.285					

Group	10 <i>aren't</i>		11 <i>where</i>		17 <i>giving</i>		19 <i>true</i>					
	R	W	R	W	R	W	R	W				
Control (n = 42)	12	30	28	14	19	23	37	5				
Experimental 1 (n = 41)	8	33	30	11	16	25	30	11				
Experimental 2 (n = 83)	34	49	70	13	48	35	77	6				
Experimental 3 (n = 40)	19	21	30	10	18	22	34	6				
	Con	-1.043	1.043	-1.724	1.724	-0.551	0.551	0.358	-0.358			
Standardised	Exp 1	-2.382	2.382	*	-0.597	0.597	-1.432	1.432	-2.763	2.763	**	
Residuals	Exp 2	1.362	-1.362	2.130	-2.130	*	2.076	-2.076	*	2.189	-2.189	*
	Exp 3	1.777	-1.777	-0.283	0.283	-0.568	0.568	-0.289	0.289			
χ^2 (df = 3)		9.062		5.425		4.717		9.148				
p		.028	*	.143		.194		.027	*			
Cramer's V		.210		.162		.151		.211				

Group	26 <i>dentist</i>		30 <i>yet</i>		36 <i>different</i>		45 <i>around</i>					
	R	W	R	W	R	W	R	W				
Control (n = 42)	25	17	11	31	26	16	28	14				
Experimental 1 (n = 41)	25	16	11	30	27	14	29	12				
Experimental 2 (n = 83)	60	23	48	35	63	20	65	18				
Experimental 3 (n = 40)	21	19	12	28	22	18	36	4				
	Con	-0.614	0.614	-2.020	2.020	*	-0.785	0.785	-1.724	1.724		
Standardised	Exp 1	-0.389	0.389	-1.897	1.897	-0.173	0.173	-1.010	1.010			
Residuals	Exp 2	2.131	-2.131	*	4.342	-4.342	***	2.235	-2.235	*	0.450	-0.450
	Exp 3	-1.624	1.624	-1.411	1.411	-1.796	1.796	2.217	-2.217	*		
χ^2 (df = 3)		5.259		18.991		6.098		7.263				
p		.154		.000	***	.107		.064				
Cramer's V		.160		.304		.172		.188				

Group	46 popular			
	R	W		
Control (n = 42)	28	14		
Experimental 1 (n = 41)	27	14		
Experimental 2 (n = 83)	73	10		
Experimental 3 (n = 40)	29	11		
	Con	-1.629	1.629	
Standardised	Exp 1	-1.741	1.741	
Residuals	Exp 2	3.250	-3.250	**
	Exp 3	-0.615	0.615	
χ^2 (df = 3)		11.152		
p		.011	*	
Cramer's V		.233		

Function words in italics.

***: $p < .001$, **: $p < .01$, *: $p < .05$

7.4 Discussion

First, as to whether steady exposure to a faster rate of speech is effective in improving word recognition by Japanese EFL listeners with lower levels of proficiency, the results of the experiment showed that half-a-year repeated sessions, in which the participants were given listening materials at mechanically compressed speech rates, were effective on their word recognition at the baseline speech rate.

Second, as to whether there is any specific speech rate at which lower-proficiency listeners fare best in word recognition, the results showed that it was most effective at the compressed rate of about 67 percent, that is, 1.5 times the baseline speech rate and that any other compressed rate was found not to have any significant effects.

These results suggest that, when L2 learners, especially lower-proficiency Japanese EFL learners, are trained by constantly listening to English at the 'right' compressed rate, there seem to be some positive effects on their recognition of words. The results of the experiment show that the 'right' rate is 1.5 times the original rate. In addition, regardless of word categories, content or function, Experimental Group 2 fared significantly better in the posttest than the other three groups.

However, the gain in function word recognition that Experimental Group 2 made between the two tests was smaller than that in content word recognition and, for all the other groups, recognition of function words deteriorated in the posttest. This may be because it takes more than improvement in bottom-up processing to successfully recognize function words, such as intervention from the top-down processing.

In recognizing function words, which are often made up only of unstressed syllables and hence quite challenging to recognize solely from

the information gained through the mere phonetic information, must oftentimes be inferred; they need to be first recognized as a chunk of several words, or a formulaic sequence (a stress unit or a phonological word, as we discussed in Chapter 3), before separated into individual words in reference to the learner's linguistic knowledge.

Therefore, it takes, at the least, some forms of top-down intervention to successfully recognize function words, as was the case in Experiment 2 (Chapter 5). In addition, the whole matter of word recognition, especially that of function words, depends on the successful interaction between top-down and bottom-up processing (as was discussed in Chapter 6), that is, on the amount of phonetic information successfully obtained from the sound stream and the linguistic knowledge the listener already has, and also on how the listener uses those information and knowledge.

It can be assumed that, in this experiment, the treatment, or some fortification of bottom-up-related skills resulted from it, helped the participants reach the level required by the top-down processing in recognizing content words, but was still not sufficient in leveraging recognition of function words.

Provided that the top-down intervention be left as it is, it takes another sort of pedagogical strategies from the perspective of bottom-up processing to enhance recognition of unstressed syllables. This should include the one related to and focused on English phonological features such as phonemes, syllable structures, and the stress-timed rhythm of the language.

Nevertheless, in the analysis of word by word, the word in which the greatest difference was found in the posttest was a function word. It may safely be said, therefore, that the speech rate that is effective in training

elementary-level Japanese EFL learners to improve word recognition of the normal speech rate (around 130 wpm) is 1.5 times the baseline rate (around 190 wpm), even though what constitutes the normal speech rate for L2 learners is still another issue that should be discussed.

Finally, as far as the group-by-group improvement in word recognition after the treatment, Experimental Group 3, who constantly listened to the textbook's CD at double the original rate, fared second best after the treatment, even though the difference was not significant. According to Dupoux and Green (1997), more highly compressed stimuli require more time for improvement than less compressed ones and performance improves with increased exposure to compressed speech. This suggests that the results obtained for the group who listened during the treatment at double the baseline rate may have turned significant if the span of the treatment had been longer.

7.5 Implications

As was discussed in the previous section, in order to effectively enhance recognition of function words, another approach from the bottom-up processing may be necessary, since the one used in this experiment, manipulation of speech rate, had limited effects on recognition of function words.

The next chapter will explore the effects of an approach focused on English phonological features; specifically, how explicit instructions on English syllable structure and its stress-timed rhythm, followed by some practice sessions of both articulation and perception on the part of learners make a difference in recognizing function as well as content words.

Notes

1. Quirk et al. (1985) was referred to in distinguishing the two word categories. Due to the contextual and syntactical functions, however, No. 28 *some* was regarded as a content word and No. 30 *yet* as a function word.
2. The title of the authorized textbook was *Vision Quest vol. 1*, published by Keirinkan.
3. In analyzing the data, an online software of ANOVA 4 was used for the three-way ANOVAs and a Microsoft add-in software for Excel was used for the two-way and one-way ANOVAs and for the chi-square tests.
4. The speed-changing software used was the one attached by the publisher of the textbook.
5. The average speech rate of the pretest material was 134 for the dialogue and 130 for the monologue, and for the posttest, 113 for the dialogue and 118 for the monologue.

Chapter 8

Experiment 5

Examining the Effects of Explicit Instructions Concerning English Phonological Features on Spoken Word Recognition

This chapter empirically examines whether it would be effective on word recognition to give explicit explanations about English phonological features: the phonemic system, the syllable structure, and the rhythm. In the experiment, the participants were given perception and articulation practices after the explanations. The experiment lasted for half a year. The effects of this treatment on word recognition will be discussed.

8.1 Introduction

In Chapter 7, it has been empirically demonstrated that spoken word recognition by Japanese EFL learners with lower levels of proficiency improved after half-a-year weekly sessions in which listening materials at the mechanically compressed rate of about 67 percent were constantly given. However, the positive effects found on recognition of function words, which are often made up of unstressed syllables pronounced weak in a stress unit, were limited. In this chapter, it will be examined how treatment in which learners are given explicit instructions on English phonological features, namely, phonemes, the syllable structure, and the stress-timed rhythm, followed by perception as well as articulation practices affects their spoken word recognition.

8.1.1 Learning Habits of Lower-Proficiency Japanese EFL Learners and the Difference in Phonological Features Between English and Japanese

Theoretical evidence, as was referred to in Chapter 3, is abundant that a kind of approach that is focused on English phonological features is expected to help enhance word recognition of Japanese EFL learners, whose mother tongue is a mora-timed language that shares very little with English in terms of phonemes, syllable structures, and rhythms.

Japanese L1 speakers, especially lower-proficiency EFL learners, tend to form in their mental lexicon acoustic image of speech, combining individual words articulated separately through phoneme-grapheme correspondence based on *romaji* GPC rules. In addition, in the process, they often insert unnecessary vowels after each consonant, further distorting the image. The acoustic image thus created as a mental representation of English speech is supposedly encoded in a mora-timed manner. When these learners articulate a string of words, they use this distorted image and when they listen, they expect to hear the same image.

This habit partly results from their dependence on a written version of the language, just as they have always been doing in their native language, which does no harm in so doing. They do not form the image by hearing those chunks of words uttered in a natural English speech, but by following the written script and transforming the written representation into their acoustic image themselves. They try to articulate English speech by applying this transformation rule, and not by copying the sound they hear. Therefore, in perceiving the speech, all they can rely on is this distorted image they have stored in their phonological lexicon.

Against this background lies the fact that their native language has a written system of *kana*, a phonogram, and that the basic unit for word

recognition is a unit of mora, which exactly corresponds to each *kana* character in the written form. This means that no problem arises even if they depend on written characters in forming their acoustic image of the language. Many lower-proficiency Japanese EFL learners try to apply to English the same rule, which they use in a language where each of its written character doubles as a basic unit for word recognition as well as its sound itself.

English is a language, on the other hand, where neither is a basic unit for word recognition clearly indicated in its written representation, nor do the sound and its written form correspond one on one. By breaking the habit of depending on the written characters for perceptual cues or applying Japanese phonological system and by forming a habit of correctly perceiving and articulating a stress unit in a stress-timed manner and also applying other rules related to English phonological system, their spoken word recognition will quite presumably be enhanced.

If words, whether content or function, can be recognized first as a stress unit, before segmenting it into individual words, thereby resulting in enhancement of their bottom-up skills, then the methods used in the previous experiments can be taken better advantage of, because a more effective interaction between the bottom-up and top-down processing can be expected.

8.1.2 Perception and Articulation

As for the relationship between perception and articulation, Lieberman (1963) says about the interconnected circuit between perception and articulation as follows:

“When a listener hears an utterance he may analyze it through a process of synthesis in which he invokes the same ordered rules that he uses for the production of speech and generates internal signals which he compares and matches with the utterance that he is listening to.” (p.173)

This means that perception and articulation are closely related and that, if a learner cannot articulate speech in a correct manner, she cannot expect as a listener to recognize an utterance articulated in the same manner. Ur (1984) also states ‘If L2 learners learn to pronounce the phonemes of the target language accurately themselves, it will be much easier for them to hear them correctly when said by someone else’ (p. 12).

The sound one cannot pronounce correctly is processed by brain as noise so that one cannot recognize an utterance unless one can articulate it correctly, which means that listening comprehension and pronunciation training are interrelated and are two aspects of the same system (Morley, 1991; Brown, 1992; Gilbert, 1995; McDonough & Trofimovich, 2009). Vanderplank (1993) and Eastman (1993) also observe that articulation practice has some positive effects on spoken word recognition.

Accordingly, it would help enhance learners’ spoken word recognition to enunciate English phonemes correctly, to copy and repeat exactly the same pronunciation of each syllable, including many consonant clusters and codas, coupled with various phonetic changes, to articulate formulaic sequences and other strings of words as a whole, and to rigidly stick to the stress-timed rhythm.

8.2 Experiment

8.2.1 Purpose

The purpose of this experiment is to examine the effects on spoken

word recognition of weekly listening sessions in which the following treatment was given: explicit explanations on English phonological features followed by perception and articulation practices. The phonological features include phonemes, especially the ones not shared by the Japanese language, syllable structures, and the stress-timed rhythm among others. Perception and articulation practices include listening, shadowing, repeating, and oral reading practices.

In the experiment, we focused more on the aspects of syllable structures and stress-timed rhythm than on phonemes, because, as studies suggest, alterations between strong and weak syllables in English play a more important role than phoneme level processing in perceiving speech and, because of these stress patterns and rhythmic structures, L2 listeners whose L1 are syllable-timed often fail to recognize words they actually know very well (Buck, 2001; Graham, 2006; Field, 2008a).

8.2.2 Participants

The participants were 76 fourth-year students at a technical college in Japan who majored in engineering. Their L1 is Japanese. The participants ranged in their levels of English proficiency from an elementary to a lower-intermediate level; their mean score of the TOEIC (IP) was 413.90 ($SD = 97.20$). They were divided into two groups according to the class they belonged to: the control group ($n = 40$) and the experimental group ($n = 36$).

8.2.3 Method

A listening comprehension test was preliminarily conducted to assess the participants' level of proficiency in listening comprehension. The test, adopted from a second-grade STEP listening test, consisted of 30 questions,

all of which were of a multiple-choice type.

For the word recognition tests, two types of transcription tests were provided both for the pretests and the posttests: cloze tests and paused transcription tests (Appendix 7). The former was a test in which words were blanked out in the script and the latter was a test in which the test takers were asked to transcribe the words they thought they had heard before the pauses, just like the ones given in Experiments 1 and 3; that is, the pauses, which lasted about 10 seconds each, were inserted at irregular intervals in the spoken text and the participants were asked to transcribe the last four to five words they thought they had heard before each pause. In both the cloze tests and the paused transcription tests, the participants listened to the recordings only once.

For materials of the pretest and the posttest, one dialogue and one monologue each for the pretest and the posttest, four different texts in total, were used. The texts were adopted from the same listening textbook¹ used in Experiments 1 and 3, which all the texts consisted of the most basic 1,000 words. For both the pretests and the posttests, the dialogues were used in the cloze tests and the monologues in the paused transcription tests.

The speech rate of the pretest material was 175 wpm in the cloze test and 146 wpm in the paused transcription test, while in the posttest it was 183 wpm for the cloze and 151 wpm for the paused transcription test.² The materials were used to test participants' accuracy in spoken word recognition so that the ones with simple vocabulary but comparatively high speech rates were adopted. The vocabulary and syntactical structures used in the texts were at the beginner level.

The cloze tests in the pretest and the posttest shared the same format and had 20 blanked-out words each, of which half of them were content

words with the other half function words.³ The participants were asked to write down one word in each blank, and discouraged from using *katakana*. All the targeted words in both the pretest and the posttest were at a junior-high school level in Japan. The participants would have had little difficulty in recognizing them, if those words had been in written scripts or articulated individually.

In the paused transcription tests, pauses were inserted in the same place in the same text across the two groups in both the pretest and the posttest. The participants were discouraged from using *katakana* when they were unsure of the spellings, and asked to use alphabet letters that they thought they had heard. Pauses were inserted 10 times in both the pretest and the posttest. Judgment of whether the participants' handwritten responses were accurate or otherwise was limited to the last four words before each pause. Therefore, 40 items, 4 each for every pause, were the maximum accurate responses possible for both the pretest and the posttest. Of the 40 items targeted, half of them, 20, were content words and the rest were function words in both tests.³ In the paused transcription tests as well, words in the targeted sections were all easy enough for the participants to recognize, if they were to be given its written script or if the words were to be pronounced individually. Sections of recordings targeted for transcription in the paused transcription posttest are shown in Table 8.1.

Each of the 20 words in the cloze tests and 40 words in the paused transcription tests for both the pretest and the posttest was graded either correct or incorrect. In grading⁴, if the sound was recognized correctly, the item was judged to be correct, even if it was misspelled in both cloze and paused transcription test. In the paused transcription test, all the items

involved were judged to be incorrect, if the word boundaries were breached.

Table 8.1.
*Sections of Recording Targeted for Transcription in the Paused
Transcription Posttest*

1	<i>to</i> visit <i>her</i> aunt	6	<i>in</i> a beautiful house
2	have <i>much</i> money <i>either</i>	7	<i>is</i> with <i>another</i> woman
3	<i>who</i> runs a school	8	<i>She</i> cries <i>and</i> says
4	<i>was</i> a student <i>there</i>	9	<i>don't</i> leave <i>me</i> alone
5	<i>are</i> going <i>to</i> marry	10	start spending time together

Function words are in italics.

All the data were computed into the percentage of correct word recognition, with the number of items correctly identified being the numerator and the total number of items targeted for transcription the denominator. Content words and function words were separately examined.

8.2.4 Treatment

After the pretest, the experiment lasted about half a year before the posttest was conducted. All the fourth-year students of the technical college, including the participants of the control and the experimental groups, were required to attend two 90-minute English sessions a week throughout the year: Scientific English (2 credits) and English Practice (2 credits). The present experiment was conducted in English Practice during the first semester of the school year.

During this half-a-year period, only the participants of the experimental group were given treatment, about 30 minutes per session, besides the normal class, while those in the control group had only normal class. In addition, it can be assumed that there was little difference between the participants of the two groups in the time spent outside class

during the period. Therefore, the following treatment was presumably the only difference between the experimental and the control groups.

The treatment the participants of the experimental group was given included explicit explanations about English phonological features as well as perception and articulation practices using actual materials after the explanations (Appendix 8). Explanations about English phonological features included the following content:

1. English phonemes, especially the ones the Japanese language does not share and some phonemic pairs that Japanese EFL learners typically have difficulty in perceiving and articulating distinctly (e.g., l/r and b/v)
2. Differences in syllable structures between open-syllable Japanese and closed-syllable English, which is featured by consonant clusters and codas (e.g., *straight*) as well as various phonetic changes such as those which often occur between codas and the first phonemes of the following word (e.g., *out of*)
3. English stress-timed rhythm in comparison with Japanese mora-timed one, especially its distinct feature of stressed syllables appearing at the same intervals irrespective of the number of unstressed syllables between the two stressed ones and resulting quick articulation of sandwiched weak syllables, coupled with practical examples of phonetic changes resulting from the rhythm (e.g., *would have been*).

Explanations were also given about how written version of English is disproportionate in length, because of this stressed-timed rhythm, with the spoken counterpart so that its articulation time is generally much shorter than the participants would expect it should be. Further, some comments

were given about unnecessary vowel insertions by Japanese L1 speakers in articulating English sentences, thereby making each word unnecessarily longer by rendering single-syllable structures multi-syllable ones, preventing some phonetic changes between a coda and the first phoneme of the next word from happening, and causing them to pronounce each word separately and independently. The participants were also told that all of these should negatively affect their spoken word recognition. Comments were also given about Japanese EFL learners' dependence on written English, or visual image, in learning English, and the participants were advised to depend more upon the sound they heard.

Perception and articulation practices, on the other hand, included those of minimum-paired words with the phonemes which Japanese L1 speakers find it difficult to distinguish (e.g., *lest* and *rest*), of sentences that contained those phonemes (e.g., *Let's read the rest of the play later.*), and of words, phrases, and sentences featuring closed-syllable structures and the stress-timed rhythm. They also included several sessions of listening, shadowing, repeating, and oral reading practices using some dialogue texts⁵.

In phonemic practices, for example, the participants were also asked to listen to the instructor articulate Japanese sentences, applying English phonemic categorization, which included those phonemes challenging for them to distinguish (e.g., *Ohiru ni kareraisu taberu* vs. *Ohilu ni kalelaisu tabelu* (I'll have curry and rice for lunch)) and to repeat and articulate the same sentences themselves.

In practices featuring syllable structures, the participants were asked to listen and repeat words and phrases with consonant clusters and/or codas (e.g., *a couple of* or *straight on*) by exactly copying the sound, while,

in practices focused on the stress-timed rhythm, they were told to listen and repeat what they had heard exactly the way they had and in the same rhythm, blurring word boundaries and not articulating each word separately and independently (e.g., *Keith should have been reading some of the documents*).

In the practices using dialogues, the participants were told, in articulation, not to insert unnecessary vowels after each consonant and stick rigidly to the English stress-timed manner, to enunciate stressed syllables clearly and with stress, while distressing others, and to copy and trace English prosodic features such as schwas in weak syllables, vowel reductions and consonant elisions, completely obliterating mora-timed pronunciation, when they repeated, shadowed, and read the texts orally. They were also asked, in reading aloud the text, to try to follow the acoustic image they had had from the CD and not to follow the written words.

One 30-minute treatment session followed a cycle of explicit explanations on phonemes, syllable structures and the rhythm, perception and articulation practices on the items used in the explanations, listening to the CD of a dialogue, repeating and shadowing practices of the dialogue, and then oral reading practices of the dialogue by the participants themselves. In one session, this cycle was repeated a couple of times, which was aimed to close the gap between the participants' articulation and the recordings of the dialogue in the CD.

The materials used in the treatment were from a series of monthly magazines⁶ and from the same listening textbook¹ used in the pretests and the posttests.

8.3 Results

8.3.1 Results of the Preliminary Listening Comprehension Test

Table 8.2 shows the descriptive statistics of the preliminary listening comprehension test (Cronbach's alpha = .760), which was conducted to assess the participants' listening proficiency preceding the experiment, and the results of an independent two-tailed *t*-test⁷.

Table 8.2.

Descriptive Statistics of the Preliminary Listening Comprehension Test and the Results of the t-Test

Group	<i>n</i>	<i>M</i>	<i>SD</i>	
Control	40	11.63	3.77	$t(74) = 0.345, p = .731 \text{ ns}, r = .041$
Experimental	36	11.28	4.98	

The results of the *t*-test showed that there was no significant difference between the control and the experimental groups in terms of listening proficiency ($t(74) = 0.345, p = .731, r = .041$).

8.3.2 Results of the Cloze Tests

Table 8.3 shows the descriptive statistics of the words correctly recognized in percentage in the pretest (Cronbach's alpha = .758) and the posttest (Cronbach's alpha = .734) of the cloze tests.

Table 8.3.

Descriptive Statistics of the Cloze Tests (Words Correctly Recognized, %, n = 76)

Group	<i>n</i>	Pretest						Posttest					
		Content		Function		Total		Content		Function		Total	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Control	40	42.25	16.81	70.00	17.32	56.13	15.50	35.25	16.88	24.25	14.64	29.75	13.63
Experimental	36	35.28	17.87	56.94	22.46	46.11	19.35	46.67	18.11	40.56	20.81	43.61	17.10

8.3.2.1 Results of a Three-Way ANOVA

A three-way ANOVA⁷ (A: groups: control/experimental, B: time: pre/post, C: word categories: content/function) was first carried out to analyze the data. The results are shown in Table 8.4. Figure 8.1 shows the means of correct word recognition in percentage for each of the three factors.

Table 8.4.

The Results of the Three-Way ANOVA for the Cloze Tests (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	280.44	1	280.44	0.313	.578	.002
S: Error (A)	66341.60	74	896.51			
B: Time	15797.66	1	15797.66	86.120	.000 ***	.103
Interaction (AB)	10800.30	1	10800.30	58.877	.000 ***	.070
Error (BS)	13574.38	74	183.44			
C: Word Categories	4943.60	1	4943.60	30.635	.000 ***	.032
Interaction (AC)	6.76	1	6.76	0.042	.838	.000
Error (CS)	11941.60	74	161.37			
Interaction (BC)	20965.00	1	20965.00	183.275	.000 ***	.136
Second-Order Interaction (ABC)	570.27	1	570.27	4.985	.029 *	.004
Error (BCS)	8464.93	74	114.39			
Total	153686.53	303				

***: $p < .001$, *: $p < .05$

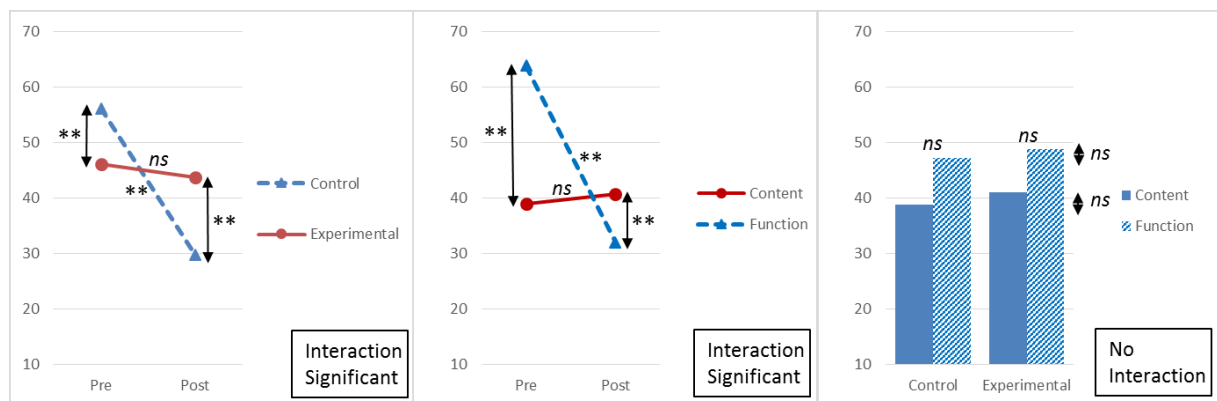


Figure 8.1. Words successfully recognized in percentage for each two of the three factors in the cloze tests. (**: $p < .01$).

The results of the ANOVA demonstrated that a significant second-order interaction was found ($F(1, 74) = 4.985$, $p = .029$, partial $\eta^2 = .004$),

so that simple interactions at each level of the three factors between all the combinations of the other two factors were examined (Table 8.5).

Table 8.5.
Simple Interactions at Each Level of the Three Factors in the Cloze Tests (n = 76)

Source		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	BC (Control)	14225.33	1	14225.33	124.357	.000 ***	.474
	BC (Experimental)	7309.94	1	7309.94	63.903	.000 ***	.244
	Error	8464.93	74	114.39			
B: Time	AC (Pre)	350.59	1	350.59	2.543	.113	.017
	AC (Post)	226.43	1	226.43	1.642	.202	.011
	Error	20406.53	148	137.88			
C: Word Categories	AB (Content)	3203.54	1	3203.54	21.513	.000 ***	.096
	AB (Function)	8167.02	1	8167.02	54.844	.000 ***	.244
	Error	22039.31	148	148.91			

***: $p < .001$

Table 8.6.
Simple Main Effects of Three Factors at Each Level of the Combinations of the Other Two Factors in the Cloze Tests (n = 76)

Source		<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups & B: Time	C (Con & Pre)	14590.66	1	14590.66	105.820	.000 ***	.311
	C (Con & Post)	2292.63	1	2292.63	16.627	.000 ***	.049
	C (Exp & Pre)	8894.74	1	8894.74	64.510	.000 ***	.190
	C (Exp & Post)	707.60	1	707.60	5.132	.025 *	.015
	Error	20406.53	148	137.88			
A: Groups & C: Word Categories	B (Con & Content)	928.42	1	928.42	6.235	.014 *	.013
	B (Con & Function)	39658.03	1	39658.03	266.315	.000 ***	.565
	B (Exp & Content)	2457.60	1	2457.60	16.503	.000 ***	.035
	B (Exp & Function)	5089.18	1	5089.18	34.175	.000 ***	.073
	Error	22039.31	148	148.91			
B: Time & C: Word Categories	A (Pre & Content)	921.07	1	921.07	2.718	.100	.008
	A (Pre & Function)	3229.53	1	3229.53	9.529	.002 **	.029
	A (Post & Content)	2469.61	1	2469.61	7.287	.007 **	.022
	A (Post & Function)	5037.56	1	5037.56	14.863	.000 ***	.045
	Error	100322.50	296	338.93			

***: $p < .001$, **: $p < .01$, *: $p < .05$

In addition, since significant simple interactions were found between

time and word categories at both levels of groups and between groups and time for both content and function words, simple main effects of the three factors at all levels were examined (Table 8.6). Each of the three simple interactions are also shown in graphs (Figures 8.2, 8.3, and 8.4).

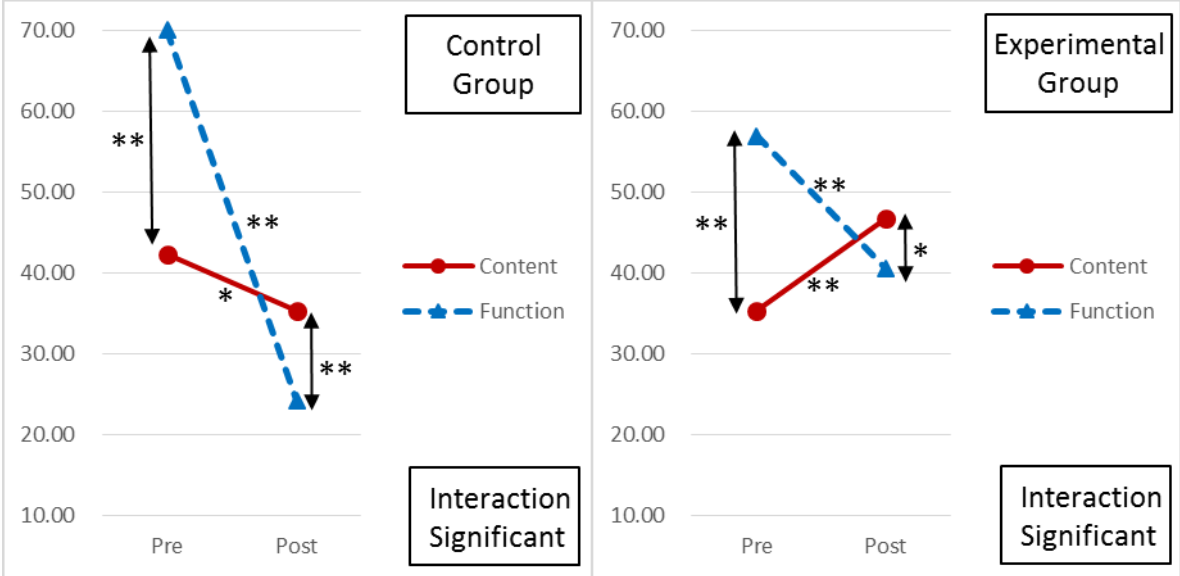


Figure 8.2. Simple interactions between time and word categories in the cloze tests. (**: $p < .01$, *: $p < .05$).

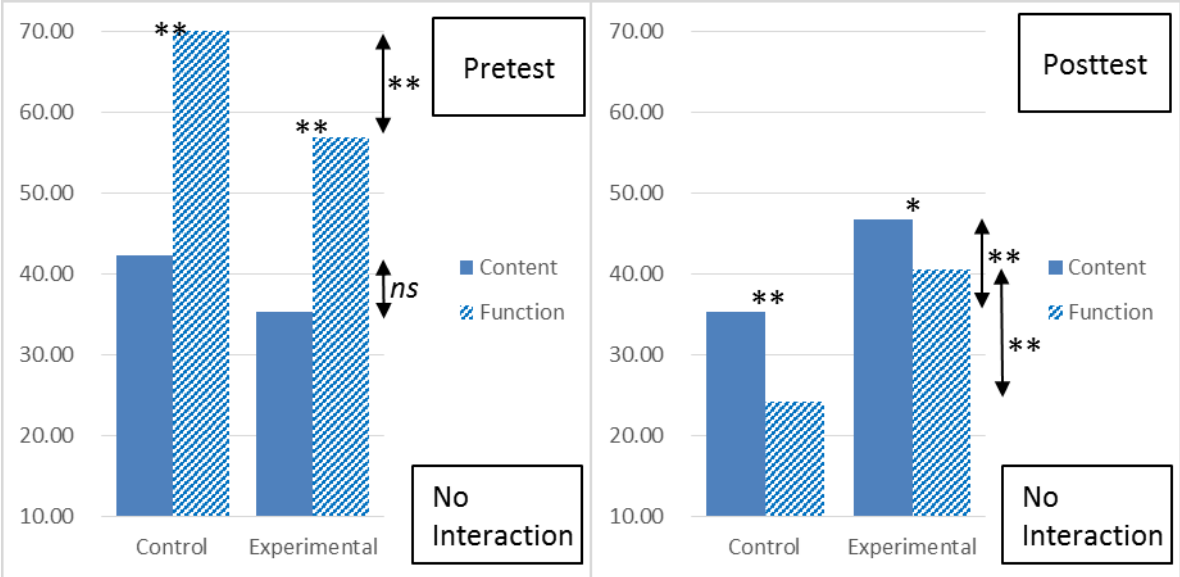


Figure 8.3. Simple interactions between groups and word categories in the cloze tests. (**: $p < .01$, *: $p < .05$).

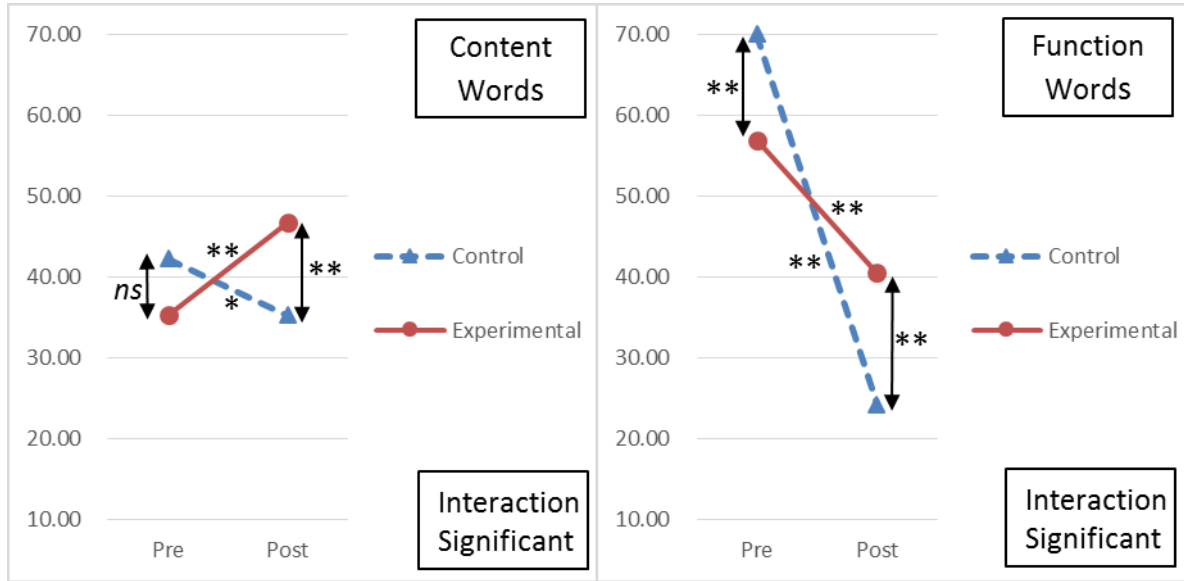


Figure 8.4. Simple interactions between groups and time in the cloze tests. (**: $p < .01$, *: $p < .05$).

First, as for the simple interaction between time and word categories, significant interaction was found both for the control ($F(1, 74) = 124.357$, $p = .000$, partial $\eta^2 = .474$) and the experimental ($F(1, 74) = 63.903$, $p = .000$, partial $\eta^2 = .244$) groups. In the control groups, recognition of both content and function words deteriorated significantly (content: $F(1, 148) = 6.235$, $p = .014$, partial $\eta^2 = .013$, function: $F(1, 148) = 266.315$, $p = .000$, partial $\eta^2 = .565$) from the pretest to the posttest. However, the deterioration was milder for content words than for function words. As a result, although in the pretest function words were recognized significantly better than content words ($F(1, 148) = 105.820$, $p = .000$, partial $\eta^2 = .311$), recognition of content words was significantly better than that of function words in the posttest ($F(1, 148) = 16.627$, $p = .000$, partial $\eta^2 = .049$).

On the other hand, as far as the experimental group is concerned, recognition of content words improved significantly ($F(1, 148) = 16.503$, $p = .000$, partial $\eta^2 = .035$) while that of function words significantly deteriorated ($F(1, 148) = 34.175$, $p = .000$, partial $\eta^2 = .073$) from the pretest

to the posttest. Consequently, for the experimental group as well, even though recognition of function words was significantly better in the pretest than that of content words ($F(1, 148) = 64.510, p = .000, \text{partial } \eta^2 = .190$), content words were recognized significantly better in the posttest ($F(1, 148) = 5.132, p = .025, \text{partial } \eta^2 = .015$).

These results indicate that function word recognition was more difficult than that of content words in the posttest, which falls in line with the results of past studies. However, in the pretest, it seems that this was not the case. Nevertheless, the difference between the two groups in the way the two factors interacted, as can also be seen from the graphs (Figure 8.2), implicates some positive effects of the treatment on the experimental group's word recognition.

Second, no significant simple interaction was found between groups and word categories both in the pretest ($F(1, 148) = 2.543, p = .113, \text{partial } \eta^2 = .017$) and in the posttest ($F(1, 148) = 1.642, p = .202, \text{partial } \eta^2 = .011$), and recognition of content words was significantly more difficult in the pretest, while in the posttest it was easier, for both groups. This corroborates the assumption that recognition of function words was easier in the pretest and more difficult in the posttest than that of content words.

Finally, between the two factors of groups and time, significant simple interactions were found both for content ($F(1, 148) = 21.513, p = .000, \text{partial } \eta^2 = .096$) and function ($F(1, 148) = 54.844, p = .000, \text{partial } \eta^2 = .244$) words. As for content word recognition, the participants in the experimental group fared significantly better in the posttest than those in the control groups ($F(1, 296) = 7.287, p = .007, \text{partial } \eta^2 = .022$), even though the difference between the two groups in the pretest was not significant ($F(1, 296) = 2.718, p = .100, \text{partial } \eta^2 = .008$). Further, although

the participants of the experimental groups fared significantly worse in recognizing function words in the pretest than those of the control group ($F(1, 296) = 9.529, p = .002, \text{partial } \eta^2 = .029$), their recognition of function words was significantly better in the posttest than that by the control group ($F(1, 296) = 14.863, p = .000, \text{partial } \eta^2 = .045$).

This indicates that, as can also be clear from the graphs (Figure 8.4), that some positive effects of the treatment were found in the experimental group both for content and function word recognition.

8.3.2.2 Results of Fisher’s Exact Tests

In order to examine the effects of the treatment in detail, Fisher’s exact tests⁷ were conducted. The numbers of right (R) and wrong (W) transcriptions by the control and the experimental groups for all the 20 words in the posttest were computed and which of the 20 words in the posttest made a significant difference in recognition between the control and the experimental groups was examined.

Table 8.7.
Words in Which There Was Significant Difference in Recognition Between the Control and the Experimental Groups in the Cloze Posttest (n = 76)

Group	1 shirt		4 look		7 would		13 don't		18 your		19 a	
	R	W	R	W	R	W	R	W	R	W	R	W
Control (n = 40)	1	39	13	27	2	38	18	22	10	30	1	39
Experimental (n = 36)	7	29	22	14	8	28	32	4	19	17	13	23
<i>p</i>	.023 *		.021 *		.040 *		.000 ***		.018 *		.000 ***	
Cramer's V	.276		.287		.254		.462		.286		.433	

Function words in italics. ***: $p < .001, *: p < .05$

Results are shown in Table 8.7. Significant differences were found in 6 words, out of which two words are content and the other four were

function words. The greatest difference was found in the word No. 13 *don't* ($p = .000 < .001$, Cramer's $V = .462$).

8.3.3 Results of the Paused Transcription Tests

Table 8.8 shows the descriptive statistics of the words correctly recognized in percentage in the pretest (Cronbach's alpha = .719) and the posttest (Cronbach's alpha = .728) of the paused transcription tests.

Table 8.8.

Descriptive Statistics of the Paused Transcription Tests (Words Correctly Recognized, %, $n = 76$)

Group	n	Pretest						Posttest					
		Content		Function		Total		Content		Function		Total	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Control	40	53.00	11.39	23.75	9.86	38.38	9.57	44.13	11.93	18.88	7.46	31.50	8.53
Experimental	36	47.64	12.56	24.17	11.40	35.90	10.63	48.89	8.75	25.14	7.77	37.01	7.51

8.3.3.1 Results of a Three-Way ANOVA

For the paused transcription tests as well, a three-way ANOVA⁷ (A: groups: control/experimental, B: time: pre/post, C: word categories: content/function) was first carried out to analyze the data. The results are shown in Table 8.9. Figure 8.5 shows the means of correct word recognition in percentage for each of the three factors.

The results of the ANOVA showed that the second-order interaction was not significant ($F(1, 74) = 2.137$, $p = .148$, partial $\eta^2 = .001$). However, significant interactions were found between groups and time ($F(1, 74) = 17.356$, $p = .000$, partial $\eta^2 = .014$), and between groups and word categories ($F(1, 74) = 4.022$, $p = .049$, partial $\eta^2 = .003$). Therefore, simple main effects of each factor at each level of the other factor for the two interactions were examined respectively (Tables 8.10 and 8.11).

Table 8.9.

The Results of the Three-Way ANOVA for the Paused Transcription Tests (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	175.30	1	175.30	0.665	.417	.002
S: Error (A)	19505.63	74	263.59			
B: Time	629.48	1	629.48	9.041	.004 **	.008
Interaction (AB)	1208.42	1	1208.42	17.356	.000 ***	.014
Error (BS)	5152.43	74	69.63			
C: Word Categories	49014.05	1	49014.05	785.829	.000 ***	.586
Interaction (AC)	250.89	1	250.89	4.022	.049 *	.003
Error (CS)	4615.56	74	62.37			
Interaction (BC)	65.63	1	65.63	1.618	.207	.001
Second-Order Interaction (ABC)	86.68	1	86.68	2.137	.148	.001
Error (BCS)	3001.81	74	40.56			
Total	83705.87	303				

***: $p < .001$, **: $p < .01$, *: $p < .05$

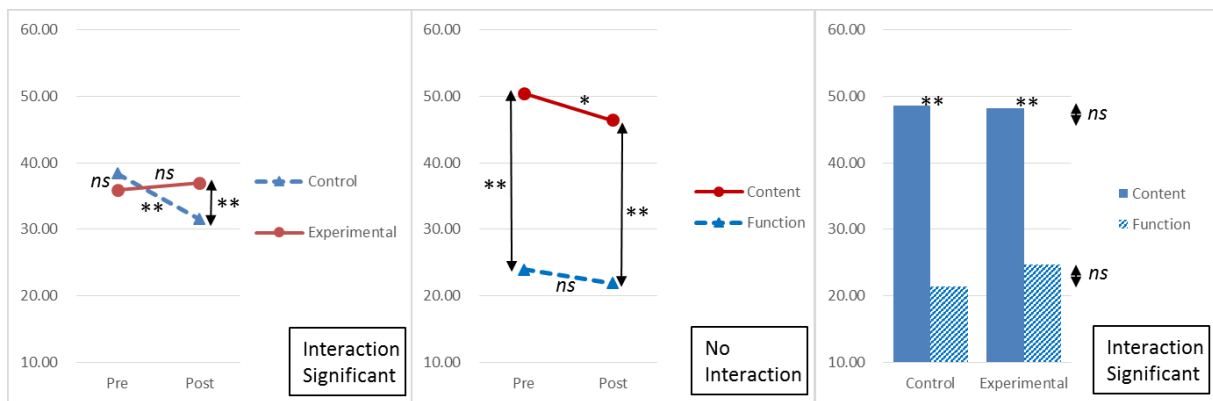


Figure 8.5. Words successfully recognized in percentage for each two of the three factors in the paused transcription tests. (**: $p < .01$, *: $p < .05$).

Table 8.10.

Simple Main Effects in AB Interaction (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups (Pre)	231.61	1	231.61	1.390	.240	.009
A: Groups (Post)	1152.11	1	1152.11	6.915	.010 **	.044
Error	24658.06	148	166.61			
B: Time (Control)	1791.12	1	1791.12	25.724	.000 ***	.256
B: Time (Experimental)	46.78	1	46.78	0.672	.415	.007
Error	5152.43	74	69.63			

***: $p < .001$, **: $p < .01$

Table 8.11.

Simple Main Effects in AC Interaction (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups (Content)	3.38	1	3.38	0.021	.886	.000
A: Groups (Function)	422.81	1	422.81	2.594	.109	.017
Error	24121.18	148	162.98			
C: Word Categories (Control)	28139.21	1	28139.21	451.149	.000 ***	.522
C: Word Categories (Experimental)	21125.73	1	21125.73	338.703	.000 ***	.392
Error	4615.56	74	62.37			

***: $p < .001$

First, as for the interaction between groups and time, the simple main effect of groups in the pretest was not significant ($F(1, 148) = 1.390$, $p = .240$, partial $\eta^2 = .009$), while in the posttest the difference between the groups was significant ($F(1, 148) = 6.915$, $p = .010$, partial $\eta^2 = .044$). Further, the simple main effect of time for the control group was significant ($F(1, 74) = 25.724$, $p = .000$, partial $\eta^2 = .256$), which means that, as is evident from the graph (Figure 8.5), word recognition by the control group significantly deteriorated from the pretest to the posttest. On the other hand, a slight gain in word recognition logged by the experimental group from the pretest to the posttest was not significant ($F(1, 74) = 0.672$, $p = .415$, partial $\eta^2 = .007$).

These results implicate that, even though there was no significant difference in the pretest between the groups in terms of word recognition, with content and function words combined, a difference emerged in the posttest and the participants of the experimental group fared significantly better than those of the control group. This suggests that there might have been some positive effects of the treatment the experimental group received on their word recognition as a whole.

Second, as for the interaction between groups and word categories, no significant difference was found between the control group and the

experimental group in the respective recognition of content ($F(1, 148) = 0.021, p = .886, \text{partial } \eta^2 = .000$) and function ($F(1, 148) = 2.594, p = .109, \text{partial } \eta^2 = .017$) words. On the other hand, there were significant differences between the recognition of content words and that of function words for both the control ($F(1, 74) = 451.149, p = .000, \text{partial } \eta^2 = .522$) and the experimental ($F(1, 74) = 338.703, p = .000, \text{partial } \eta^2 = .392$) groups. The interaction was barely significant ($p = .049$) and it can be assumed that, in both the pretest and the posttest, recognition of function words was far more difficult than that of content words across both the groups, which falls in line with the past studies, since there was no significant interaction between time and word categories ($F(1, 74) = 1.618, p = .207, \text{partial } \eta^2 = .001$) and the main effect of word categories was significant ($F(1, 74) = 785.829, p = .000, \text{partial } \eta^2 = .586$), which is also evident from the graphs (Figure 8.5).

8.3.3.2 Results of Two-Way ANOVAs for Content and Function Words

In order to more thoroughly examine the effects of the treatment on spoken word recognition, recognition of content words and that of function words were separately examined, using two-way mixed ANOVAs⁷ (A: groups: control/experimental, B: time: pre/post). Tables 8.12 and 8.13 show the results of the ANOVAs for recognition of content words and that of function words respectively. Figure 8.6 shows the means of correct content and function word recognition in percentage by the two groups at the pretest and the posttest.

Table 8.12.

The Results of the Two-Way ANOVA for Content Word Recognition in the Paused Transcription Tests (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	3.38	1	3.38	0.017	.896	.000
S: Error (A)	14667.67	74	198.21			
B: Time	550.81	1	550.81	8.772	.004 **	.026
Interaction (AB)	971.20	1	971.20	15.467	.000 ***	.046
Error (BS)	4646.56	74	62.79			
Total	20921.05	151				

***: $p < .001$, **: $p < .01$

Table 8.13.

The Results of the Two-Way ANOVA for Function Word Recognition in the Paused Transcription Tests (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups	422.81	1	422.81	3.310	.073	.030
S: Error (A)	9453.51	74	127.75			
B: Time	144.30	1	144.30	3.044	.085	.010
Interaction (AB)	323.91	1	323.91	6.833	.011 *	.023
Error (BS)	3507.67	74	47.40			
Total	13876.32	151				

*: $p < .05$

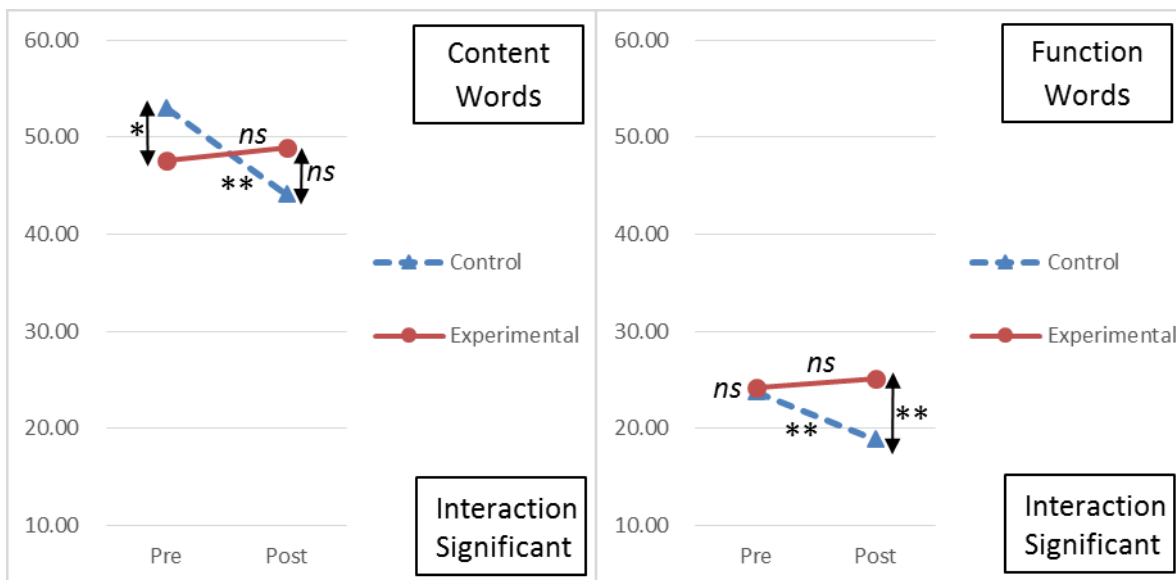


Figure 8.6. Content and function words successfully recognized in percentage at the pretest and the posttest. (**: $p < .01$, *: $p < .05$)

The results of the ANOVAs show that significant interactions were found for both content ($F(1, 74) = 15.467$, $p = .000$, partial $\eta^2 = .046$) and function ($F(1, 74) = 6.833$, $p = .011$, partial $\eta^2 = .023$) word recognition. Therefore, simple main effects for both the interactions were analyzed (Tables 8.14 and 8.15).

Table 8.14.

Simple Main Effects in the Interaction for Content Word Recognition (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups (Pre)	544.58	1	544.58	4.173	.043 *	.034
A: Groups (Post)	430.00	1	430.00	3.295	.072	.026
Error	15268.69	117	130.50			
B: Time (Control)	1575.31	1	1575.31	25.088	.000 **	.252
B: Time (Experimental)	28.13	1	28.13	0.448	.505	.005
Error	4646.56	74	62.79			

** $: p < .01$, * $: p < .05$

Table 8.15.

Simple Main Effects in the Interaction for Function Word Recognition (n = 76)

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>	η_p^2
A: Groups (Pre)	3.29	1	3.29	0.038	.847	.000
A: Groups (Post)	743.42	1	743.42	8.489	.004 **	.065
Error	10684.22	122	87.58			
B: Time (Control)	475.31	1	475.31	10.027	.002 **	.119
B: Time (Experimental)	17.01	1	17.01	0.359	.551	.004
Error	3507.67	74	47.40			

** $: p < .01$

First, as for content words, recognition by the control group deteriorated significantly from the pretest to the posttest ($F(1, 74) = 25.088$, $p = .000$, partial $\eta^2 = .252$), while the experimental group's recognition only slightly improved during the same period even though the difference between the pretest and the posttest was not significant ($F(1, 74) = 0.448$,

$p = .505$, partial $\eta^2 = .005$). As a result, although, in the pretest, the control group fared significantly better in content word recognition ($F(1, 117) = 4.173$, $p = .043$, partial $\eta^2 = .034$), the experimental group fared marginally better in the posttest than the control group ($F(1, 117) = 3.295$, $p = .072$, partial $\eta^2 = .026$).

Second, as for function words, there was no significant difference in recognition between the two groups at the pretest ($F(1, 122) = 0.038$, $p = .847$, partial $\eta^2 = .000$). However, the control group's recognition deteriorated significantly from the pretest to the posttest ($F(1, 74) = 10.027$, $p = .002$, partial $\eta^2 = .119$), while the experimental group's recognition in the posttest was only slightly better than in the pretest, even though the difference in recognition between the two tests was not significant ($F(1, 74) = 0.359$, $p = .551$, partial $\eta^2 = .004$). Therefore, significant difference in function word recognition between the two groups emerged at the posttest ($F(1, 122) = 8.489$, $p = .004$, partial $\eta^2 = .065$).

These results indicate that, as can also be evident from the graphs in Figure 8.6, some positive effects of the treatment were found in the experimental group both for content and function word recognition in the paused transcription tests as well.

8.3.3.3 Results of Fisher's Exact Tests

In order to analyze in which word differences in recognition between the two groups were found in the posttest, Fisher's exact tests⁷ were conducted for the paused transcription tests as well. The numbers of right (R) and wrong (W) transcriptions by the control and the experimental groups for all the 40 words in the posttest were computed.

However, significant difference was found only in two words, both

content words, out of 40: *visit* in No. 1 ($p = .031$, Cramer's $V = .269$), and *beautiful* in No. 6 ($p = .022$, Cramer's $V = .272$).

8.4 Discussion

First, given the results of the ANOVAs for both the cloze tests and the paused transcription tests, it is likely that the treatment given to the experimental group had some positive effects on their spoken word recognition. In addition, those positive effects were not only found on recognition of content words but also on that of function words.

Second, the results of the Fisher's exact tests for the cloze posttest indicate that out of the 20 words targeted for transcription, significant differences were found in 6 words: 2 content and 4 function words. Of these words, No. 7 *would* in 'I think it would look good on you,' No.18 *your* in 'Take your time,' and No. 19 *a* in 'I'm not in a hurry' would be worthy of attention. The first phrase 'I think it would look good on you' has a rhythm of WSWWWSWW and the targeted word consisted of one weak syllable. This is exactly the case where, due to English stress-timed rhythm, three weak syllables 'it would look' sandwiched between two strong syllables 'think' and 'good' were pronounced in quick succession with three syllables attached to one another. In order to successfully recognize the word *would*, it takes the listener to first recognize the stress unit as a whole before separating the unit and identifying each word in it by taking advantage of phrasal, grammatical, contextual and other related knowledge. There is no denying that the treatment given to the participants of the experimental group might have helped them recognize the word.

Likewise, the second phrase 'Take your time' consisted of three syllables, which have a rhythm of SWS. This is another example, in which

the targeted word *your* was made up of one weak syllable sandwiched between two stressed ones. The third phrase ‘I’m not in a hurry’ has a rhythm of WSWWSW, in which the targeted word *a*, one-syllable word, falls on the second weak syllable of the two unstressed ones sandwiched between the stressed syllables. This is still another typical example of English phrases uttered in a stress-timed manner, in which the listener is required to recognize a weak syllable in a sequential articulation of one stressed syllable and several unstressed ones bunched together to keep the rhythm from getting broken. The fact that there were significant differences found in recognition of these words between the two groups implicates that there were some positive effects of the treatment given to the experimental group on their recognition of not only stressed syllables but also weak ones.

On the other hand, in the analyses of the post-paused-transcription test, significant difference between the groups were found only in two words: *visit* in No. 1 in ‘Lucy comes to visit her aunt,’ and *beautiful* in No. 6 in ‘They all go to a big dance party in a beautiful house.’ While the rhythm in the former was SWSWSWWS and in the latter it was WSWWWSSSWWSWWSW, both words contained a syllable enunciated with stress. The paused transcription test might have been too challenging for the participants to correctly catch unstressed syllables articulated quickly between the intervals of stressed syllables.

In the treatment, participants of the experimental group were given some instructions focused on English phonological features and, following these explicit explanations, they were asked to listen to sentences and dialogues, paying attention to those features. They also articulated themselves those sentences and dialogues, trying not to deviate from these English phonological rules, especially the stress-timed rhythm of the

language.

It can safely be said that this kind of treatment can have positive effects on spoken word recognition regardless of word categories. This means that the treatment is presumably effective in strengthening Japanese EFL learners' ability to recognize elusive unstressed syllables, which are usually articulated very quickly in a stress-timed language like English. The results also corroborate the claim by Eastman (1993) and Vanderplank (1993) that it would be beneficial to spoken word recognition, especially recognition of unstressed syllables, for a syllable-timed speaker to undergo some forms of perception and articulation practices, strictly following the English stress-timed rhythmic patterns, preceded by explicit explanation on the English phonological and rhythmic rules.

When Japanese EFL learners try to acquire English vocabulary, they typically undergo the following steps; they first learn the spelling of a word with its Japanese translation attached and then they learn how to pronounce the word in isolation. One of the greatest differences between L1 speakers of English and Japanese EFL learners in the way they acquire vocabulary is that Japanese EFL learners are first given each word cut out from the context, not sequences of words with the information of their prosody, while English native speakers usually acquire each of those words by first hearing those sequences contained in the context with all the prosodic features attached. This may be why Japanese EFL learners have difficulty in recognizing each syllable or word, especially unstressed ones or function words, contained in the sequences articulated in a stress-timed manner. It can be presumed, therefore, that, given that English stress-timed rhythm plays a critical role in the way the spoken text is articulated, successful spoken word recognition cannot be expected without instructors

folding this element into some form of teaching method. Not so, especially if the listener is a speaker of a mora-timed language.

Notes

1. The materials used for these experiments and treatment were all different, even though they were from the same listening textbook, *Kyukyoku-no-eigo-listening* (Ultimate English listening) series level 1, published by ALC Press.
2. Materials used in the pretest were a dialogue of 175 words with the recording time of 60 seconds for the cloze test and a monologue of 219 words in 90 seconds for the paused transcription test. In the posttest, a dialogue of 162 words in 53 seconds was used for the cloze test and a monologue of 206 words in 82 seconds for the paused transcription test. In terms of wpm, there was a difference in rates between the dialogues and the monologues. However, all the materials used were from the most difficult level, as was the case in Experiment 3, of the three levels available in the listening textbook. According to Tauroza and Allison (1990), dialogues and monologues in spoken English are basically different in speech rates. They claim that the average speech rate of dialogues in British English is around 210 wpm while that of monologues around 140 wpm. They say that dialogues are faster than monologues because, unlike in monologues, speakers usually fight for the control of the floor in conversational situations. They further claim that there tend to be many more simple words used in dialogues than in monologues so that monologues tend to have more syllables per word, which means that there is no such big difference in speech rate in terms of spm.

3. Quirk et al. (1985) was referred to in distinguishing the two word categories.
4. The way the transcription of each targeted word was judged to be correct or otherwise was in conformity with the method used in Experiment 4 in the cloze tests and with Experiment 3 in the paused transcription tests.
5. The texts used were only dialogues because a body of literature shows that the English stress-timed rhythm, the alteration of stressed and unstressed syllables, is more distinct and that phonetic changes occur more frequently in dialogues than in monologues (Kohno, 1993; Osada, 2004; Jones, 2008). For reference, the average speech rate of the dialogues used was 160 wpm.
6. The magazines used in the treatment were a series of *English Journals*, published by ALC Press.
7. In analyzing the data, an online software of ANOVA 4 was used for the three-way ANOVAs and a Microsoft add-in software for Excel was used for the two-way ANOVAs, for the *t*-test and for the Fisher's exact tests.

Chapter 9

Conclusion

This chapter concludes the present study. First, we discuss the results of the experiments in Chapters 4 to 8. Then, the major findings of this study will be stated. Finally, implications and limitations of the study will be given.

9.1 General Discussion

First, as the results of Experiment 1 (Chapter 4) show, recognition of function words is more challenging than that of content words for lower-proficiency Japanese EFL learners. The results of the experiment also implicate that speech rate is an important variable in recognition of content and function words.

Second, in Experiment 2 (Chapter 5), treatment in which Japanese translations were given before dictation practices and instructions were provided to make inferences from them about the text the participants would soon hear had positive effects on spoken word recognition. This might well have resulted from some form of reinforcement on the top-down processing, through application of such strategies as semantic and contextual inferences.

In addition, the treatment was no less effective in enhancing the recognition of function words than that of content words. The enhancement in the recognition of function words may well have been caused by the increased amount of content words recognized through the application of top-down strategies, coupled with some reference to internalized linguistic

knowledge, which also constitutes one of the top-down strategies. However, in making inferences on function words they failed to recognize, there must have been no other recourse for them but to rely on the grammatical and phrasal knowledge they already had, because they had not been given any reinforcement concerning such linguistic knowledge.

Third, in Experiment 3 (Chapter 6), even though grammatical and phrasal knowledge was given in the treatment, no positive effects on word recognition were observed when the speech rate was around 185 wpm. At a slower speech rate of around 130 wpm, on the other hand, some positive effects were witnessed mainly on recognition of content words.

However, effects on function word recognition was only limited. In utilizing top-down strategies, sufficient amount of linguistic knowledge is a must. In addition, in order to make inferences on elusive weak syllables, one must rely considerably on one's knowledge about grammar and phrases, or formulaic sequences. Therefore, it can be inferred that, at the higher rate, bottom-up processing, perception of the speech stream itself, was not sufficient enough to catch even stressed syllables. Hence, it was impossible to recognize content words as well as sequences of words that contain the stressed syllables, still less function words. At the slower rate, it was possible to recognize content words in the stress unit. However, the recognition of content words did not necessarily lead to that of function words, even with the fortified knowledge about the phrases in question.

The reasons would be threefold why the participants, even at the slower rate, were unable to make sufficient inferences on elusive weak syllables from the information they had about the content words. First, unlike in Experiment 5 (Chapter 8), the treatment did not involve instructions concerning English phonological features and the participants

would have hardly imagined that spoken texts in English should first be recognized as stress units, or chunks of words which contain stressed syllables, and that only later should those units be divided into individual words. Therefore, they might have had no idea about how they should utilize the knowledge on the information they successfully perceived.

Second, unlike in Experiment 2 (Chapter 5), the treatment did not involve instructions on the importance of making inferences. The participants of the experimental group in Experiment 3 (Chapter 6), therefore, might not have expected that missing function words must be inferred. Third, in the treatment, handouts were given about the related grammatical and phrasal knowledge. The participants were given explicit explanation about it and asked to repeat the sentences that contained those phrases and grammatical items after the instructor many times. However, they did not listen to those sentences articulated by a native speaker of English in a stress-timed manner at a natural, moderately high speech rate. Therefore, it is highly likely that those grammatical items and phrases, or formulaic sequences, were not appended with prosodic information unique to those sequences, which means that those kinds of added knowledge were not just 'readily available' even if they tried to recognize the spoken text by the stress unit and then to divide the unit into individual words through inferences.

In addition, the participants in Experiment 2 (Chapter 5) listened to an English spoken text at the moderately high speech rate of around 170 wpm during the treatment period. That is, they had opportunities to perceive English texts spoken in the stress-timed manner and at the natural rate for about two months and a half, which might have led to the results that positive effects on spoken word recognition was found even at

the rate of around 170 wpm. On the other hand, the participants in Experiment 3 (Chapter 6) did not have opportunities to listen to CDs of English articulated in a natural way at a natural rate during the treatment, even though they had oral reading practices, using the handouts about the related grammatical and phrasal knowledge. Consequently, positive effects on spoken word recognition were observed only in the case of content words and also at the speech rate of around 130 wpm.

One cannot deny the importance, and possible positive effects on spoken word recognition, of reinforcing knowledge about grammar and phrases. However, it is highly likely that it must be augmented by some perception, and quite possibly articulation, practices. In addition, the spoken texts given should be the ones articulated in a natural stress-timed manner and at a moderately high speech rate. Given the results of Experiments 2 (Chapter 5) and 5 (Chapter 8), it may also be important to give listeners knowledge about English phonological features including the stress-timed rhythm and to tell them that inferences play a greater role in spoken word recognition.

Fourth, the results of Experiment 4 (Chapter 7) show that continuous exposure to speech delivered at a high-speed rate of about 1.5 times the normal rate had positive effects on spoken word recognition of speech delivered at the original baseline rate. However, here again, effects on recognition of function words were limited. Even though the participants did constantly listen to English spoken text at the compressed rate of about 190 wpm on average, they were not instructed about English rhythmic features or how English speech should be recognized. They probably did not have any idea whether speech should be segmented by the chunk of stress units or whether missing unstressed syllables can only be inferred if they

use their linguistic knowledge.

Fifth, on the other hand, in Experiment 5 (Chapter 8), the treatment had positive effects on both content and function words. Both the results of the statistical analyses indicated that the explicit explanations given to the participants on the English stress-timed and other phonological features with the subsequent perception and articulation practices focused on those features enabled them to recognize spoken English words in a more ‘natural’ way; to segment speech by the stress unit and to recognize individual words in the unit by turning to their internalized linguistic knowledge. The treatment did not involve any of the augmentation concerning grammatical and phrasal knowledge, unlike in Experiment 3 (Chapter 6), which means that each participant had to make inferences on missing weak syllables on their own, utilizing the knowledge they already had at the time. However, in the perception and articulation practices using dialogues, they are quite likely to have learned some of the expressions, often-used phrases and formulaic sequences, with prosodic information attached, even though materials used in the treatment did not share the same expressions with the posttest, as was the case in Experiment 3 (Chapter 6).

On balance, the treatment applied in Experiment 5 (Chapter 8) was effective, probably because they did perception and articulation practices in a way similar to those that would have been gone through by L1 speakers of English: they listened and copied the sound not by the individual word but by the chunk of words, after they were given the explicit instruction. Nevertheless, the results of the statistical analysis on the individual words was not sufficient to corroborate the effectiveness on recognition of function words. One of the possible reasons for this is that the amount of perception and articulation practices was fairly small. Weekly treatment that lasted

about 30 minutes would not have been sufficient to make a significant difference in the analyses on individual word recognition.

As we have seen in Chapter 3, English speech is supposed to be parsed by the unit of several words which contains both stressed and unstressed words. It also seems that recognition process begins with that of stressed syllables followed by recognition of the word, mainly content, which has that stressed syllable, and then finally the process goes into the search for other weak syllables and function words that have these weak syllables in them. In parsing English speech, therefore, Japanese EFL learners, whose L1 is mora-timed and who are unfamiliar to this parsing unit and the rhythm, quite presumably find it challenging to recognize unstressed syllables, which most of the function words are made up of.

It is true that, in spoken word recognition, both bottom-up and top-down processing is necessary. However, it must begin first with the processing of the sound that listeners have perceived and the sound they hear in English spoken text has all the phonological and prosodic features that are unique to English language. Given that the Japanese language scarcely, if ever, shares these features, perception of English speech is quite challenging for Japanese EFL learners, especially those with lower levels of proficiency, even before activating top-down strategies.

It seems, therefore, that what is first to be addressed is to have learners get accustomed to English phonological and prosodic features, especially its stress-timed rhythm. Accordingly, it is very important for instructors in Japan to try to make their students, first and foremost, perceive the sound by the chunk of a stress unit as a whole. It goes without saying, however, that is not all that has to be addressed. In order to break down the unit, learners need such linguistic knowledge as grammar,

phrases or formulaic sequences. In addition, in order to give them such knowledge with prosodic features attached, instructors must give their students a lot of perception and articulation practices. It is also important to tell them to make inferences on missing information, based on the words they have successfully recognized. Last but not least, speech rate is yet another important factor. It is necessary for instructors to have them listen to English speech at a moderately high speech rate.

9.2 Major Findings of the Study

The primary goal of the present study is to propose some effective teaching methods of enhancing spoken word recognition by Japanese EFL learners with lower levels of proficiency by both utilizing top-down and bottom-up strategies. Major findings of the study can be summarized in the following four points.

First, from the perspective of the bottom-up processing, it will be effective to have learners get accustomed to English natural speech, which is articulated in a distinct stress-timed manner with stressed syllables enunciated strong and other syllables destressed. In addition, to make learners practice the speech themselves, accurately copying the sound without getting English phonological features distorted (e.g. no insertion of extra vowels) and repeating, shadowing, and oral reading with the English stress-timed rhythm. They should practice articulating, for example, word sequences or chunks such as *would have been possible* or *out of the blue*, not as a group of separate individual words, but as one unit bound together with weak syllables completely destressed and reduced, with consonant clusters intact and vowels in schwas, or maybe sometimes deleted.

When Japanese EFL learners, especially those with lower levels of proficiency, try to remember these sequences or phrases, they quite often depend solely on their visual image, that is, their written representations, and try to link their written image or spellings with meanings or Japanese translation. In addition, if instructors ask them to listen and repeat, to attach phonetic information with those sequences and phrases and to read them aloud in a proper pronunciation when they try to remember them, then these learners very often end up reading aloud each individual words separately, and sometimes quite haltingly, reconstituting unstressed syllables to their full salient forms and inserting unnecessary vowels between consonant clusters and after codas. They should, instead, and it is necessary for the instructors to require them to, link and bind the words together and enunciate those sequences and phrases as a whole, not a group of separate words but one unit, and in the observance of proper English phonological rules, not inserting extra vowels and pronouncing each phoneme, especially consonants (vowels can be reduced to schwas), correctly, without distorting them into different ones in accordance with the Japanese phonemic system. They should also be told to articulate each sequence, and for that matter longer sentences as well, with certain syllables stressed and others destressed, keeping the proper English stress-timed rhythm. In order for that to happen, explicit instructions on English phonological and prosodic knowledge should be fully given to the learners, preceding the perception and articulation practices.

Second, utilization of higher-than-normal speech rate, about 1.5 times faster, should also be recommended to shore up the bottom-up processing. The results of the present study show that constant exposure to speech delivered at a compressed rate of about 67 percent is effective in enhancing

spoken word recognition. When instructors use in class mechanically compressed speech, they have to be careful about how fast it should be.

Third, from the perspective of the top-down processing, it seems that giving listeners meanings, Japanese translation of the script, before dictation practices has some positive effects on their subsequent spoken word recognition. If they constantly keep doing these practices, it will help them form a habit of accessing the meanings from what they hear and making inferences on the words they have missed, which enables them to use top-down strategies effectively, even without a scaffolding of the meanings given in advance. It also seems to be important, when giving them Japanese translation, to instruct them to guess English they will soon hear. This might help them pay attention to the forms they are going to listen to, which will lead to successful recognition of not only content but also function words.

Fourth, even though there is no denying that giving as much grammatical and phrasal knowledge as possible will certainly help listeners in spoken word recognition because such knowledge would enable them to guess words they cannot instantly recognize, its effects would be limited, should they not be given any reinforcement from the bottom-up processing, such as the one related to speech rate or to English phonological and prosodic features. Such knowledge would become most effective when given with prosodic information. Listeners must acquire the knowledge through a lot of perception and articulation practices, which also must be aided by the instructor's explicit explanation on the English phonological features and its unique rhythm. In addition, in implementing such practices, it would be best to use the CDs in which sequences of words or sentences are articulated at a moderately high speech rate and in a stress-

timed manner proper to the English language. Furthermore, instructions on the importance of making inferences by taking advantage of such knowledge would also be helpful.

9.3 Implications and Limitations of the Study

9.3.1 Implications for English Education in Japan

When spoken word recognition is improved, hopefully close to the automatic level, one of the major differences between the skills necessary in reading and listening will be bridged. In other words, it will enable listeners to perceive the sound they hear not as noise but as meaningful linguistic signals on the same level as they would in visually following along the written characters on the paper. Figuratively speaking, it is as if the listener is trying to enjoy an English film with English subtitles on; the script is on the screen over the period the line is uttered.

It is true, nonetheless, that enhanced spoken word recognition does not guarantee enhanced comprehension of the text, since there are other factors involved, the biggest of which would be whether the text can be processed in real time or not. Considering, however, that one of the greatest hurdles Japanese EFL learners face in listening comprehension is whether they can successfully segment speech and recognize words in it, as was discussed in Chapter 2, adopting effective pedagogical methods in enhancing spoken word recognition is a path for instructors of the English language in Japan to take.

What is unfortunate for Japanese EFL listeners, however, whose mother tongue is mora-timed and the unit of its word recognition is said to correspond to one mora, is that listening materials in general, used in Japanese educational environment, are often less stress-timed than it

should be. Jones (2008) observes that there exist listening materials which are in a sense ‘over-pronounced’ (p. 18), where weak forms are stressed and the rhythm of speech is distorted. Naturally, the words in these over-pronounced and less stress-timed spoken texts are easier for Japanese EFL learners to recognize and, therefore, the texts themselves are less challenging to comprehend, since their rhythm are distorted toward one of a syllable-timed language, where weak syllables are stressed.

Furthermore, according to Yanagawa (2016), in the Center Listening Test, which is used for gate-keeping purposes by all state-run and public universities and more than 90 percent private universities across Japan, and, therefore, influential not only for high-school students but also for many educational institutions, the spoken texts used do not represent the nature of real-life speeches, especially in their speech rate and lack of reduced forms. In addition, local English teachers in high schools across Japan seem to feel positive about the current practice of omitting reduced forms and enouncing every word rather clearly (Yanagawa, 2016). In other words, they feel little qualm about using speeches that are in a sense ‘de-stress-timed’, hence distorted. The lack of ‘naturalness’ in the Center Listening Test may reflect the current practice of English teaching at upper-secondary schools in Japan (Yanagawa, 2016), where the speech rate of recorded materials can be slowed down and include fewer reduced forms to keep the speech closer to that of a syllable-timed rhythm. All in all, it seems that the current clarity in articulation in the Center Listening Test as well as many of the listening materials used in high schools is well supported by high-school teachers across Japan. This may be partly because, in terms of spoken word recognition, the teachers think their students can only cope with those ‘quasi-syllable-timed’ English speeches,

and that authentic listening texts are too demanding for their students (Jones, 2008). Consequently, when many of these students first have the opportunity to hear such authentic spoken texts in TOEIC tests after graduating from their high schools, they are embarrassed by the naturalness of the speech and fare far worse than they did in the Center Listening Test (Mitsubishi, 2015).

Taking the above into consideration, implications that can be drawn from the present study is that the use of authentic materials, which fully reflect the English stress-timed nature and other phonological features, should not be avoided. Otherwise, positive effects on spoken word recognition, which naturally leads to more successful comprehension, should not be expected. Japanese EFL learners should not be deprived of the opportunity not only to attune to natural stress-timed rhythm but also to brush up both top-down and bottom-up listening strategies that should be helpful in recognizing words in such authentic spoken texts. With quasi-syllable-timed listening materials prevalent in English classrooms in Japan, such strategies are less likely to be developed.

9.3.2 Limitations of the Study

There are, however, limitations to this study. First, effects of speech rate manipulation should be examined further. For example, effectiveness of longer exposure to speech delivered at different compressed speech rates on spoken word recognition should be explored.

Second, this study dealt only with Japanese EFL learners with fairly low levels of proficiency. It did not verify if the methods adopted would be useful in enhancing spoken word recognition by learners with higher levels of English proficiency.

Finally, this study did not verify if the methods proved effective in the experiments would lead to enhanced comprehension. Higher-proficiency learners, who are expected to have a rich knowledge about grammar and formulaic expressions and to have high degree of reading skills, are likely to enjoy enhanced comprehension, given the newly elevated level of spoken word recognition. However, in the case of lower-proficiency learners, an elevated level of spoken word recognition does not necessarily lead to better comprehension. They supposedly have much smaller orthographic as well as phonological lexicon, have insufficient syntactical knowledge, and lack the ability to process the incoming information in real time. Therefore, in order to enhance their listening comprehension, pedagogical methods which are effective in enhancing comprehensive linguistic competence in English should also be added, such as reinforcement of their grammar and vocabulary. It might also help to have them read the written script and tell them to comprehend in the English syntactical order while at the same time listening to its phonetic representation.

Concerning these limitations mentioned above, further studies would be necessary. Nevertheless, the present study proposed some effective methods in enhancing Japanese EFL learners' spoken word recognition that can be adopted in English classrooms. It is hoped that this study becomes a stepping stone for further study on L2 learners' spoken word recognition and listening comprehension.

References

- Adank, P., & Devlin, J. T. (2010). On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. *Neuroimage*, *49*, 1124–1132.
- Alderson, J. C. (2005). *Diagnosing foreign language proficiency: The interface between learning and assessment*. London: Continuum.
- Bard, E. G., Shillcock, R. C., & Altmann, G. T. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics*, *44*(5), 395–408.
- Berne, J. E. (1998). Examining the relationship between L2 listening research, pedagogical theory, and practice. *Foreign Language Annals*, *31*, 169–190.
- Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly*, *24*, 746–753.
- Bond, Z. S., & Garnes, S. (1980). Misperceptions of fluent speech. In R. A. Cole (Ed.) *Perception and production of fluent speech*. (pp. 115–132). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Brown, A. (1992). Twenty questions. In A. Brown (Ed.), *Approaches to pronunciation teaching* (pp. 1–17). London: Macmillan.
- Buck, G. (1992). Listening comprehension: Construct validity and trait characteristics. *Language Learning*, *42*(3), 313–357.
- Buck, G. (2001). *Assessing listening*. Cambridge: Cambridge University Press.
- Byrnes, H. (1984). The role of listening comprehension: A theoretical base. *Foreign Language Annals*, *17*, 317–329.
- Carrier, K. (1999). The social environment of second language listening:

- Does status play a role in comprehension? *The Modern Language Journal*, 83, 65–79.
- Chaudron, C. (1988). *Second language classrooms*. Cambridge University Press.
- Cook, M., & Liddicoat, A. J. (2002). The development of comprehension in interlanguage pragmatics: The case of request strategies in English. *Australian Review of Applied Linguistics*, 25, 19–39.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language*, 25(4), 385–400.
- Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech segmentation by bilinguals. *Cognitive Psychology*, 24(3), 381–410.
- Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, 14(1), 113–121.
- Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language*, 33(6), 824–844.
- Cutler, A., & Otake, T. (2002). Rhythmic categories in spoken-word recognition. *Journal of Memory and Language*, 46(2), 296–322.
- Dahan, D., & Magnuson, J. S. (2006). Spoken word recognition. *Handbook of Psycholinguistics*, 2, 249–284.
- Date, M. (2016). Does phonetic instruction for Japanese preschoolers at the beginner level help develop their phonemic categorization? *Journal for the Science of Schooling*, 17, 167–181.
- Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly

- compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology*, 23, 914–927.
- Eastman, J. K. (1993). C-words and F-words: The importance of distinguishing content and function in teaching second language listening comprehension. *System*, 21(4), 495–502.
- Field, J. (1999). Bottom-up and top-down. *ELT Journal*, 53(4), 338–339.
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57, 325–334.
- Field, J. (2008a). Emergent and divergent: A view of second language listening research. *System*, 36, 2–9.
- Field, J. (2008b). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly*, 42(3), 411–432.
- Flowerdew, J. (Ed.). (1994). *Academic listening: Research perspectives*. Cambridge University Press.
- Foulke, E. (1968). Listening comprehension as a function of word rate. *Journal of Communication*, 18(3), 198–206.
- Fujimoto, K. (2014). On-in chishiki ga nihonjin EFL gakushusha no listening ni ataeru eikyo nit suite [Effects of phonological knowledge on Japanese EFL learners' listening skills]. *JACET Kansai Journal*, 16, 90–102.
- Fujinaga, M. (2002). Why is English listening comprehension difficult for Japanese college students?: An analysis of errors in students' dictation. *Economic Theory* 306, 1–22.
- Geschwind, N. (1972). Language and the brain. *Scientific American*, 226, 76–83.
- Gilbert, J. (1995). Pronunciation practice as an aid to listening comprehension. In D. J. Mendelsohn, & J. Rubin (Eds.), *A guide for the*

- teaching of second language listening* (pp. 97–112). Carlsbad, California: Dominic Press.
- Glisan, E. W. (1988). A plan for teaching listening comprehension: Adaptation of an instructional reading model. *Foreign Language Annals*, 21(1), 9–16.
- Goh, C. (1999). How much do learners know about the factors that influence their listening comprehension? *Hong Kong Journal of Applied Linguistics*, 4(1), 17–40.
- Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System*, 28, 55–75.
- Golomb, J. D., Peelle, J. E., & Wingfield, A. (2007). Effects of stimulus variability and adult aging on adaptation to time-compressed speech. *The Journal of the Acoustical Society of America*, 121, 1701–1708.
- Goodman, K. S. (1973). The psycholinguistic nature of the reading process. In K. S. Goodman (Ed.), *The psycholinguistic nature of the reading process*. (pp. 15–26). Detroit: Wayne State University Press.
- Goto, A. (2016). Eigo gakushusha no eigo listening ni prosody joho ga oyobosu eikyo: Pitch no henka ni shoten wo atete [The effects of prosodic cues on listening activities by Japanese EFL learners: Focusing on pitch shift]. *Journal of the Chubu English Language Education Society*, 45, 15–22.
- Graham, S. (2006). Listening comprehension: The learners' perspective. *System*, 34(2), 165–182.
- Griffiths, R. T. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning*, 40, 311–336.
- Griffiths, R. T. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly*, 26, 385–390.

- Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, 28(4), 267–283.
- Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics*, 38(4), 299–310.
- Grosjean, F. & Gee, J. P. (1987). Prosodic structure and spoken word recognition. In U. H. Frauenfelder, & L. K. Tyler (Eds.), *Spoken word recognition* (pp. 135–155). Amsterdam: Elsevier Science Publishers.
- Harmer, J. (1983). *The practice of English teaching*. London: Longman Group Limited.
- Hasan, A. S. (2000). Learners' perceptions of listening comprehension problems. *Language Culture and Curriculum*, 13(2), 137–153.
- Hayashi, T. (1991). Interactive processing of words in connected speech in L1 and L2. *International Review of Applied Linguistics*, 29, 151–160.
- Herron, D. T., & Bates, E. A. (1997). Sentential and acoustic factors in the recognition of open-and closed-class words. *Journal of Memory and Language*, 37(2), 217–239.
- Ito, H. (1989). Eigo listening no konnansa no bunseki [Analysis of English listening difficulties]. *Journal of the Chubu English Language Education Society*, 19, 99–104.
- Ito, H. (1990). Comprehension gap between listening and reading among Japanese learners of English as a foreign language. *Annual Review of English Language Education in Japan*, 1, 13–27.
- Jones, D. (2008). Is there any room for listening? The necessity of teaching listening skills in ESL/EFL classrooms. *Foreign Language Teaching Forum*, 7, 15–22.
- Kadota, S. (Ed.). (2003). *Eigo no mental lexicon* [*The English mental*

- lexicon*]. Tokyo: Shohakusha.
- Kadota, S. (2012). *Shadowing ondoku to eigo shutoku no kagaku: Input kara output e* [*The science of shadowing, oral reading, and English acquisition*]. Tokyo: Cosmopier.
- Kaiser, S. (1996). Basic problems in kanji education: Are there phonogram-kanji processing differences? *Journal of Science Education in Japan*, 20, 101–102.
- Kajiura, M. (2016). Enhancing the processing speed of L2 speech: The effect of practice using faster rate (compressed) speech and transcript reading on listening proficiency. *JACET Journal*, 60, 117–135.
- Kano, N., & Saito, M. (1997). The effects of speech speed and pauses on recognition of English words and listening comprehension. *Language Laboratory*, 34, 13–31.
- Kelch, K. (1985). Modified input as an aid to comprehension. *Studies in Second Language Acquisition*, 7(1), 81–90.
- Koda, K. (2005). *Insights into second language reading*. Cambridge University Press.
- Kohno, M. (1993). Hito wa oto no nagare wo donoyounishite rikaisurunoka [How a person understands the flow of sound]. In I. Koike (Ed.), *Eigo no hearing to sono shidou* [*Teaching English listening*] (pp. 19–55). Tokyo: Taishukan Publishing.
- Krashen, S. (1982) *Principles and practice in second language acquisition*. Oxford: Pergamon Press.
- Kubozono, H. (2013). On-in ron [Phonology]. In K. Mihara, & K. Takami (Eds.), *Eigo gaku no kiso: Nichi ei taisho* [*The basics of English linguistics: A contrastive study of English and Japanese*] (pp. 1–30). Tokyo: Kuroshio Publishers.

- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: The MIT Press.
- Lieberman, P. (1963). Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech*, 6(3), 172–187.
- Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, 39(3), 155–158.
- Lund, R. J. (1990). A Taxonomy for teaching second language listening. *Foreign Language Annals*, 23, 105–115.
- Lund, R. J. (1991). A comparison of second language listening and reading comprehension. *The Modern Language Journal*, 75(2), 196–204.
- Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review*, 79, 487–509.
- McBride, K. (2011). The effect of rate of speech and distributed practice on the development of listening comprehension. *Computer Assisted Language Learning*, 24(2), 131–154.
- McDonough, K., & Trofimovich, P. (2011). *Using priming methods in second language research*. London: Routledge.
- MEXT (2009). *Kohto gakko gakushu shido yoryo* [Course of study for upper secondary school]. Kyoto: Higashiyama Shobo.
- MEXT (2013). *Gurohbaruka ni taioushita eigo kyouiku kaikaku jisshi keikaku* [Implementation plan of English educational reform to cope with globalization]. Retrieved Sep. 2, 2015, from http://www.mext.go.jp/b_menu/houdou/25/12/__icsFiles/afieldfile/2013/12/17/1342458_01_1.pdf
- Mitsubishi, M. (2015). TOEIC(R) IP test to 2015 nen daigaku nyushi center

- shiken no score hikaku bunseki chosa [Score comparison study: A descriptive analysis of TOEIC(R) IP and National Center Test for university admission in 2015]. *The JACET 54th (2015) International Convention Book*. 66.
- Monrad-Krohn, G. H. (1957). The third element of speech: prosody in the neuro-psychiatric clinic. *The British Journal of Psychiatry*, *103*(431), 326–331.
- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, *25*(3), 481–520.
- Murao, R. (2006). Gating wo ohyoshita listening nouryoku no yoin bunseki [Analyses of factors in listening by using gating paradigm]. *STEP Bulletin*, *20*, 61–76.
- Nakamura, T. (2012). The effect of prosodic boundary in understanding English sentences by Japanese EFL learners. *Second Language* *11*, 47–58.
- Narita, H. (2013). *Nihonjin ni sououshii eigo kyouiku* [*English education suitable for the Japanese*]. Tokyo: Shohakusha.
- Nishino, T. (1992). What influences success in listening comprehension? *Language Laboratory*, *29*, 37–52.
- Noro, T. (2006). Developing a construct model of “listening Stress”: A qualitative study of the affective domain of the listening process. *Annual Review of English Language Education in Japan*, *17*, 61–70.
- Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1209–1228.
- Oakeshott-Taylor, J. (1977). Information redundancy, and listening comprehension. In R. Dirven (Ed.), *Hörverständnis im*

- Fremdsprachenunterricht. [Listening comprehension in foreign language teaching.]* Kronberg/Ts.: Scriptor.
- Okazaki, M. (1993). The way of the teaching of kanji (Chinese character) on Japanese language education. *Artes Liberales* 52, 11–28.
- Oller, J. W., Jr. (1971). Dictation as a device for teaching foreign language proficiency. *ELT Journal*, 25(3), 254–259.
- Oller, J. W., Jr. (1979). *Language tests at school: A pragmatic approach*. New York: Addison-Wesley Longman Ltd.
- Oller, J. W., Jr. & Streiff, V. (1975). Dictation: A test of grammar-based expectancies. *ELT Journal*, 30, 25–36.
- O'Malley, J. M., Chamot, A. U., & Kupper, L. (1989). Listening comprehension strategies in second language acquisition. *Applied Linguistics*, 10(4), 418–437.
- Osada, N. (2001). What strategy do less proficient learners employ in listening comprehension?: A reappraisal of bottom-up and top-down processing. *Journal of Pan-Pacific Association of Applied Linguistics*, 5, 73–90.
- Osada, N. (2004). Listening comprehension research: A brief review of the past thirty years. *Dialogue*, 3(1), 53–66.
- Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language*, 32(2), 258–278.
- Peters, A. M. (1983). *The units of language acquisition*. Cambridge University Press.
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. London: Longman.
- Richards, J. C. (1983). Listening comprehension: Approach, design,

- procedure. *TESOL Quarterly*, 17(2), 219–240.
- Rivers, W. M. (1966). Listening comprehension. *The Modern Language Journal*, 50, 196–204.
- Rivers, W. M. (1971). Linguistic and psychological factors in speech perception and their implications for teaching materials. In P. Pimsleur & T. Quinn (Eds.), *The psychology of second language learning* (pp. 123–134). London: Cambridge University Press.
- Rost, M. (1991). *Listening in action*. New Jersey: Prentice Hall.
- Rost, M. (2002). *Teaching and researching: Listening*. London: Pearson Education Limited.
- Saito, Y. (2003). Investigating the construct validity of the cloze section in the Examination for the Certificate of Proficiency in English. *Spain Fellow Working Papers in Second or Foreign Language Assessment, English Language Institute, The University of Michigan*, 1, 39.
- Saussure, F. de. (1959). *Course in general linguistics*, Charles Bally and Albert Sechehaye (Eds.), Wade Baskin (Tr.), The Philosophical Library: New York.
- Shirai, Y. (2013). *Eigo wa motto kagakuteki ni gakushu shiyo* [*Scientific guide to learning English*]. Tokyo: Chukei Press.
- Takanashi, Y. (1982). Aspects of listening comprehension. *Language Laboratory*, 19, 1–12.
- Takashima, H. (1998). Accuracy of spoken word recognition as a predictor of listening comprehension for Japanese learners of English. *Annual Review of English Language Education in Japan*, 9, 87–95.
- Takei, A. (2002). *Eigo listening ron* [*Introduction to English listening*]. Tokyo: Kagensha.
- Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied*

- Linguistics*, 11(1), 90–105.
- Thompson, I. (1995). Assessment of second/foreign language listening comprehension. In D. J. Mendelsohn & J. Rubin (Eds.) *A Guide for the teaching of second language listening* (pp. 31–58). San Diego, CA: Dominie Press.
- Underwood, M. (1989). *Teaching listening*. London: Longman.
- Ur, P. (1984). *Teaching listening comprehension*. Cambridge: Cambridge University Press.
- Vandergrift, L. (2004). Listening to learn or learning to listen? *Annual Review of Applied Linguistics*, 24, 3–25.
- Vandergrift, L. (2007). Recent developments in second and foreign language listening comprehension research. *Language Teaching*, 40, 191–210.
- Vandergrift, L., & Goh, C. (2012). *Teaching and learning second language listening*. New York: Routledge.
- Vanderplank, R. (1993). ‘Pacing’ and ‘spacing’ as predictors of difficulty in speaking and understanding English. *ELT Journal*, 47, 117–125.
- Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge University Press.
- Yamaguchi, T. (1997). Nihonjin EFL gakushusha no chokairyoku to goninchisokudo nitsuite [Japanese EFL learners’ listening comprehension and word recognition rate]. *Journal of the Chugoku-Shikoku Society for the Study of Education*, 43, 104–109.
- Yamaguchi, T. (1999). Effects of training in rapid word recognition on listening comprehension: An analysis of translation-task data from Japanese EFL learners. *Annual Review of English Language Education in Japan*, 10, 83–91.
- Yamaguchi, T. (2001). Importance of word recognition in listening

- comprehension of English as a foreign language. *Bulletin of the Graduate School of Education, Hiroshima University. Part 1, Learning and Curriculum Development* 50, 17–24.
- Yamauchi, Y. (2014). Multifaceted images of Japanese EFL learners' listening anxiety attributable to instructional factors. *International Journal of Curriculum Development and Practice*, 16(1), 1–11.
- Yanagawa, K. (2016). Examining the authenticity of the center listening test: Speech rate, reduced forms, hesitation and fillers, and processing levels. *JACET Journal*, 60, 97–115.
- Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics*, 18, 49–68.

Appendices

Appendix 1

The scripts of the posttest in Experiment 2 (Words blanked out for transcription are in bold type with function words in italics)

[Dialogue]

W: You remember I'm going on my ⁽¹⁾**first** business trip next week, ⁽²⁾**right**?

M: Of course I ⁽³⁾**remember**. You've been talking about ⁽⁴⁾**it** for weeks.

W: Well, I'll ⁽⁵⁾**be** gone for two weeks. ⁽⁶⁾**You're** going to have to ⁽⁷⁾**learn** how to do-the-wash.

M: ⁽⁸⁾**Don't** worry. How difficult can ⁽⁹⁾**it** be? You put the ⁽¹⁰⁾**clothes** in and turn it ⁽¹¹⁾**on**.

W: There's more ⁽¹²⁾**to** it than that. First, ⁽¹³⁾**you** have to sort ⁽¹⁴⁾**the** clothes.

M: What?

W: Yes, you separate ⁽¹⁵⁾**the** white things and the ⁽¹⁶⁾**dark** things. You wash white ⁽¹⁷⁾**things** in hot ⁽¹⁸⁾**water**.

M: OK. That's easy.

W: Then ⁽¹⁹⁾**you** wash the dark things ⁽²⁰⁾**in** cold water.

M: Why?

W: If you ⁽²¹⁾**use** hot water, the colors ⁽²²⁾**will** change.

M: Got it. ⁽²³⁾**Is** there anything ⁽²⁴⁾**else**?

W: You can wash ⁽²⁵⁾**wool** clothes in the ⁽²⁶⁾**machine** using the gentle ⁽²⁷⁾**cycle**. But don't put ⁽²⁸⁾**wool** things in the machine ⁽²⁹⁾**that** dries clothes.

M: Why?

W: ⁽³⁰⁾**Everything** will become ⁽³¹⁾**smaller**.

M: I guess ⁽³²⁾**washing** clothes is harder ⁽³³⁾**than** I thought.

W: I'm ⁽³⁴⁾**sure** you can handle ⁽³⁵⁾**it**.

[Monologue]

On October 8, 1871, a fire started in a small building in central Chicago. Chicago is famous ⁽³⁶⁾**for** its strong ⁽³⁷⁾**wind**. That hot autumn, the ⁽³⁸⁾**wind** was very strong. There ⁽³⁹⁾**had** been no rain ⁽⁴⁰⁾**for** a long time ⁽⁴¹⁾**so**

Chicago's buildings and (42)**streets** were very dry. Most (43)**of** the buildings were made (44)**of** wood. The fire quickly (45)**grew** as the hot wind (46)**forced** it to spread to (47)**the** north. Soon it became (48)**more** and more violent, burning (49)**everything** it touched. It jumped (50)**across** the Chicago River, burning (51)**ships** and wood and coal (52)**yards** in its path. The (53)**fire** continued to burn (54)**its** way through the (55)**central** business district, where it (56)**destroyed** hotels, department stores Chicago's (57)**city** hall, the opera house, (58)**theaters** and churches, as well (59)**as** thousands of homes. The (60)**people** ran to Lake Michigan (61)**to** get away from the (62)**fire** and hot wind.

When (63)**the** fire finally ended (64)**on** October 10th, it (65)**had** destroyed an area (66)**about** six kilometers long (67)**and** one kilometer (68)**wide**. Ninety-thousand people (69)**lost** their homes. In (70)**total**, the fire killed (71)**between** two and three hundred (72)**people**. The city quickly worked (73)**to** provide food and water (74)**for** the people who had (75)**lost** their homes. Very (76)**soon**, new brick and (77)**stone** homes were (78)**built**. The downtown area (79)**grew** larger as businesses began (80)**building** tall buildings. The city (81)**quickly** grew and soon (82)**became** a center for (83)**both** business and industry.

Examinations (84)**after** the fire could not (85)**discover** exactly how the fire (86)**began**. There was a newspaper story (87)**that** Mrs. Catherine O'Leary was (88)**in** the small building (89)**where** the fire started. The (90)**story** said she put a (91)**lamp** on the floor and (92)**her** cow kicked it over, (93)**starting** the fire. In a (94)**children's** song, the cow said, "(95)**There'll** be a hot time (96)**in** the old town tonight." (97)**Years** later, the reporter who (98)**wrote** the story said that (99)**he** made it up to (100)**make** the story more exciting.

Appendix 2

An example of the scripts and their Japanese translation used in the dictation practices during the treatment of Experiment 2 (Words targeted for transcription are in bold type and in italics)

[Scripts]

Hello and welcome to your first day *in* the accounts *department*. I'm Martha Stone and I'm the *head of* this department. Because you are all new workers, we will *start you off with* checking numbers. We are responsible *for making* sure all the numbers *entered* into the computer system *agree* with the *figures reported* by our branch offices all over the country. These blue forms are *order forms that* come from the branch offices. *These* yellow forms are office *copies*. It's very important that all the numbers are the *same*. You will work *in pairs*. Please check these forms and then check them again, and give them to the *other* person *in* your pair *who will* check them a *third* time. When *you've* finished checking them, please *pass them to* Julie Sanders. She's *standing* over there. If you have any questions, *come and ask either* Julie or me. Don't try to *guess*! I know this work is *not* very exciting but it is very important *for* the company. *Once* you have *learned* how to do it, we will give you more interesting things *to* do. Always remember that every *job* is important, and you have to start *at the* beginning.

[Japanese Translation]

こんにちは、そして会計課で初日を迎えた皆さん、ようこそ。私はマーサ・ストーン、この課の責任者です。皆さんは新入社員なので、数字の確認から始めてもらいます。コンピューター・システムに入力されたすべての数字が、全国の支店が報告した数字と一致しているかどうかを確認するのが、私たちの仕事です。この青い用紙が支店から来る注文書です。この黄色い用紙は控えです。全ての数字が同じであることが非常に重要です。皆さんには2人1組で作業してもらいます。これらの用紙をチェックして、それからもう一度チェックし、ペアの相手に渡してその人が3度目のチェックをします。チェックが終わったら、それをジュリー・サンダースに渡してください。彼女はあそこに立ってい

ます。質問があったら、ジュリーか私に聞いてください。推測しようとしては
いけません。この仕事は、大して面白いものではないことは承知していますが、
会社にとってはとても重要です。この仕事の進め方を覚えたら、もっと面白い
仕事を与えます。どんな仕事も大切であって、一から始める必要があるという
ことを常に心しておいてください。

Appendix 3

The scripts of the posttest in Experiment 3 (Sections targeted for transcription are in bold type with function words in italics)

[Dialogue]

A: Hello. This is Ticket Central. How can I help you?

B: I'd like two tickets for the music show at 7 o'clock tomorrow night at Sunshine Theater.

A: I'm sorry. There ⁽¹⁾***are no more tickets*** for that show until next month.

B: Next month! But I'm leaving next week!

A: Well, if you like that kind of music, there is the same kind of show at the Roxy Theater. The musicians are famous and their music is wonderful. In fact, I'm going to that show tomorrow. We still have tickets for the show on Thursday, the day after tomorrow.

B: That sounds interesting. Thank you very much for ⁽²⁾***telling me about it***. What time does the show start?

A: It begins at 8 o'clock.

B: And where is the Roxy Theater?

A: It's near Grand station on Second Street.

B: Oh! That's very near my hotel.

A: There's a very good French restaurant beside the Roxy Theater if you want to have dinner before the show. It's called 'Le Petit Chateau.'

B: You are ⁽³⁾***giving me great help!*** Thank you.

A: I'm glad I can help because we like the same kind of music. I really think ⁽⁴⁾***you'll enjoy the show***.

B: How much are the tickets?

A: ⁽⁵⁾***For the best seats***, they're \$30 dollars each. There are cheaper seats for \$25 and \$20.

B: I like to be ⁽⁶⁾***close to the musicians***, so I'll take two \$30 dollar tickets.

A: Great. Can I have your name, please?

B: I'm Keiko Tanaka. That's K-E-I-K-O T-A-N-A-K-A.

A: Thank you. You can pick up the tickets and pay for them at the desk at the Roxy Theater. The desk opens at 5 o'clock and you should get your

tickets any time before 7:30, because ⁽⁷⁾*if you are late*, they may sell ⁽⁸⁾*them to someone else*.

B: I understand. Thank you so much!

A: Enjoy the show.

[Monologue]

Good morning, everyone. I hope you all had a nice weekend. Before we begin today's meeting, I have several things to tell you. ⁽⁹⁾*As you all know*, the machines we ⁽¹⁰⁾*ordered from ABC Company* did not arrive on time. I called them four times, and they had no good ⁽¹¹⁾*excuse for arriving late*. This now means our ⁽¹²⁾*business plan is off* so I don't think that we should use ABC Company any more in the future.

The second thing I want to talk about is Susan Wesley. She has moved to a new job in Arizona. She worked very well, and we ⁽¹³⁾*will all miss her*. Marc Connolly is the person who will be ⁽¹⁴⁾*taking over her job*. Marc worked in our Washington office for 12 years before he came here. Please make him welcome and help him learn about what we do here.

Third, we will have a meeting with the people ⁽¹⁵⁾*from our parent company* on Thursday, September 27. Please read these important papers before the meeting. There will be many questions, and we must answer them.

The last thing I want to say is good news. The company picnic will be held this Saturday at Greenway Park. We will have many good things ⁽¹⁶⁾*to eat and drink*. There will be games for your children to play. We will also have a baseball game between our team and the team from Rockland Company. The company picnic is always great fun, so please bring your families to the park around 12 o'clock on Saturday.

Thank you for listening to me. Now let's begin the meeting.

Appendix 4

The handouts given to the experimental groups as the treatment in Experiment 3

Grammar Phrase Sheet	
• あなたと話したかったんです。 I wanted to talk to you.	• 宿題をするのに忙しい。 I'm busy doing my homework.
• 彼女の叔母は入院中です。 Her aunt is now in the hospital.	• 駅へ迎えに行くよ。 I'll pick you up at the station.
• 彼は突然気分が悪くなった。 He suddenly got very sick.	• ジャあまたね。 Talk to you soon.
• 彼女の姉は病院で勤務している。 His sister works at the hospital.	• グランドキャニオンはアメリカでもっともすばらしい自然の景観の一つだ。 The Grand Canyon is one of the most wonderful natural places in the US.
• それはお気の毒に (それは残念ですね)。 That's too bad.	• その川は長さ 300 キロだ。 The river is 300 kilometers long.
• 彼が大丈夫だといひんですが。 I hope he will be all right.	• その湖の最も狭い部分は幅 300 メートルだ。 The narrowest part of the lake is 300 meters across.
• 昨夜、私の夫は私に対してひどかったのよ。 Last night my husband was mean to me.	• その岩は太陽に照らされるととても美しく見える。 The rock looks very beautiful under the sun.
• 今朝、夕食の後で食べるための素敵なケーキを作ったのよ。 This morning I made a nice cake to eat after dinner.	• 何千年もの間、ネイティブアメリカンはここに住んでいた。 For thousands of years, Native Americans lived here.
• これが今実行すべき計画です。 This is the plan to carry out now.	• 彼らは籠を作るのがうまい。 They are good at making baskets.
• それを作るのにとっても時間がかかったのよ。 It took me a long time to make it.	• その町の周りや中には多くの建物がある。 There are a lot of buildings in and around the city.
• ケーキはおいしくなかった。 The cake didn't taste good.	• これが川へと下っていく道です。 This is the way to get down to the river.
• 彼の言葉は私の気持ちを傷つけた。 His word hurt my feelings.	• 結局彼らはあきらめた。 In the end they gave up.
• とにかくそれを売りたいんだ。 I just want to sell it.	• 楽しみでスーパーへ行く。 I go to the supermarket for fun.
• 君の気持ちはわかるよ。 I know how you feel.	• 川を渡るボートがある。 There are boats to go across the river.
• 君の弟はときどき私を怒らせる。 Your brother sometimes makes me angry.	• キャンプ用のテントがいくつかある。 There are several tents for camping.
• 彼女をここへ招待するたびに、彼女は来れない言い訳をするのよ。 Every time I invite her to come here, she finds an excuse not to come.	• 川下りのボートツアーもある。 There are boat trips down the river.
• 君はタバコをやめるべきだ。 You should stop smoking.	• その飛行機は峡谷の上を飛ぶ。 The plane flies over the canyon.
• 彼女にパーティーに来るように頼んだ。 I asked her to come to the party.	• 訪問客は車でそこまで行ける。 Visitors can get there by car.
• 彼女が黒髪の高くて黒い男と一緒にいるのを見たわ。 I saw her with a tall man with black hair.	• その場所をよく知っているガイドを雇えるよ。 You can hire guides who know the place very well.
• 彼女は毎週末違う男とデートしている。 She goes out with a different man every weekend.	• その場所に行ったことのある者は誰もいない。 None of them have ever visited the place.
• 僕なら彼女とは結婚しないな。 If I were you, I would not marry her.	• それがどれほど大きく美しいか理解することは難しい。 It is very difficult to understand how big and beautiful it is.
• 賢明な男性なら彼女とは結婚しないでしょう。 A wise man would not marry her.	• 御用をお伺いしましょうか (いらっしやいませ)。 How can I help you?
• 彼女は仮に結婚したとしても、家にずっといることはないだろう。 She wouldn't stay home even if she did marry.	• 明日の 7 時からのミュージカルのチケットを 2 枚欲しいのですが。 I'd like two tickets for the music show at 7 o'clock tomorrow.
• 彼女は自由時間をすべて買い物に使っている。 She spends all her free time shopping.	

<ul style="list-style-type: none"> • そのミュージカルの空席はもうありません。 There are no more vacant seats available for the musical. • 来月出発します。 I'm leaving next month. • これは君が先週見たのと同じ種類のショーだ。 This is the same kind of show you saw last week. • 実は、来週彼のところに行く予定です。 In fact I'm going to visit him next week. • 明後日の金曜日のショーのチケットなら何枚かあります。 There are some tickets left for the show on Friday, the day after tomorrow. • おもしろそうですね。 That sounds interesting. • 知らせてくれてありがとう。 Thank you very much for letting me know about it. • 貴重なアドバイスをありがとう。 You are giving me great advice. • お役にたててうれしいです。 I'm glad I can help. • 音楽を楽しんでいらしてください。 I hope you'll enjoy the music. • いい席は1枚30ユーロです。 For good seats, the tickets are 30 euros each. • そのコンビニは駅に近い。 The convenience store is close to the station. • 一番いい席のチケットを2枚ください。 I'll take two tickets for the best seats. • レジでチケットを受け取れます。 You can pick up the tickets at the check-out counter. • そこで支払ってください。 Please pay for them there. • 遅刻すると、彼らは怒るよ。 If you are late, they'll be angry. • このことは他の誰にも話さないでください。 Don't give this information to anyone else. • きっと誰かがあなたの前に持っていったに違いないわ。 Someone else must've taken it before you. • 皆さん週末はいかがでしたか。 I hope you all had a nice weekend. • 会議の前にお知らせがあります。 I have some things to tell you before the meeting. • ご存知の通り、来週私はフランスに向けて出発します。 As you know, I'm leaving for France next week. • あなた方はみなさん、これに加入しなければなりません。 You all have to sign up here. • これが、我々が紀伊国屋に注文した本です。 These are the books we ordered from Kinokuniya. • 注文したコピー機は遅れずに到着した。 The copying machines we ordered arrived on time. • 遅刻のもっともな理由はないね。 You have no good excuse for being late. • 我々の将来計画はやりなおしです。 Our plan for the future is now off. 	<ul style="list-style-type: none"> • これからは彼の言うことは聞かなくてよいでしょう。 I don't think we should listen to him any more in the future. • 2点目は、彼女の異動についてです。 The second thing I want to talk about is her move to a new job. • 我々はみんな、アメリカにいる彼がなつかしい（いなくて寂しい）。 We all miss him in the US. • 彼がこの仕事の責任者です。 This guy is the person who is responsible for the job. • あなたの同僚の一人があなたの仕事を引き継ぎます。 One of your colleagues will be taking over your job. • 彼はここへ来る前はABC会社で働いていました。 He worked for ABC company before he came here. • 彼が新しい仕事を覚えるのを手伝ってあげてください。 Please help him learn about the new job. • ここでやっていることはそんなに難しくありません。 What we do here is not very difficult. • 親会社の重役たちと、会議があります。 We will have a meeting with the executives from our parent company. • 彼らからたくさん質問が出るでしょうし、我々はそれに答えなければなりません。 There may be many questions from them and we must answer them. • 最後の1点はよい知らせです。 The last thing I want to say is good news. • パーティーは来週日曜にあります。 The party will be held next Sunday. • 何か食べ物を持ってきてください。 Please bring in something to eat. • 何か冷たいものが飲みたい。 I want something cold to drink. • 飲んだり食べたりする時間です。 It's time to eat and drink. • ゲームもたくさんあります。 There are a lot of games for you to play. • これは内緒ですよ。 This is just between you and me. • そのパーティーは楽しいでしょう。 The party will be great fun. • ピクニックは楽しかった。 The picnic was a lot of fun. • 日曜日、公園に昼ごろ来てください。 Please come to the park around noon on Sunday.
--	---

Appendix 5

The scripts of the posttest in Experiment 4 (Words blanked out for transcription are in bold type with function words in italics)

[Dialogue]

1. Wow! There are so many people ⁽¹⁾*here*, Maiko. / Yeah, this is one of the most ⁽²⁾**popular** beaches in Japan. / I've never seen ⁽³⁾*such* a crowded beach before. I ⁽⁴⁾*should* take a picture to show to my family in Canada. / I'll take a picture with you ⁽⁵⁾*in* it. Stand over there.
2. Hello? / Hi, Julia. It's Dan. I'm calling to ask you about our ⁽⁶⁾**math** homework for tomorrow. / Oh. Can I call you back ⁽⁷⁾*in* an hour? / Sure, no problem. Are you ⁽⁸⁾**busy**? / Yeah. I have to finish ⁽⁹⁾**making** dinner. My parents ⁽¹⁰⁾*aren't* home tonight.
3. Excuse me. ⁽¹¹⁾*Where* is the cafeteria at this university? / There are two of them. The one ⁽¹²⁾*in* Smith Hall is better. / Great. Where is that? / Do you see ⁽¹³⁾*that* tall building over there? That's it. You'll ⁽¹⁴⁾**see** the cafeteria as soon as you enter it.
4. Did you sleep ⁽¹⁵⁾**well** last night, Pam? / Yeah, Dad. I dreamed I was a ⁽¹⁶⁾**famous** singer. I was ⁽¹⁷⁾**giving** a concert. / Well, ⁽¹⁸⁾*you're* good at singing. Maybe your dream will come ⁽¹⁹⁾**true**. / That would be nice. But I'll need to take singing lessons ⁽²⁰⁾**first**.
5. Clarksville Public Library. / Hi. I think I ⁽²¹⁾**left** my coat on a chair when I was ⁽²²⁾*there* this morning. / What floor were you ⁽²³⁾**on**, sir? / The ⁽²⁴⁾**third** floor. I was sitting ⁽²⁵⁾*near* the newspaper section.

[Monologue]

6. Dr. Tanaka is a ⁽²⁶⁾**dentist** in Tokyo. Recently, many foreign ⁽²⁷⁾**patients** have been coming to his clinic. ⁽²⁸⁾**Some** of them can't speak Japanese, and Dr. Tanaka wants to communicate with them in English. As a ⁽²⁹⁾**result**, he has started taking English lessons. Dr. Tanaka can't speak very well ⁽³⁰⁾*yet*, but he is studying hard.
7. Nicole bought a ⁽³¹⁾**used** car last month. ⁽³²⁾*While* she was driving it yesterday, she noticed something strange. Whenever she ⁽³³⁾**turned** a

corner, the car made a loud noise. But when she drove ⁽³⁴⁾**straight**, the car sounded normal. Nicole ⁽³⁵⁾**plans** to take it to a repair shop this weekend.

8. Tim's family moved to a ⁽³⁶⁾**different** town this week. Tim will ⁽³⁷⁾**start** going to his new high school next Monday. Before then, he has to do some work ⁽³⁸⁾**that** his new teachers gave him to help him catch ⁽³⁹⁾**up** with his classmates. Tim is ⁽⁴⁰⁾**looking** forward to making new friends at the school.
9. Attention, ladies and gentlemen. The Springfield Summer Parade will begin shortly. I'd like to ⁽⁴¹⁾**ask** everyone standing in the street ⁽⁴²⁾**to** take a few steps back. Please move onto the sidewalk so ⁽⁴³⁾**that** the people marching in the parade can get ⁽⁴⁴⁾**through**. Thank you for your cooperation.
10. Today, many people ⁽⁴⁵⁾**around** the world enjoy skateboarding. Skateboards became ⁽⁴⁶⁾**popular** in California in the 1960s. At the time, there were many people in southern California ⁽⁴⁷⁾**who** went surfing. But ⁽⁴⁸⁾**when** the waves at the beach were not big enough for surfing, surfers looked ⁽⁴⁹⁾**for** activities to do on land. They started to buy skateboards and ride ⁽⁵⁰⁾**them** on sidewalks and roads.

Appendix 6

An excerpt from the textbook used by the participants in Experiment 4

Lesson 2 文型と動詞／興味・関心

Are you interested in Japanese anime? 日本のアニメに興味がある？

Model Conversation [K: Kaito / J: John]

K1: John, are you interested in Japanese anime?

J1: Yes, in fact, anime **gave** me the motivation to study Japanese. **In other words**, it brought me to Japan.

K2: Oh really? I didn't know you were such an anime fan. What is your favorite one?

J2: It is definitely *Ponyo*.

K3: Yes, Miyazaki Hayao's movies **are** fantastic! The endings of his movies always **make** us happy.

J3: Yeah, I **agree**.

K4: **By the way**, do you **like** Japanese comics? I'll **lend** you some if you like. You can learn Japanese through manga.

J4: That's a great idea!

K1: ジョン、日本のアニメに興味がある？

J1: うん、実はアニメのおかげで日本語を勉強する気になったんだ。言い換えれば、アニメが僕を日本に連れてきたってこと。

K2: へえ本当に？君がそんなにアニメ好きとは知らなかったよ。君の好きなアニメは何？

J2: 間違いなく、『ポニョ』だよ。

K3: なるほど。宮崎駿監督の映画はすばらしいよね！最後にはいつも僕たちを幸せな気分にしてくれる。

J3: 本当にそのとおり。

K4: ところで、君は日本のコミックは好き？もしよければいくつか貸してあげるよ。漫画を通して日本語を学ぶことができるよ。

J4: それはすごくいい考えだね！

Listening Task 1. T / F 2. T / F 3. T / F

Pronunciation リズム

英語を聞いて、強く発音されている単語を()に書かれた数だけ○で囲みなさい。

1. He takes a walk in the morning. (3)
2. I called him and invited him to dinner. (3)
3. There are some comics for children in the library. (3)

Function 言い換える／話題を変える

1. She's a vet, **I mean**, an animal doctor.

彼女は獣医、つまり動物のお医者さんです。

2. They said no. **In other words**, they turned down our offer.

彼らはいいえと言った。言い換えれば、彼らは私たちの申し出を断ったということだ。

3. **By the way**, how did you get here?

ところで、ここまでどうやって来たの？

Grammar 1 文型と動詞①

A 主語と動詞／目的語／補語

1. Ms. Ito teaches English. 伊藤先生は 英語を 教えている。

主語(S) 動詞(V) 目的語(O) S O V

2. My sister is a college student. 私の姉は 大学生 です。

S V 補語(C) S C V

主語(S), 動詞(V), 目的語(O), 補語(C)以外の語句は修飾語。

Ms. Ito teaches English at school.
↑
修飾語

S: 「～は」「～が」
V: 「～する」「～である」
O: 動作の対象となる。「～を」「～に」
C: 主語・目的語の情報を補う語。

B SV (第1文型)

3. Kate lives in New York. ケイトはニューヨークに住んでいる。

S V

4. There is a tree in the yard. 庭に1本の木がある。

V S

3. 動詞(V)の後に目的語(O)も補語(C)もこない。in New York は修飾語。

4. 否定文は be 動詞の後に not を置く。There isn't a tree in the yard.

疑問文は (be 動詞+there+主語?) の語順。Is there a tree in the yard?

注意 〈There+be 動詞〉の be 動詞は、後ろにくる名詞が単数なら is / was、複数なら are / were となる。

主語+動詞

There+be 動詞+主語
「～がある[いる]」

C SVC (第2文型)

5. He is famous. 彼は有名だ。

S V C

6. He became a doctor. 彼は医者になった。

S V C

補語(C)は主語(S)を説明する。

S=C の関係が成り立つ。

(6. は He is a doctor. という関係)

SVC で使われる動詞

be 動詞(～である) keep(ずっと～である)
remain(～のままである) become, get(～になる)
grow([次第に]～になる) look(～に見える)
seem(～に思われる) sound(～に聞こえる)
smell(～においがする) taste(～の味がする)など

主語+動詞+補語

D SVO (第3文型)

7. He bought a new watch yesterday.

S V O

彼は昨日、新しい腕時計を買った。

動詞(V)の後に目的語(O)がくる。yesterday は修飾語。

注意 SVC と SVO の違い

SVC He became a doctor. S=C (He is a doctor.)

SVO He bought a new watch yesterday. S≠O (× He is a new watch.)

主語+動詞+目的語

Knowledge is power. (知識は力なり。)

Exercises 1

1 各文の下線部の語句は S, V, O, C のどれにあたるか、下線部の下に書き入れなさい。A

1. We play badminton during the lunch break.
2. After meals my dog sleeps for three or four hours.
3. The view from the room is beautiful.

2 [] 内から動詞を選び、英文を完成させなさい。B C D

1. There _____ five rooms in my house.
2. The bells _____ at 7 every morning.
3. My aunt _____ a journalist.
4. Do you _____ his phone number?
5. _____ your textbook to page 5.
[open / ring / is / are / remember]

3 () 内の語句を並べかえて英文を完成させ、文型が SV か SVC か SVO かを答えなさい。B C D

1. (a / was / pet shop / there) here before. _____ []
2. (my / broke / cup / I / favorite). _____ []
3. (in / a teacher / became / she) 2001. _____ []
4. (sat / in / he / the chair / down). _____ []
5. (looks / for / Meg / young) her age. _____ []

4 日本語に合うように、() 内の語句を並べかえて英文を完成させなさい。総合

1. パーティーにはたくさん人がいましたか。
(at / many / there / the party / were / people)?

2. 彼は服を着替え、部屋を出た。
(left / and / his clothes / changed / he / the room).

3. そのボートは4人が乗るには小さそうだ。
(people / small / seems / the boat / four / for).

Use it あなたが普段の生活で行うことを、3つ書いてみよう。

(例) I usually get up at 6:45. I go to school by bike with my friends. I take a bath after dinner.

Expressing 興味・関心

STEP 1 Emily と Kaito の会話を聞いて、下記のリストから語句を選び、吹き出しの中の下線部に補いなさい。

Emily: I'm interested in 1_____. My favorite 2_____ is 3_____. I like 4_____ very much. I was impressed when I 5_____.

Kaito: I'm interested in 1_____. My favorite 2_____ is 3_____. I like 4_____ very much. I was excited when I 5_____.

1	2	3	4	5
sports	sport	soccer figure skating	Messi Asada Mao	saw his/her last performance
music	musician	Michael Jackson Beyoncé	"Black or White" "Single Ladies"	heard this song
movies	actor	Brad Pitt Audrey Hepburn	<i>Seven</i> <i>Roman Holiday</i>	saw this movie
reading	writer	Murakami Haruki J.K. Rowling	<i>Norwegian Wood</i> the <i>Harry Potter series</i>	read this work

STEP 2 STEP 1 のリストを参考にして、あなたが興味・関心のある人や物について、英語でメモを完成させなさい。

- What are you interested in? _____
- What [Who] is your favorite _____? _____
- Write more information about it [him/her].

STEP 3 STEP 2 のメモを参考にして、あなたが興味・関心のあることについてクラスで発表する文章を書きなさい。

Hello, everyone. I'll tell you about (a _____). My favorite (b _____) is (b _____). (c _____).

Thank you for listening.

Appendix 7

The scripts of the cloze and the paused transcription posttest in Experiment 5 (Words blanked out and sections targeted for transcription are in bold type with function words in italics)

[Cloze Posttest]

W: I've had my eye on this pink ⁽¹⁾**shirt** for a long time. What do you think?

M: It looks nice ⁽²⁾*on* you.

W: Do you like these ⁽³⁾**green** pants?

M: Not really. To be honest, they make you ⁽⁴⁾**look** fat.

W: Oh. That's not good. I'll ⁽⁵⁾**try** a different pair. You should try that ⁽⁶⁾**blue** shirt. I think it ⁽⁷⁾*would* look good on you.

M: I don't ⁽⁸⁾**really** like the color. Blue looks good on you, ⁽⁹⁾*not* on me. I like green.

W: OK. Try this green shirt.

M: I don't really want ⁽¹⁰⁾*any* more shirts.

W: Well, you need a ⁽¹¹⁾**new** business suit.

M: The suits I ⁽¹²⁾**have** are fine.

W: If you ⁽¹³⁾*don't* want to buy anything, ⁽¹⁴⁾*why* did you come to the store with me?

M: Because you ⁽¹⁵⁾**asked** me to come! It wasn't my ⁽¹⁶⁾**idea**! I'll go wait for you ⁽¹⁷⁾*in* that coffee shop.

W: Are you sure?

M: Take ⁽¹⁸⁾*your* time. I'm not in ⁽¹⁹⁾**a** hurry. I'll go buy a book to read ⁽²⁰⁾*while* I wait.

W: OK. See you later.

[Paused Transcription Posttest]

One day, Lucy comes ⁽¹⁾*to visit* **her aunt**, Mrs. Jennings. Lucy doesn't ⁽²⁾**have much money, either**. She lives in Plymouth with her uncle ⁽³⁾**who runs a school**. Lucy tells Elinor that Edward went to her uncle's school. When Edward ⁽⁴⁾**was a student there**, he and Lucy fell in love. Lucy tells

Elinor a secret: she and Edward ⁽⁵⁾**are going to marry**. Elinor is very sad but she promises to keep Lucy's secret.

Soon, Mrs. Jennings, Lucy, Elinor and Marianne go to London. They all go to a big dance party ⁽⁶⁾**in a beautiful house**. Marianne sees Willoughby but he doesn't want to talk to her. Marianne discovers that he ⁽⁷⁾**is with another woman** and he is planning to marry her!

Lucy is not very clever. She tells Fanny that she is going to marry Edward. Fanny and Edward's mother both become very angry. Mrs. Ferrars takes away all of Edward's money and gives it to his younger brother Robert.

Marianne is so sad about losing Willoughby that she gets very sick. The doctor thinks she is going to die. Elinor is very sad. ⁽⁸⁾**She cries and says**, "Please ⁽⁹⁾**don't leave me alone**." But Marianne starts getting better. Soon she and Colonel Brandon ⁽¹⁰⁾**start spending time together**.

Appendix 8

Examples of the materials used in the treatment of Experiment 5

[Material Examples Featuring Phonemes, Syllables, and Rhythms]

No.1 l/rを素早く切り替える!

Target Sentence

Track 40

Larry met Clara in Honolulu. They have a really happy relationship. They plan to get married and move to rural Florida next year.

(ラリーはクララにホノルルで会いました。彼らは実に幸せな関係を築いています。彼らは来年には、結婚してフロリダの田舎に引っ越すつもりです)

Part 1 で l と r の発音を確認しましたが、場合によっては、1 つの単語の中で、あるいは 2 語以上がひと固まりで発音される場合に、l と r が母音を挟んで隣接する場合があります。l と r の口の動きを連続して行うため、慣れないと滑らかに言えず、l と r を互いに入れ替えてしまうこともあります。また、数はそれほど多くありませんが、1 つの単語の中で、2 つ以上の l や r が母音を挟んで連続する場合があります。例えば、temporary (一時の) であれば、temporal_y と言ってしまう (実際には日本語のラ行になる) 間違いがよく見られます。

いずれにしても、l と r の発音ができていれば、練習次第で克服できます。どんな組み合わせでも正しく発音できるように、l/r と日本語の「あいうえお」を組み合わせた練習をしてみましょう。

「ら り り る る れ れ ろ ろ」

上記の子音の部分を、英語の l や r に置き換えて発音します。「r あ l あ r い l い……」といった具合です。r/l、l/r、r/r、l/l の、すべての組み合わせで言ってみましょう (Track 41)。慣れてきたら、スピードを上げて練習してみてください。その後、以下の練習をしましょう。

Exercises

literary	literally	celery	parallel
lullaby	Maryland	necessarily	lyrics
merrily	temporary	rolling	gallery
lily	particularly	primarily	rewrite

◆ literary 文学の ◆ literally 文字どおりに ◆ celery セロリ ◆ parallel 平行な ◆ lullaby 子守歌
◆ lyrics 歌詞 ◆ merrily 楽しく、陽気に ◆ primarily 第一に ◆ rewrite 書き直す

1 Track 42

上記の単語一覧を見ながら CD を聞き、ポーズの部分で、丁寧にリピートしましょう。

2 Track 42

単語一覧を見ずに CD を聞き、耳を頼りに、ポーズの部分でリピートしましょう。

3 Track 43

単語一覧を見ながら CD を聞き、リズムに合わせて、ナレーションと同じタイミングで発音しましょう。

[注意] テンポについていくのが精いっぱい、発音があいまいになってしまう場合は、CD を使わず、確実に発音できるテンポでゆっくりと練習しましょう。

Review Target Sentence を、CD を参考に、滑らかに発音できるよう、何度も音読しましょう。 Track 40

No.2 thをサ行で読まない

Target Sentence

Track 30

I used to think Timothy and Thelma were brother and sister.

(ティモシーとセルマは兄妹だと思っていました)

[s]の音は、日本語にも存在するので問題ありませんが、thのスペリングで表される [θ] の音は、日本語にはありません。中級者以上であれば、[θ] の発音は練習した経験があるはずですし、意識しているときはそんなに間違えることもないでしょう。ただ、語中や語末、あるいは算用数字に隠れてしまう (thousand など) 場合に、意識が低くなって、[θ] を [s] に置き換えて発音してしまうことがあります。

まず、[θ] の音の作り方を確認します。「トゥ、トゥ、トゥ……」と言いながら、舌先が歯茎に当たる感触を確かめます。次に、舌先が当たる部分を徐々に前に移動して、歯茎ではなく前歯の裏または先に当ててください。その状態で息を出します。これが「澄んだ (無声の) th」 [θ] の音です。舌を上前歯の裏に軽く当て、その間

から息を出すのがポイントです。それほど響く音ではないので、無理に音を出そうとせず、舌先を歯に当てることを意識しましょう。

比較的簡単にはずですが、必要なときにいつでもすぐに出せるようになるには、慣れが必要です。また、[s] の音とはっきり区別ができるよう、ここで練習しておきましょう。

なお、[θ] と同じ口の形で「うー」と声を出せば、「濁った (有声の) th」 [ð] の音が出ます。濁った th は、声の振動が舌先や唇に伝わって、くすぐったい感じがします。[θ] を [s] と混同するように、[ð] は [z] の音と混同することがあります。ここでは練習しませんが、こちらも注意が必要です。

Exercises

thick

sick

thaw

saw

sink

think

faith

face

sought

thought

sing

thing

thank

sank

mouse

mouth

◆ thaw 溶ける、打ち解ける ◆ sought ★ seek (探す、求める) の過去分詞形。

1 Track 31

上の単語一覧を見ながら CD を聞き、ポーズの部分で、丁寧にリピートしましょう。

2 Track 31

単語一覧を見ずに CD を聞き、耳を頼りに、ポーズの部分でリピートしましょう。

3 Track 32

単語一覧を見ながら CD を聞き、リズムに合わせて、ナレーションと同じタイミングで発音しましょう。

【注意】テンポについていくのが精いっぱい、発音があいまいになってしまう場合は、CD を使わず、確実に発音できるテンポでゆっくりと練習しましょう。

Review

Target Sentence を、CD を参考に、滑らかに発音できるよう、何度も音読しましょう。

Track 30

1. First of all, I've two or three things I'd like to say.

(まず、2つ3つ、お伝えしたいことがあります)

発音のコツ

First of all は、ひと息で発音します。

「フアー・スト・ヴォール」のように聞こえたら OK です。

2. I'll let you know my views before Thursday.

(木曜までに、私の見解をあなたにお伝えします)

発音のコツ

I'll let の語尾と語頭の [l] はつなげる感じで、息を舌の両側に流しながら発音します。

know の二重母音 [ou] はしっかりと、view の [v] と before の [f] は息を長め・強めに発音します。

3. I'd like to thank everyone for coming at such short notice.

(急な連絡にもかかわらず、お集まりくださった皆さまに、感謝いたします)

発音のコツ

I'd like は、d と l をつなげる感じで発音します。thank everyone for は th、v、r、f の子音を意識して発音すると、歯切れよく聞こえます。notice の [ou] はアとオの中間の音です。

4. Please help yourself to the food. Is there anything you don't eat?

(お食事はご自由にどうぞ。何か苦手なものはありますか?)

発音のコツ

please の pl は子音の連続なので、p の後、舌を上につけるだけで自然と [l] は発音されます。

help や yourself の [l] は弱く、また、オに近い音です。

5. Finally, before you go, I'd just like to say, you're all very welcome to visit us again.

(最後に、お帰りになる前に、ひとつお伝えておきたいのですが。ぜひ、皆さんでまたいらしてください)

発音のコツ

コンマでは基本的に、ひと呼吸置きます。you're all very welcome はひと息で発音するのがコツ。

very の [v] を長め・強めに言うと、より感謝の気持ちが込められます。

ラ行音を持つ部分(下線部)に
[l]の音を当てて読んでください。
その後[r]の音でも
読んでみましょう。

A: おひる何食べる?
カレーライス?
それともラーメン?

B: 激からラーメンか、
レバニライスがいいなー。

A: それにサラダも! きゆうりと
レタスのつてるのがいい!

B: あ、カラアゲとエビチリも
食べる!

次の英文の下線部には、
[l] [r]の音が含まれています。
注意して読んでください。

1. The leader is
the wrong person
to read the letter.
2. The light trucks
in the right lane are
in a long row.

[訳] 1. その指導者は、その手紙を読むべき人ではない。 / 2. 右車線の軽トラックは、長い列の中にいた。

音声を聞いて、強く読まれている語
に○を付けてください(正解は p.
51 の左上)。

1. He traveled to Paris
and Berlin on business.
2. Keith should have
been reading some
of the documents.

[訳] 1. 彼は、出張でパリとベルリンに行った。
 / 2. キースは、書類のいくつかを読んでいた
はずだ。

音声を聞いて、それぞれの文で、最
も目立つように読まれている語に○
を付けてください。

1. Leonard received a card.
2. Leonard received
a long black pencil.
3. Leonard received
a couple of
contemporary novels.

[訳] 1. レナードは、カードを受け取った。 / 2.
レナードは、長い黒鉛筆を受け取った。 / 3.
レナードは、現代小説を2、3冊受け取った。

She went to the station



to meet her father.



[An Example of the Dialogues for Perception and Articulation Practices]

W: Good morning. Did you have a bad night last night?

M: Why are you asking me that?

W: Oh, it's nothing.

M: You're laughing at me! What's so funny?

W: Nothing's funny.

M: You're still laughing! Do I have egg on my face or something?

W: Well, your hair looks really bad today. Was it wet when you went to bed?

M: Yes, it was. I had to work late last night. When I got home, I took a shower and went straight to bed. I was really tired. Does my hair really look bad?

W: Go to the bathroom and look in the mirror. You'd better do something or everyone will laugh at you.

M: You don't happen to have a hat I could wear, do you?

W: おはよう。夕べはひどかったの？

M: どうしてそんなことを聞くの？

W: あら、なんでもないわ。

M: 僕を見て笑っているじゃないか！何がそんなにおかしいの？

W: 何もおかしくないわよ。

M: まだ笑ってるじゃないか！僕の顔に卵か何かついてる？

W: あのね、髪型が今日はすごく変なのよ。寝るとき、まだ濡れてたの？

M: ああ、そうだったよ。夕べは遅くまで残業しなきゃいけなかったから。家に帰ってシャワーを浴びて、すぐに寝てしまったんだ。とても疲れていたから。髪型、本当に変？

W: 洗面所に行って鏡で見みなさいよ。何とかした方がいいわよ。みんなに笑われるから。

M: ひょっとして僕がかぶれるような帽子、持ってないよね。