THE ROLE OF IMPLICIT MEMORY IN SECOND-LANGUAGE SPEECH PROCESSING: AUDITORY PRIMING IN JAPANESE LEARNERS OF ENGLISH

by

Noriko Matsuda

松田　紀子

A Dissertation

Presented to

The Graduate School of Language, Communication, and Culture

Kwansei Gakuin University

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

March 2017

Doctor of Philosophy Dissertation

The Role of Implicit Memory in Second-language Speech Processing: Auditory Priming in Japanese Learners of English

by

Noriko Matsuda

Members of Evaluation Committee

Major Advisor:

Associate Advisor:

Associate Advisor:

Associate Advisor:

ACKNOWLEDGMENTS

I would like to extend my appreciation to all the instructors at the Graduate School of Language, Communication and Culture at Kwansei Gakuin University.

First and foremost, I would like to thank Professor Shuhei Kadota, my mentor, for his continuous support, patience and encouragement. He has been a great role model for me. Without his instruction, invaluable advice, and constructive feedback, this dissertation would not have been possible.

I would also like to thank Professor Naoya Hase, who provided me emotional support during this research, as well assisted me in the always challenging task of finding participants for this study. I am also grateful to Professor Hiromi Otaka for his valuable tips regarding phonetic analyses and his ability to encourage me in my work as a researcher. My sincere thanks also goes to Professor Keiichi Ishikawa for his insightful comments and questions which inspired me to widen my research.

I'd also like to extend many thanks to the doctoral and master course students in Professor Kadota's seminar.

In addition, I wish to thank Hope Kron, an English conversation school instructor for spending a considerable amount of time and effort on checking my thesis and giving me feedback.

I also want to show my appreciation to the participants of my experiments: undergraduate and graduate students at Kwansei Gakuin University and Kindai University.

I would like to give my sincere thanks to my students, since their eagerness to learn inspired me to do research in the SLA field.

There are many others whom I cannot name individually who contributed directly and indirectly to my dissertation. I would like to express my sincere appreciation to all of them, because without their help, this dissertation would not have seen completion.

Last but certainly not least, I would like to express my gratitude to Kwansei Gakuin University for granting me a scholarship which has provided me an opportunity to thrive in a rich academic environment.

ABSTRACT

The Role of Implicit Memory in Second-language Speech Processing: Auditory Priming in Japanese Learners of English

by
Noriko Matsuda

The present study investigated the role of implicit memory in speech processing of Japanese EFL learners using auditory priming experiments. In addition, perceptual learning training was conducted based on the outcomes of the auditory priming experiments to explore efficient ways of enhancing second language (L2) perceptual processing.

Implicit memory is closely related to language learning and acquisition as it is said to be the foundation of language. Moreover, the use of implicit memory, the ability to derive what one has learned without conscious recollection, is indispensable in real world situations. Three auditory priming experiments were conducted based on previous research to verify the involvement of implicit memory in L2 word perception and to examine the application of the exemplar-based language model (EBM).

The goal of Experiment 1 was to investigate auditory word priming in Japanese EFL learners and native English speakers and decipher the features of the L2 priming effect. Both groups showed priming effects, indicating the existence of a common

mechanism for acquisition or learning of both L1 and L2. Experiment 2 aimed to examine the influence of speaker variability on the priming effect in Japanese EFL learners. The results showed that L2 learners could not process linguistic information and paralinguistic information separately, suggesting the possibility of a mechanism that allocates larger amounts of cognitive resources to processing meaning, as in L1 speech perception development. Experiment 3 compared the priming effect using natural human speech versus synthetic speech. Results showed that lower proficiency learners of L2 can gain a certain perceptual learning effects with using synthetic speech; however, great learning effects were seen when participants were exposed to the same natural human voice.

In order to shed some light on effective repetition methods that would help Japanese EFL learners in gaining L2 speech knowledge, Experiment 4 examined the effects of auditory word repetition on online performance and Experiment 5, on offline performance. The results revealed that more repetition led to swifter responses of L2 words, and that vocal repetition rather than subvocal repetition following semantic tasks helped learners to produce each word more accurately and rapidly in both the priming experiment (online) and recognition task (offline).

The results of Experiments 1 through 3 verified the involvement of implicit memory in L2 language learning and the possible application of EBM. Moreover, the overall findings of Experiments 4 and 5 consistently underscored the importance of well-planned perceptual learning for Japanese EFL learners. The results of this study consistently showed L2 learners' sensitivity to perceptual information. As this might be due to lack of exemplars in L2 speech knowledge, learners should expose themselves to a large amount of L2 input and it should be varied in order to build a robust representation of L2 speech. This study suggests that the need for accumulating a wide base of exemplars is likely to have a significant influence on L2 learning. Therefore, providing opportunities for acquiring a variety of exemplars with efficient perceptual learning methods should be considered a critical issue for English education in Japan.

Contents

Introduction

The main purpose of the present study is to explore the cognitive processes related to implicit memory, which serves as the basis of language, examine its role in second language (L2) acquisition, and discuss pedagogical implications. In this paper, L2 language processing in Japanese learners of English is analyzed from the perspective of perceptual learning, especially as it is important in relation to implicit memory. In foreign language education, particularly English, the attainment of proficiency centered on speaking has taken on new urgency with the advancement of globalization. This study focused on speech processing involving the auditory priming effect, which is thought to be a universal mechanism facilitating speech acquisition.

By repeating a particular action, one is able to do it quicker, more naturally, and more efficiently. This is a function of implicit memory. It has long been known that this is the type of memory that serves as the basis of language acquisition and learning (Schacter & Tulving, 1994). For this reason, the researcher believes that research into implicit memory has the potential for important implications not only in one's mother-tongue (L1), but also regarding L2 learning. For better understanding, an overview of the classification of types of memory is given.

According to an information processing concept in cognitive psychology, the memory process can be divided into three stages: encoding, storage and retrieval (Melton, 1963). At the encoding stage, the information that comes into our memory system by sensory input is changed into a form that can be stored. The next stage, which can be called memory storage, is information retention in sensory, short-term and long-term memory. Memory retrieval is the final stage of the process where previously stored and encoded information is accessed again.

Ohta (2011) showed various types of memory based on multiple memory systems theory (e.g., Tulving, 1987) (Figure 1). When a stimulus from the outside world enters the sensory memory, it is retained for only a very short length of time (at most a few seconds). Information to which one's attention is drawn is entered into the short-term memory store, where it is retained for a short length of time. The retention period is usually around one minute, but can be extended slightly longer through rehearsal or elaboration.[1] The information is stored and processed by the working memory (short-term memory). From here, part of the information is transferred to the long-term store, where it will be retained for anywhere from several minutes to the rest of one's life. There are two kinds of long-term memory: episodic and semantic, and an overlap of these two types is autobiographical memory. These are also called declarative memories as they can be expressed linguistically. Non-declarative memory contains priming memory[2] (from the perception to semantic levels) and procedural memory (memory of processes related to skill learning). Semantic memory, priming memory and procedural memory

---

[1]Elaboration means using knowledge already possessed to give meaning to something (Craik & Tulving, 1975).

[2]Tulving and Schacter (1990) called it presemantic perceptual system (PRS) which is described later in this study (1.2.1.).

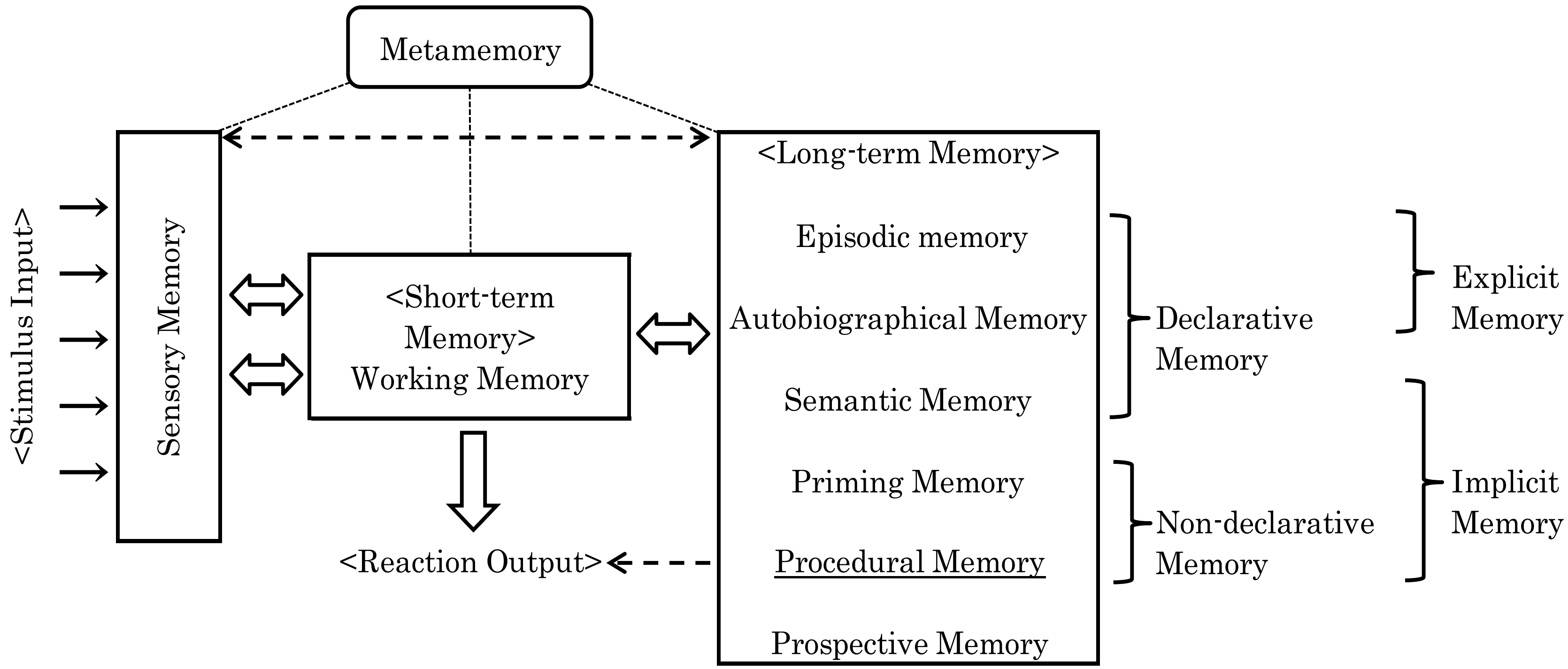

Figure 1. Classification of memory. Revised from Ohta (2011).

together form implicit memory. Implicit memory is a type of memory that does not require conscious recollection (remembering of episodes as one's own experiences [Graf & Schacter, 1985]). Since episodic memory and autobiographical memory are accompanied by conscious recollections, they are types of explicit memory. Other types of memory also exist, such as prospective memory (memory of acts to be executed in the future), metamemory (related to all types of memory), and so on. The exchange of information within implicit memory is shown in Figure 1. It is important to recognize the point where the line extends directly from sensory memory to long-term memory.

As stated before, implicit memory is deeply related to repetition. Repetition, or rehearsal, is said to be a fundamental form of learning as well as an effective language learning method that is indispensable for achieving proficiency and automatic language use. Research on shadowing[3] or repeating[4] has advanced in Japan due to the fact that it is regarded as an efficient learning method to develop learners' phonetic perception and articulation abilities (Kadota, 2007; Tamai, 2005). Though phonetic perception and articulation abilities are indispensable to becoming a fluent communicator in a second language (L2), most learners in Japan are not taught to enhance them in secondary school. In general, adult Japanese EFL learners, who are now in

---

[3]In this study, shadowing is defined as an immediate word-for-word repetition task which requires learners to repeat the speech of someone while listening (e.g., Torikai, Tamai, Someya, Tanaka, Tsuruta, & Nishimura, [2003]).

[4]In the present study, repeating is defined as a verbatim repetition task that requires a pause before repetition of speech.

their 20's and above, started studying English from junior high school using the grammar-translation method in their first language (L1) and focusing on passing paper-based entrance exams. As a result, the amount of phonetic L2 input was limited, causing dissociation between learners' phonological knowledge and actual spoken words.

According to the Educational Testing Service, the average TOEFL iBT score of Japanese learners was the fifth-worst out of 36 Asian countries in 2015. More precisely, the average score of the listening section was the fifth-worst and the speaking section was the lowest (Educational Testing Service, 2015). While other factors, such as the total number of examinees, should be taken into account when reading the data, the results seem to show that limited auditory input caused several problems in the Japanese learners' processing of spoken English.

There has been continuous innovation in English education in Japan to combat this issue. Since 1987, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) has assigned native English teachers as assistant language teachers in public schools (JET Programme, 2014), contributing to an increase in auditory L2 input. Moreover, since a listening exam was introduced in 2006 for the English exam of Daigaku Nyugakusha Senbatsu Daigaku Nyushi Center Shiken (literally, the University Candidate Selection University Admission Center Test), English classes in schools are likely to have increased the amount of auditory L2 input. In addition, English has become a compulsory subject in elementary school from the 5th-grade since April of 2011, further increasing the amount or time period of L2 input for young

EFL learners.

The recent introduction of a new entrance exam system using the Test of English for Academic Purposes (TEAP) or the Global Test of English Communication (GTEC) might cause drastic changes in English education in Japan if their use by universities becomes widespread.[5] In 2013, the educational panel of the ruling Liberal Democratic Party suggested that Japanese universities should use TOEFL scores as one of the criteria for college enrollment, which served as a trigger to introduce the new system. In the age of globalization, the government is preparing an enormous investment into educational reform in order to bolster economic growth in the future. Since the tests can evaluate listening, speaking, reading and writing skills, teachers and schools are now under pressure to transform classroom English education. Though these educational reforms may be positive steps toward solving the problem of limited L2 input in the future, the learners' gap between phonological knowledge and actual spoken words is still likely to persist.

The second purpose of this study is to investigate the effects of L2 auditory repetition on speech processing in order to suggest effective methods of repetition that suit Japanese EFL learners to gain L2 speech knowledge. Specifically, empirical studies using the auditory word priming paradigm were conducted to understand

---

[5]The main developers of TEAP are Sophia University and the Eiken Foundation of Japan, while GTEC is an online English test developed by Berlitz Corporation and Benesse Corporation. The tests are designed to test students who learn English as a second language, and consists of four sections: reading, listening, writing, and speaking.

the cognitive processes of speech learning. Auditory priming is said to be a learning mechanism related to automatization in decoding. Clarifying the underlying mechanism of automatization is indispensable to explore the pedagogical implications of repetition in language learning. In addition, this study hopes to accumulate speech processing data of Japanese EFL learners for further studies.

The contents of this paper are as follows: Chapter 1 provides background information of implicit memory, auditory priming research, and auditory repetition. Chapter 2 shows the aims and hypotheses of the five experiments conducted in this study. Chapters 3 through 6 report on Experiments 1 through 5, conducted based on previous research presented in Chapter 1. Experiment 1, dealt with in Chapter 3, aims at investigating auditory word priming in Japanese EFL learners and native English speakers. Chapter 4 covers Experiment 2, which seeks to monitor the priming effect on Japanese EFL students considering the influence of contextual details, namely speaker variability, in L2 input. Chapter 5 describes Experiment 3, which looks at the priming effect in natural human speech and synthetic speech in order to assess the applicability of text-to-speech (TTS) synthesized technology in English education in Japan. Experiment 4, explained in Chapter 6, explores the effects of auditory word repetition on online performance using a priming experiment. The same chapter deals with Experiment 5, which investigates the effects of auditory word repetition on offline performance using a recognition task. Chapter 7 provides a discussion of the results of these five experiments from the view point of implicit memory and perceptual learning. The final section includes the conclusion and implications for English education in Japan and some issues for further research.

Chapter 1
Previous Studies

1.1 Implicit Memory Research and Theoretical Implications

There are two prominent types of human memory: implicit and explicit. Implicit memory does not require any explicit recollection of former experience and it plays an important part for humans in perceiving speech sounds. For instance, people usually recognize a voice on the phone without actually seeing the person calling them since they retain some acoustic properties as implicit memory.

In the mid-1970s, implicit memory was brought to light due to research into memory in amnesiacs (Warrington & Weiskrantz, 1974), and by the 1980s there were many studies in this field, particularly on priming effects. Theoretical explanations of the phenomenon were proposed after various psychological verification experiments. According to these studies, the processing of linguistic information and non-verbal information (paralinguistic information) is performed at the perceptual level, which then becomes implicit memory.

The main features of implicit memory as understood by psychological experiments are (1) long-term persistence, (2) sensitivity to perceptual information that lacks meaning and to changes in modality, and (3) not affected by aging (Roediger & Mcdermott, 1993). Various pieces of sensory information persist for each modality over a long time period. Memories of people's faces, for example, are believed to be retained on a monthly basis based on the number of times they were seen (Sloman, Hayman, Ohta, Law & Tulving, 1988). It is therefore highly probable that people hold various types of information for a long time at a level where semantic processing is not performed.

Although some rare cases have been documented, such as the savant syndrome and the hyperthymestic syndrome, where the affected are said to be capable of remembering all episodes, the majority of people are not conscious of the huge amount of sensory information typically accumulated. Some researchers believe this suggests that healthy people possess similar storage capabilities but they cannot be used as explicit memory (Kuroda, 2010; Terasawa, 2016). Since this paper deals with linguistic information, it is worth considering what kind of language model can be proposed from the series of studies on implicit memory.

According to the Usage-Based Model (UBM) (e.g., Langacker 1987, 2000, 2009; Kemmer & Barlow, 2000), a well-known language memory model, language is acquired through concrete linguistic experience, and language knowledge is constructed from a huge network of 'schemas' based on language expressions. When patterns that repeatedly occur in linguistic experiences are turned into knowledge, they become established as 'units,' which become abstracted 'schemas,' while the actual situations that occur in real life are their 'instantiations.'

A rival model is the exemplar-based language model (EBM). Considering the results of implicit memory research where individual exemplars of perceptual information are retained for a long time (e.g., Gahl & Yu 2006; Johnson 2005, 2006; Pierrehumbert, 2001; Port, 2007), an appropriate model for the present study appears to be EBM. According to this model, language is considered to be an accumulation of exemplars, not abstracted information, making this a highly descriptive language model of implicit memory.

EBM defines innumerable quantities of features as qualities in individual exemplars, creating structures that correspond to UBM 'schemas' by organizing these features. The basis of EBM is the idea that every exemplar is memorized. Both this concept and the outcomes of the research into implicit memory are compatible. In this paper, the applicability of EBM as an L2 language model will be analyzed through several experiments. This study mainly examines data pertaining to the perception of L2 since the implicit memory phenomenon emerges at the perceptual level. The following sections will address priming studies (1.2) and perceptual learning (1.3).

1.2 Auditory Word Priming

1.2.1. Implicit Memory and Auditory Word Priming

Priming is defined as the "facilitative effects of an encounter with a stimulus on subsequent processing of the same stimulus (direct priming) or a related stimulus (indirect priming)" (Tulving, Schacter, & Stark, 1982, p. 336). Priming occurs because people commonly use previous information to carry out their daily routines smoothly and efficiently. Tulving and Schacter (1990) first pointed out that direct priming, or repetition priming, was a constructive concept of implicit memory. Moreover, its relative insensitivity to the type of processing (e.g., semantic or nonsemantic) in the study phase suggested the existence of a presemantic perceptual system in human memory. Tulving and Schacter (1990) called it the "perceptual representation system" (PRS). Tulving (1995) stated that human memory consists of three main implicit memory systems (procedural memory, PRS,[6] and semantic memory) and two main explicit memory systems (primary memory and episodic memory).

While most studies were designed to show the effects of the visual PRS, studies on the auditory PRS were limited (e.g., Church & Schacter, 1994; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Pilotti, Gallo, & Roediger, 2000; Schacter & Church, 1992). To verify the existence of the auditory PRS, Schacter and Church (1992), as well as Church and Schacter (1994), conducted a priming experiment by manipulating the type of processing of stimuli in the study phase.

The types of processing can also be referred to as the levels of processing (LOP). The LOP framework, proposed by Craik and Lockhart (1972), explained the different levels of information processing in the stages of perception, encoding, storage, and retrieval (usage). In particular, they attempted to intellectualize the reason for high retention scores of deeper information, such as semantic level information, compared to shallower information, such as phonemic level information. They explained that deeper information could be retained longer in human memory and recalled more swiftly compared to shallower information, because semantic encoding of incoming verbal information could be integrated with existing knowledge (elaboration: Craik & Tulving, 1975). The LOP framework, has a significant influence on human memory research to this day. The importance of retrieval factors, as well as encoding factors, has been stressed in a series of LOP studies (Craik & Tulving, 1975; Bower & Winzenz, 1970; Walsh & Jenkins, 1973). Similarly, Morris, Bransford,

[6]It is called priming memory in Figure 1.

and Franks (1977) stated that memory performance is not only determined by the levels of processing but also by the relationship between how information is initially encoded and subsequently retrieved. They claimed that semantic encoding was usually very effective because the retrieval processes of recall and recognition also involved semantic processing (transfer-appropriate processing = TAP principle). It is similar to the encoding specificity principle focusing on the interaction between encoding and retrieval processes (Tulving, 1979; Tulving & Thomson, 1973). However, Craik (2002) stated that the concept of TAP and encoding specificity seem complementary.[7] Importantly, this framework can be applied to the explicit memory tasks such as word recall or word recognition.

Auditory word priming used in the present study is a form of direct priming or repetition priming, and it is an implicit memory task. It is said to be a mechanism that supports spoken-word processing and learning (Church & Fisher, 1998; Church & Schacter, 1994; McDonough & Trofimovich, 2009; Schacter & Church, 1992; Trofimovich, 2005). In typical auditory word priming experiments, participants listen to a set of spoken words as stimuli for the encoding phase (the study phase) of the experiment. In the second phase (the test phase), they are tested using a set of both previously heard and unheard stimuli (new words). Most participants show significantly more rapid and accurate processing of repeated words compared with new words in the test phase in both L1 studies (Bassili, Smith, & MacLeod, 1989; Onishi, Chambers, & Fisher, 2002; Pilloti,

[7]Although the framework is said to include some debatable points such as a lack of depth measurement, the simple framework still helps to provide a better understanding (Craik, 2002).

Bergman, Gallo, Sommers, & Roediger, 2000; Schacter & Church, 1992) and L2 studies (Trofimovich 2005, 2008; Trofimovich & Gatbonton, 2006; Woutersen, de Bot, & Weltens, 1995). Listeners seem to encode and store a number of details, such as acoustic properties, when they are exposed to spoken words, and the memory of the details appears to promote reprocessing (e.g., repetition, recall). This sequence of phenomena was called "auditory word priming," and was said to originate at the perceptual level of speech. Auditory priming effects in different processing conditions (semantic or nonsemantic) showed no significant difference in the successful priming experiments of L1 research (Church & Schacter, 1994; Pilotti et al., 2000; Schacter & Church, 1992). Auditory word priming is also said to be an indicator of "[the] listeners' sensitivity to the formal (as opposed to meaningful) properties of language" (Trofimovich & Gatbonton, 2006, p. 521). Its nonsemantic nature is one of the traits of auditory priming in L1.

1.2.2. Auditory Word Priming in L1

There are four important traits of auditory word priming in L1: developmentally constant, long lasting, stimulus specific and nonsemantic nature (McDonough & Trofimovich, 2009; Trofimovich, 2005). The first characteristic is its developmentally constant nature. Church and Fisher (1998) recorded auditory word priming affects young children, while Pilotti and Beyer (2002) observed the effects on older persons (from 65 to 88 years of age). These results showed that the robustness of auditory word priming remained, regardless of age. The second characteristic is the long-lasting nature of auditory word priming. According to studies, the effects were said to last for minutes (Church & Schacter, 1994) or even weeks (Goldinger, 1996). It can be presumed that the effects become a part of long-term memory. The third characteristic is its stimulus-specific nature. While speaking and understanding spoken words, listeners seem to encode and store a large number of details regarding what they hear, such as the speaker's voice, intonation, and pitch. The information is then available at a later time to comprehend speech and to recite some of the words. For instance, research showed that repeated words spoken in a previously heard voice could be processed faster than the same words spoken by a different person (Goldinger, 1996; Sheffert, 1998). Speaker variability seems to affect the priming effect in L1. Finally, several L1 auditory word priming studies revealed insensitivity to encoding manipulation in the study phase as mentioned in the previous section. Although listeners' attention was manipulated according to the different types of processing in these experiments, the effects were nearly the same (Church & Schacter, 1994, p. 527; Schacter & Church, 1992, p. 926). Its non-semantic nature is peculiar to perceptual priming. Considering these features, it is probable that auditory word priming provides some support for processing spoken words in L1.

1.2.3. Auditory Word Priming in L2

There are several L2 experimental studies of auditory word priming (Trofimovich 2005, 2008; Trofimovich & Gatbonton, 2006; Woutersen, Cox, Weltens, & de Bot, 1994; Woutersen, de Bot, & Weltens, 1995), which show the auditory word priming effect in processing L2 words. However, the long-lasting and developmentally constant nature of auditory word priming has not been confirmed in L2 studies. It is more complicated to verify them in an L2 setting because the proficiency levels can vary among individuals at any age.

Regarding the stimulus-specific nature, research demonstrated that learners were over dependent on minute context-specific information of spoken L2 words compared with L1 (Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Goldinger, 1996; Trofimovich, 2005). According to Trofimovich (2005), the priming effect of L2 learners (20 learners of Spanish whose L1 was English) could be seen only when the words were spoken in the same voice. This suggests that speaker variability might drastically affect the priming effect in L2 (discussed further in 1.2.5.).

Moreover, semantic processing at the encoding stage seemed to reduce the priming effects of L2 learners, at least at the beginning of their learning (Kirsner & Dunn, 1985; Trofimovich & Gatbonton, 2006[8]). Trofimovich and Gatbonton (2006) explained the result by using a memory study theory called the "transfer-appropriate-processing" (TAP) principle (Morris, Bransford, & Franks, 1977) as "a mismatch between information-processing demands on learners at the time of study and at the time of testing" (p. 529). According to the TAP principle, the auditory priming effect under the focus-on-meaning condition can be smaller because learners are not required to perform any semantic processing during the test phase.

Trofimovich and Gatbonton (2006) demonstrated this in their experiment with 60 L2 learners of Spanish who were native English speakers. The participants were asked to rate the clarity of each word in the focus-on-form condition and the pleasantness of each word in the focus-on-meaning condition in the study phase. In the test phase, an

[8]Trofimovich and Gatbonton (2006) found that a focus-on-meaning condition decreased the priming effect only for low pronunciation accuracy learners.

auditory repetition task requiring participants to only listen to a series of words and repeat each word quickly and accurately was used.

Trofimovich (2008) also explained the nonsemantic nature of the L2 priming effect, stating that attention to word meaning might decrease learners' sensitivity to phonological details because their short-term phonological memory capacity is limited.

In addition, priming effects decreased greatly when learners were exposed to a combination of different voices and semantic processing (Trofimovich, 2008). Previous research in L2 implies that auditory word priming may be involved much differently in L2 spoken word processing and learning.

There were some issues to be addressed in the priming methodology of previous L2 studies. When participants measured the pleasantness of words in the focus-on-meaning condition, some of them may have used their episodic memory[9] of words, while others did not. As previously noted, episodic memory is one of the explicit memory systems while priming is one of the implicit memory systems (Tulving, 1995). It is necessary to devise a method that enables participants, particularly L2 learners, to be less affected by explicit memory and process words in a unified manner.

1.2.4. Auditory Word Priming in Japanese EFL Learners

Although auditory priming shows a clear effect on L2 word processing, there have been limited auditory priming studies with Japanese EFL learners. Sugiura and Hori (2012) conducted an auditory

[9]Episodic memory involves personal memories (e.g., memory about who, when, where, and what) and varies among different people (See Figure 1).

priming experiment in the same manner as previous studies (Trofimovich, 2005, 2008; Trofimovich & Gatbonton, 2006) using both L1 and L2 words. This study demonstrated that "Japanese learners of English use auditory priming to facilitate spoken-word processing," regardless of word familiarity and language types (Japanese or English) in the stimuli and the learners' proficiency. Because there is minimal understanding of L2 auditory word priming in Japanese EFL learners, it is worthwhile to examine whether and to what extent it is involved in L2 word processing. Furthermore, when it comes to the stimulus specific nature, little is known concerning Japanese EFL learners. Accordingly, it is also worth considering the effects of minute context-specific information of spoken L2 words, such as speaker variability.

1.2.5. Auditory Word Priming in Speaker Variability

This section covers the stimulus-specific nature of auditory priming, particularly investigating speaker variability.

When listening to the news on TV in one's L1, one seldom experiences sudden difficulty understanding what is being said when announcers change. However, many learners of a second language have trouble understanding a new speaker. During language learning and acquisition, it is impossible to correctly understand what is being said if one is unable to ignore the variations in prosody and pronunciation between different speakers in order to identify and retain vocabulary patterns and associate meanings with those patterns. In spoken language, listeners cannot separate the linguistic information or content from the acoustic elements (paralinguistic information), such as differences between speakers' voices and emotional inflections. Therefore, exploring how such variations are processed is an important topic in considering the mechanisms of spoken language learning and acquisition.

1.2.5.1. Previous studies of Speaker Variability in L1

Based on perception and cognitive research of L1, it is widely believed that the capacity to identify common linguistic information from different speakers is present in early infancy, before infants acquire the ability to link words with meaning. Research on L1 speech perception development in infants has shown that at two months of age, the learning of syllables does not proceed very well when infants are exposed to multiple speakers (Jusczyk, Pisoni, & Mullennix, 1992). According to Houston and Jusczyk (2000), common linguistic information can be recognized between speakers of the same gender at seven and a half months of age, but not between speakers of different genders. Processing information with no influence of speaker variability becomes possible at ten and a half months of age. Infants begin to be able to process linguistic information independently from the different acoustic features of individual speakers' voices.[10] At 12 months of age, for the most part, they are able to do this quite well.

Once infants are able to independently process linguistic information in this way, their attention to information relatively less important for understanding spoken language, such as acoustic features in utterances, is inhibited for some time. One such example is native speakers of Japanese, who at the age of two have difficulty learning new words similar to words they already know but differing in pitch accentual

---

[10]There are various views on whether the processing is entirely independent. Some researchers argue that there is some interaction, while others insist that there is none (Ikeda & Haryu, 2016).

patterns, while they can learn such words easily at the age of three (Yamamoto & Haryu, 2016). In addition, despite the fact that children may be sensitive to emotional prosody in utterances in early infancy, from infancy to later childhood the phenomenon of lexical bias comes into play: they prioritize the linguistic content of utterances over the manner of speaking to infer the speaker's feelings (Friend & Bryant, 2000). While a speaker's way of speaking is given more importance as children grow older, this change overlaps precisely with the period of development of the central executive, which is believed to control human attention (Chevalier, 2015; Cowan, Morey, AuBuchon, Zwilling, Gilchrist, 2010; Jerger, Martin, & Pirozzolo, 1988). In bilingual children, this development starts earlier, and they are able to discern speakers' feelings early on (Yow & Markman, 2011).

While this research field continues to be subject to many debates concerning L1 speech perception development,[11] as suggested in this overview, some important findings have been made. As children's knowledge of their L1 develops, they become able to independently process paralinguistic information (modularity of processing). Further, children appear to allocate limited cognitive resources to more important linguistic information as they develop the ability to switch attention.

As previously mentioned, the priming effect of L1 seems to be affected by speaker changes to some extent. However, various L1 studies

[11] It has been pointed out that there is a possibility that not all studies on this subject are looking at the same factors. Many studies on infants exposed them to specific sounds and (continued) looked at gaze duration as the response. For older children, methods such as exposure to specific sounds were used while attempting to elicit responses from the children. These studies differ because gaze duration measures implicit processing capability, while monitoring children's answers measures explicit processing capability.

on adults have demonstrated the robustness of their ability to cope with speaker variability. One study has shown that even in a slightly noisy environment speaker adaptation occurred after listening to only five syllables spoken by a single speaker (Kato & Kakei, 1988). There are some representative phonological research models that explain this phenomenon; namely, models based on the idea of *speaker normalization* (e.g., Ames & Grossberg 2008; Johnson, 2005) and the *Exemplar Model* (e.g., Pisoni, 1997; Hintzman 1986; Nosofsky 1991; Goldinger, 1998; Pierrehumbert, 2001). In addition, models combining both ideas have been developed recently (e.g., Hawkins & Smith, 2001; Hawkins, 2010). *Speaker normalization* assumes the existence of abstract, standardized representations, while the *Exemplar Model* assumes the existence of cognitive representations derived from the accumulation and integration of examples from experience. Moreover, adults' capacity to flexibly cope with various environmental changes in their L1 is said to be due to perceiving speech hierarchically (top-down processing) through comprehensive use of not only its acoustic aspects, but also a variety of information from memory, experience, and knowledge.

1.2.5.2. Previous studies of Speaker Variability in L2

Variations in paralinguistic information in L2, in contrast to those in L1, are known to be a factor imposing cognitive loads on L2 learners' speech processing. As described in the previous section, some L2 priming studies have shown that priming effects were greatly reduced if a word that the participants had learned once was repeated by a different voice. Thus, it is likely that L2 learners who do not develop phonological information databases in the L2 may have difficulty independently processing the acoustic features and linguistic information conveyed by different speakers. In related research, one study looking at the relationship between bilingual proficiency levels and speaker recognition has shown that familiarity of the target language was closely associated with speaker recognition, suggesting the possibility that L2 speech processing and speaker recognition may be linked (Bregman & Creel, 2014). This study also suggested that, from the perspective of L2 learning, exposure at an early stage may be important for the formation of L2 phonological representations. Moreover, a number of studies have shown that for both adults and children, more robust representations can be formed by being exposed to different L2 speakers during the early stages of learning (Lively, Logan, & Pisoni, 1993; Kingston, 2003; Rost & McMurray, 2009, 2010).

Based on research on L1 speech perception development in children, we may need to consider the possibility that L2 speech containing paralinguistic information may be processed quite differently depending on the kinds of linguistic information to which the limited cognitive resources are allocated. While cognitive resource allocation involves the development of the attention switching function of the central executive in children, in adults the central executive function itself can be assumed to be adequately developed. Therefore, for adult L2 learners, it is highly likely that cognitive resource allocation in L2 processing is determined by the focus of the learner's attention. In L1 research, there were no differences in the priming effects (perceptual learning effects) when participants listening to vocabulary words focused their attention on either the sound or the meaning of the words. In contrast, L2 studies have suggested that focusing on meaning results in negative priming effects, depending on the learner's proficiency.

A priming experiment using speaker variability and attention focus as independent variables in both L1 and L2 showed no effects of speaker variability on the priming effects in L1, while in the L2 speaker variability caused negative effects and no priming effect was seen regardless of whether attention was focused on sound or meaning (Trofimovich, 2005). In a similar L2 study, using length of stay in the L2 country as well as attention focus and speaker variability as independent variables, the priming effect could be seen only in the longer-stay group when attention was focused on sound (Trofimovich, 2008). These two studies used audio recordings of the speech of six native speakers. When the speaker was changed in the test phase, speech of a speaker of the other gender was used. Further, for the sound-focused task, participants were asked to rate the sound clarity of each word, and for the meaning-focused task, they were asked to rate the pleasantness of word meaning (i.e., to rate how fun the meaning of each word was). The implication of these studies was that when the participants focused on the sound of words, change of speakers did not affect speech processing of participants who had had a long exposure to the L2; however, when participants focused on meaning, a change of speakers greatly reduce the priming effect. We can predict that for Japanese EFL learners who are in environments with little exposure to English speech input, speaker changes will reduce the priming effect regardless of the attention focus. Furthermore, very importantly, the combined effect of focusing attention on meaning and speaker variability is likely to produce a large decrease in the priming effect.

Natural human speech includes speaker variability. Unfortunately, as mentioned above, there is a lack of L2 speech input in English education in Japan. The use of text-to-speech (TTS) synthesized technology is expected to remedy this problem to some extent. In fact, the number of applications of synthesized speech software in English language classroom has increased in recent years. However, it still remains unclear whether synthetic speech has similar learning effects as natural human speech for effectively learning a second language. The next section discusses the priming effect when using synthetic speech.

1.2.6. Auditory Word Priming in Natural Human Speech and Synthetic Speech

1.2.6.1. Background

TTS (text-to-speech) synthesizing software that allows teachers and students to freely create foreign speech has an enormous potential to solve the problem of limited second language (L2) input. Several cases of speech synthesis in English-language classrooms have been reported following the rapid advancements in speech synthesis technology in recent years (Azuma, 2010; Kataoka & Ito, 2013). These cases reveal a variety of potential advantages, from the possibility of developing different kinds of speech learning material to broadening educational activities. Adding and editing data is simplified using speech synthesizing software, which could lighten the workload for teachers by eliminating the need to contact native speakers individually and record and edit their voices. In addition, the use of synthesized TTS is not limited to learning activities as it also has the potential of aiding in research, such as in conducting psycholinguistic experiments, again, because it is easy to control the necessary stimulation. However, despite these various applications, there are few studies on the application of TTS for foreign-language classrooms or comparative studies to natural human speech (Azuma, 2010; Kashiwagi, Kang, & Ohtsuki, 2008).

To be able to automatically process phonetic input (a conscious process that increases in speed after repeated drills and transitions into an unconscious process), which is the basis of the spoken language process, it is essential for foreign language learners to be able to correctly decipher what they hear in the target language. As a result, there have been a large number of studies in the past ten years in the field of foreign language education in Japan, focusing on the effects of training that facilitates the perceptual process, such as shadowing (a training method wherein learners immediately repeat what they hear) (e.g., Kadota, 2007, 2015; Tamai, 2005). Therefore, understanding the benefits of synthesized TTS for this type of training offers the potential of using synthesized TTS to improve the listening skills of Japanese learners of English. With this background, the researcher conducted a priming experiment to compare and investigate the perceptual learning[12] effects of using synthesized TTS and natural human speech.

1.2.6.2. Previous Studies on Priming

Previous studies of auditory priming have shown that learners memorize the acoustic properties of a voice and use the information unconsciously (Trofimovich & Gatbonton, 2006). This represents a learning effect at the perception level. Auditory

[12]Perceptual learning effects can be defined as the changes in perceptual (or sensory) systems, as observed through behavior, such as fast and accurate recognition of the target word. The conception of perceptual learning will be discussed in the next section (1.3).

priming is known to be a universal mechanism that aids in language acquisition. Additionally, it has been suggested that this mechanism may also work in the acquisition of languages other than L1 (McDonough & Trofimovich, 2009).

As discussed in the previous section, studies focused on L1 recorded no visible differences in the priming effect when listening to vocabulary, whether the focus was on the sound or the meaning of the material presented (McDonough & Trofimovich, 2009; Trofimovich, 2005). However, contrary to acquiring L1, the few studies that have focused on L2 indicate that there is a negative impact on the priming effect based on a person's proficiency when focusing on meaning (Trofimovich, 2005, 2008; Trofimovich & Gatbonton, 2006). These studies explain that "L2 learners may not benefit from repeated experiences with spoken words, at least early in their L2 development or after a relatively brief experience with the L2, when they engage in a meaningful, semantic processing of words." (= no perceptual learning effect) (McDonough & Trofimovich, 2009, p. 30). The subjects of these studies were L2 learners in auditory-input-rich ESL environments. These studies also had various definitions for proficiency. Trofimovich (2008) defined the barometer of proficiency as the length of residence in the country where L2 is the national language, while Trofimovich and Gatbonton (2006) defined it as the degree of pronunciation ability. The auditory priming effect itself can also be seen in studies where subjects were Japanese students in English in EFL environments dissimilar from other ESL environments (Sugiura & Hori, 2012). However, the researcher could not locate detailed studies of the auditory priming effect on EFL learners that considered both proficiency and focus when students were listening to vocabulary.

1.2.6.3. Previous Studies on Synthetic Speech

The most popular kind of speech synthesis technology in use today is rule-based speech synthesis known as corpus-based speech synthesis technology based on a large-scale database from natural voices such as from professional announcers. It "generates synthesized speech by editing the voice waveform segment data and varying it for intonations and such according to synthesis rules established beforehand" (Watanabe, Iwaki, Kaneyasu, & Miki, 2006). This is characterized by speech that feels authentic because it connects fragments of natural human speech. The TTS synthesis software used in this experiment also uses this method.

There is continuing research into intelligibility and comprehensibility in synthetic speech. Studies on intelligibility relate to this study especially because the study objective is to understand the perceptual learning effect; however, multiple studies are being conducted to find contributing factors, such as how age differences in students effects the outcomes (e.g., Drager, Reichle, & Pinkoski, 2010; Pinkoski-Ball, Reichle, & Munson, 2012) or repetition effects (e.g., Koul & Clapsaddle, 2006; McNaughton, Fallon, Tod, Weiner, & Neisworth, 1994; Reynolds & Jefferson, 1999) in one's native language. In addition, speaking or speech rate, noise, linguistic context, and practice effects have all been presented as factors that influence speech intelligibility (Axmear et al., 2005, p. 245). However, speech rate has been found to be an especially important factor that influences not only intelligibility, but also comprehensibility (Jones, Berry, & Stevens, 2007).

Few studies exist that focus on speech intelligibility in L2. Axmear et al. (2005) assigned repetition tasks to monolingual and bilingual children that revealed that intelligibility was higher for natural voices than synthetic ones and intelligibility of synthetic speech was lower in bilingual children than in monolingual children. Similar results were obtained with adults in Venkatagiri's study (2005), even though written and not repetition tasks were assigned.

Hirai and O'ki (2011) focused on the comprehensibility of synthetic speech with Japanese learners of English. This study indicated that although comprehensibility among learners tended to be higher with natural speech, synthetic speech was perceived to be almost the same as natural. Moreover, the "experience effect" influenced the comprehensibility of synthetic speech after hearing the speech once. Despite this, a higher percentage of students with low proficiency (25.0%) preferred synthetic speech compared to students with higher proficiency levels (8.3%). The authors believe this is due to the fact that "synthetic speech is read at a constant speed in all sections of the speech, and each word is regularly segmented," making it easier for the "lower proficiency listeners" to listen to it (p. 13). The authors argue that their study shows that synthetic speech can be used for English education.

Based on previous studies of L2 speech intelligibility, the perceptual learning effect can be expected to be greater when using natural speech rather than synthetic speech especially for students with higher proficiency levels. Also, it is likely that unnatural features of synthetic speech, such as steady reading speed and regular segmentation, will influence the preferences of higher proficiency level students and reduce the perceptual learning effects of synthetic speech.

One study that investigated auditory priming in Japanese learners of English showed the presence of priming effects when using recorded natural human speech (Sugiura & Hori, 2012). Although this study did not compare recorded natural human speech and synthetic speech, the researcher believes it is possible to compare the learning effect of using both speeches at the perception level by controlling various factors including speech rate.

As the auditory priming effect is a learning effect at the perceptual level, created by exposure to speech, the effect of repeated drills, or repetition can be discussed in the auditory priming paradigm. We will overview the repetition effect in the next section.

1.3 Perceptual Learning and Repetition

This section covers some of the previous research on auditory word repetition in L2. More specifically, previous empirical studies of repetition in the auditory word priming paradigm are described. In each section, current outcomes arising from the studies are pointed out. Before discussing repetition research, a definition of perceptual learning must be provided.

Perception is the basis of information processing related to all cognitive processing. According to some information processing models, perceptual learning is when new associations are made between sensory impressions and the memories stored in the brain; i.e., when the brain interprets new stimuli and reclassifies them. According to Goldstone (1998, p. 585), perceptual learning of speech is “relatively long-lasting changes to an organism’s perceptual system that improve its ability to respond to its environment and are caused by this environment.”

This study mainly analyzes L2 vocabulary, rather than phonemes or syllables, because there is a high possibility that the units of verbal recall are phonetically ‘words’. Moreover, unlike phonetics, which places emphasis on discussing the representation of sensory input, this study considers perceptual learning from the point of view of cognitive psychology. Therefore, the focus is on analyses of L2 word processing not only at the prelexical level, but also representations from sensory input to word recognition.

As long-lasting changes to an organism’s perceptual system are caused by frequent exposure or massed repetition, repetition effects should be considered from the view point of language learning.

The importance of repetition has been emphasized since the days of the audio-lingual method (Lado, 1964), and even accepted by researchers supporting communicative language teaching (Allen, 1983; Littlewood, 1981). In Japan, Takeuchi (2000) insisted that “repetitive practice is an indispensable learning style to establish and automatize basic language skills in the early stages of foreign language education” (p.131).

Learning a second language (L2) includes not only acquiring knowledge, but also the types of skill learning specific to linguistic performance (McLaughlin, 1987). McLaughlin (1987) stated that learning involves “the automatization of component sub-skills” (p.133); for example, phonetic perception (the first stage of decoding) is considered to be a sub-skill, or a lower-level listening skill. Investigation into the process of decoding in Japanese EFL learners may contribute to the development of more efficient ways of acquiring speech knowledge of L2 in an EFL setting. In addition, as previously noted, skill learning relates deeply to the procedural memory in implicit memory and the investigation may aid in understanding the role of implicit memory in language learning.

1.3.1. Decoding and Automatization

According to Field (2008), the refinement of decoding skills in second-language (L2) learners is of utmost importance. Decoding assumes the form of a matching process that includes “translating the speech signal into speech sounds, words and clauses, and finally into a literal meaning” (p. 125). Although the process is automatized in the first language (L1), for inexperienced L2 learners, the process is still complicated even at the perceptual level. This is because the ability to recognize the sounds of the target language, as well as the amount of known vocabulary, is limited. In fact, Goh (2000) revealed that five out of ten L2 listening problems reported by inexperienced learners were related to perceptual processing. As Field (2008) noted, because a high degree of automatization in decoding is necessary to become an expert, the attainment of decoding skills is a critical issue to be addressed.

Interpretations of automatization differ among researchers as a result of different views on the process. DeKeyser (2001) summarized theories of L2 skill learning into three approaches: rule-based, item-based, and the limited conversion of the two approaches. The rule-based approach, exemplified by a series of studies by Anderson (1976, 1983), argues that automaticity is the transformation of declarative knowledge into procedural knowledge through practice. On the other hand, the item-based approach regards automaticity as memory retrieval. According to Logan (1988), "automatization reflects a transition from algorithm-based performance to memory-based performance" through consistent practice (p. 493).

Although the limited conversion of the two approaches (Anderson, 1993; Delaney, Reder, Staszewski, & Ritter, 1998; Rickard, 1997) seems to compensate for the shortcomings of each approach, a wide gap remains between them. However, several of these studies used the same characteristics as criteria for describing automatization, in spite of having different views. Some of these characteristics include that automatization must be fast, capacity-free, unintentional, have little interference from and with other processes, unconscious, and as a result of consistent practice (DeKeyser, 2001, p. 128). With respect to these characteristics, previous empirical studies of repetition based on the auditory word priming paradigm must be discussed, because this paradigm concerns the effect of repetition with spoken input. In addition, it might also help us to understand the complicated learning process of L2 speech perception.

1.3.2. Empirical Studies of Repetition

The researcher classified the related studies of repetition into three categories: effects of the number of repetition, effects of the repetition method and processing orientation. Since the idea of processing orientation is based on the auditory priming paradigm, the details can be found in the previous section (1.2).

1.3.2.1. Effects of the Number of Repetitions

Repetition is considered to be a fluency-building[13] task that increases the speed and efficiency of cognitive performance (Schneider & Chein, 2003). In repetition experiments, the same stimuli are repeatedly presented and the reaction time (RT) gradually decreases as the number of repetitions increases. As previously noted, this is referred to as the repetition (direct) priming effect. On a broader scale, all repetition can be seen within this repetition priming paradigm. However, while participants tend to respond faster as the number of repetitions increases, this improvement of performance has been shown to be more drastic with the first few repetitions (Grant & Logan, 1993; Hu, Liu & Zhang, 2010; Salasoo, Shiffrin & Feustel, 1985). Moreover, Terasawa, Yoshida and Onishi (2008) found that learning English words more than 5 times a day appears to have no effect for memory retrieval of Japanese EFL learners. Thus, four times a day is likely to be enough for L2 vocabulary learning.

In summary, previous studies suggest that as the number of

[13]Fluency is defined as "the rapid, smooth, accurate, lucid, and efficient translation of thought or communicative intention into language under the temporal constraints of on-line processing" (Lennon, 2000, p.26).

repetition increases, the participants' responses accelerate and correct word retrieval increases. The number of effective exposures appears to be small — likely around four times a day.

1.3.2.2. Effects of the Repetition Method

When people try to retain information, they unconsciously use an inner rehearsal process, or subvocal rehearsal (subvocal repetition), in a phonological loop[14] of working memory as a learning system (Baddeley, Thomson, & Buchanan, 1975). To memorize L2 words, however, students usually use overt rehearsal (vocal repetition). Both types of repetition enhance perceptual fluency, though it has been suggested that vocal repetition uses several sensory organs resulting in multiple retrieval cues. Therefore, vocal repetition enables better retention (multimodality theory, as in Bäckman & Nilsson, 1984, 1985, or multiple cues effect, as in Ohta, 2016). In addition, vocal repetition is said to provide opportunities for auditory self-perception (Baker & Trofimovich, 2006). Thus, these repetition methods, vocal and subvocal repetition, may have different beneficial effects on learner retention and phonetic development of L2 words.

The above mentioned studies suggest that vocal repetition may shorten word processing time and decrease the error rate more

---

[14]The Phonological loop, or articulatory loop, is one of the slave systems in a multi-component working memory system (Baddeley & Hitch, 1974). The system is said to temporarily hold verbal information while retrieving required phonological information from long-term memory. In 1986 Baddeley presented a new phonological loop model with two parts (Osaka, 2002): a phonological short-term store and a subvocal rehearsal mechanism, or articulatory control process.

so than subvocal repetition.

1.3.2.3. Effects of the Processing Orientation

In order to understand the effects of processing orientation, the auditory word priming paradigm must be restated. As mentioned in the previous section (1.2), several auditory word priming studies showed L1's insensitivity to the processing type of the study phase. Although listeners' attention was manipulated in these experiments (e.g., focusing on the sound or meaning of the words), the priming effects were found to be almost equal (Church & Schacter, 1994; Schacter & Church, 1992). However, in several L2 studies, semantic processing at the encoding stage seemed to diminish the priming effects for beginners (Kirsner & Dunn, 1985; Trofimovich & Gatbonton, 2006). This implies that the level of word processing affected the response of L2 learners. As stated previously, the level of word processing affected explicit memory. The following section will address explicit memory as it is relevant to auditory repetition.

1.3.3. Repetition and Word Memory Retrieval

Recognition memory is a subcategory of episodic memory and is therefore categorized as declarative knowledge and explicit memory (Figure 1). People are said to be able to recognize previously encountered items using recognition memory.

It is common when taking an L2 vocabulary quiz to have a sense of having seen a word but not remember its meaning. On the other hand, some students remember not only the meaning of the word but also precisely where it is written in the textbook.

This phenomenon has been explained by the dual-process models[15] which state that recognition memory consists of two processes: familiarity and recollection (Atkinson & Juola, 1974; Diana, Reder, Arndt, & Park, 2006; Jacoby 1996; Mandler, 1980; Wixted, 2007; Yonelinas, 2002). A number of experiments showed dissociation, or at least partial independence, of the two processes (using behavioral methods: Gardiner & Richardson-Klavehn, 2000; Jacoby, 1991; Rugg & Yonelinas, 2003; etc.; or using neuroimaging and neuropsychology data: Skinner & Fernandes, 2007 for review; Waidergoren, Segalowicz, & Gilboa, 2012; Westerberg, Paller, Weintraub, Mesulam, Holdstock, Mayes, & Reber, 2006, etc.). Familiarity has defined as the sense of having encountered the item without remembering any detailed information (i.e., the meaning of the word). It is assumed to be a fast, automatic, stochastic process (Stenberg, Hellman, Johansson, & Rosén, 2009), irrespective of age (Chung & Light, 2009; Light, Prull, La Voie, & Healy, 2000; Yonelinas, 2002), and sensitive to perceptual fluency (Rugg & Yonelinas, 2003). In contrast, recollection has been defined as using detailed information associated with the item (i.e., the meaning of the word and episodic memory). It is said to be a slow, consciously controlled, contextual process, subject to age, and sensitive to elaborative encoding (see the references of familiarity above).

There are some other important frameworks or principles relating to recognition memory, such as the levels-of-processing

[15]There are single-process models in which familiarity is said to be the basis for recognition; however, some proponents admit the necessity of the other process to fully explain the phenomena (Malmberg, Holden, & Shiffrin, 2004).

(LOP) framework, (Craik and Lockhart , 1972), transfer-appropriate processing (TAP) principle (Morris, Bransford, & Franks, 1977), and the encoding specificity principle (Tulving, 1979; Tulving & Thomson,1973), discussed in the previous section (2.1). Experiments manipulating processing levels at the encoding stage found that deep processing affects recollection more than familiarity (Dobbins, Kroll & Yonelinas, 2004; Rugg & Yonelinas, 2003) and since recollection is sensitive to elaborative information and is a consciously controlled process, deep processing is expected to reduce errors in recognition tasks. On the other hand, frequent exposures, or repetition, made familiarity more likely (Hasher & Zacks, 1984; Hasher, Zacks, Rose, & Sanft, 1987) or affected both aspects (Chung & Light, 2009; Hintzman & Curran, 1994; Kelley & Wixted, 2001; Malmberg, Holden, & Shiffrin, 2004). It is plausible that the characteristics of tasks at the encoding stage affected information retrieval, creating mixed results. Since familiarity is sensitive to perceptual fluency, a repetition task mainly building perceptual fluency will yield familiarity-dominant recognition, which will result in fast responses in recognition tasks.

1.4 Summary

This chapter discussed the application of EBM as a linguistic implicit memory model, and suggested that the auditory word priming paradigm may hold an important clue to unveiling the role of implicit memory in language learning, based on empirical studies of auditory priming. In the studies of L2, negative effects could be seen if learners previously engaged in semantic tasks in auditory priming experiments, giving rise to a working theory called the TAP theory. Some methodological issues of previous studies were stated, with a discussion on the need to devise an alternative experimental method that could eliminate the effects of explicit memory, more precisely, episodic memory. Since there is little auditory priming research with Japanese EFL learners, the researcher underscored the importance of rectifying this gap. In addition, the researcher pointed out that speaker variability appears to reduce L2 priming effects because L2 learners are sensitive to perceptual changes and discussed the effects of the use of synthetic speech to supplement L2 input-poor environment of Japanese EFL learners.

The researcher next summarized empirical studies of L2 repetition classified by the effects of number of repetition, method of repetition and processing orientation. The researcher pointed out that learners were able to respond faster as an effect of repetition training, but the effect was limited in the first few times. Moreover, the researcher raised the possibility that vocal repetition had some positive effects on learners' speech processing and memory retrieval as it provides some memory retrieval cues and opportunities for auditory self-perception. Processing orientation was addressed though a discussion of the auditory word priming paradigm and recognition memory, while providing empirical studies for both L1 and L2.

These previous studies provided some significant insights, as well as shed light on some issues that need to be further examined. This led to the formulation of several research questions and hypotheses, followed by the construction of five experiments to be described in the following chapters.

Chapter 2
Purposes and Hypotheses of the Study

The primary goal of this study is to explore the cognitive process of implicit memory, which facilitates language acquisition and learning, and explores its functions in L2 learning in order to provide some pedagogical implication for English language education. Five experiments were conducted to attain this goal. This chapter describes purposes and hypotheses of each experiment derived from the previous studies.

2.1 Purpose and Hypotheses of Experiment 1

The goal of Experiment 1 is to reveal the characteristics of Japanese EFL learners' speech processing in order to illustrate the role of implicit memory in L2 speech perception compared to that of L1. Therefore, in contrast to Sugiura and Hori (2012), this study focuses on the difference in the priming effect through repetition between Japanese EFL learners and native English speakers.

The present study addresses the following two research questions:

(1) Can auditory word priming be seen in Japanese EFL learners' processing of L2 words?

(2) Do different types of processing at the encoding stage affect L2 speech processing in Japanese EFL learners?

The following two hypotheses are proposed:

(1) Auditory word priming can be seen in speech processing of both Japanese EFL learners and native English speakers.

(2) Semantic processing in the encoding stage diminishes the

priming effect in Japanese EFL learners compared to native English speakers.

2.2 Purpose and Hypotheses of Experiment 2

An auditory priming effect has previously been demonstrated in Japanese EFL learners (Sugiura and Hori, 2012). However, as far as the researcher is aware, no such studies have been conducted while taking into account speaker variability and cognitive resource allocation (attention focus while listening to vocabulary words). The researcher posed the following research questions and conducted a priming experiment with the purpose of examining what effect the paralinguistic variation of speaker variability has on L2 speech processing among Japanese EFL learners:

(1) Is L2 speech processing in Japanese EFL learners affected by speaker variability?

(2) Is L2 speech processing in Japanese EFL learners affected by the combination of speaker variability and cognitive resource allocation?

The following hypotheses were formulated based on the previous research:

(1) L2 priming effect will decrease with speaker variability.

(2) If attention is focused on meaning while listening to vocabulary words in the presence of speaker variability, L2 priming effect will decrease.

2.3 Purpose and Hypotheses of Experiment 3

In order to compare learning effects at the perceptual level using both synthetic speech and natural human speech, this study investigated auditory word priming in Japanese EFL learners to answer the following research questions:

(1) Is there a difference in the perceptual learning effects when using natural human speech rather than synthetic speech?

(2) Does the perceptual learning effect change based on the learner's focus when listening to speech?

(3) Does the perceptual learning effect change with the proficiency level of the learner?

The following hypotheses were constructed based on the previous studies on priming and synthetic speech to answer the above research questions:

(1) The perceptual learning effect will be greater with natural human speech than with synthetic speech.

(2) The perceptual learning effect will decrease when using natural human speech and focusing on meaning.

(3) The perceptual learning effect will be greater using natural human speech, but decrease using synthetic speech for learners with high proficiency levels.

2.4 Purpose and Hypotheses of Experiment 4

The objective of this experiment was to investigate the effects of word-based L2 auditory repetition on online speech processing. As mentioned in the previous chapter, this study tried to verify which specific factors might help learners to perceive and produce L2 sounds. Experiment 4 was conducted to examine the following research questions:

(1) Does auditory word repetition change participants' processing of words as in previous repetition studies?

(2) Are the participants' responses affected by two different methods, vocal repetition and subvocal repetition?

(3) Are the participants' responses influenced by two different levels of processing, phonological (nonsemantic) and semantic, in the study phase?

In order to investigate the effects of auditory word repetition, reaction time (RT) and error rate are examined in an online method.

The following hypotheses are proposed:

(1) RTs and error rate will decrease as the number of repetitions increases.

(2) Vocal repetition will accelerate RTs, while decreasing error rate as compared to subvocal repetition.

(3) The nonsemantic task will show shorter RTs and lower error rates compared with the semantic task.

2.5 Purpose and Hypotheses of Experiment 5

The goal of Experiment 5 was to investigate the effects of word-based L2 auditory repetition on offline performance. To examine how auditory word repetition affected retention of presented words, the recognition method was used. Experiment 5 was conducted after Experiment 4. In addition, Experiment 5 investigated the effects of encoding factors, semantic vs. nonsemantic and vocal vs. subvocal repetition, on word memory retrieval to answer the following research questions:

(1) Does the level of processing at the encoding stage affect the accuracy of word memory retrieval?

(2) Does perceptual fluency building affect the speed and

accuracy of word memory retrieval?

The following hypotheses were constructed based on previous studies:

(1) Deep processing at the encoding stage will reduce the error rate in the recognition task.

(2) Repetition will accelerate the recognition of words and above all, vocal repetition will enhance recollection of words.

Chapter 3
Auditory Word Priming Effect in L1 and L2

3.1 Experiment 1

3.1.1 Participants

The participants in this experiment included 48 Japanese EFL learners and 40 native English speakers. The former group consisted of Japanese undergraduate students and graduate students (18 men and 30 women) at a university located in the Kansai area of Japan, and the experiment for them was conducted from April to May 2010. Their ages ranged between 18 and 24, and they were enrolled in different faculties. None of them had spent more than three months in an English speaking country. The average scores of the Test of English for International Communication (the *TOEIC*® Listening and Reading test) reported by the 30 students was 569.06 (*SD* = 154.73), which ranged from 280 to 940. The native English speakers consisted of overseas students[16] (13 men and 27 women) with the length of residence in an English-speaking country as follows: *M* = 20.21, *SD* = 2.17. Like the first group, they were enrolled in different faculties of the same university. Their ages ranged from 18 to 27 and they participated in this experiment in June–July and September–October 2011. All of the participants reported had normal hearing and vision at the time of the experiment.

---

[16]Participants included 33 students from the United States, six from the United Kingdom, and one from Canada.

3.1.2 Materials

Before describing the details of the experiment, the auditory priming research methodology should be discussed. To devise a method that enables participants to process words in a unified manner, a rhyme judgment task and synonym judgment task were conducted in the study phase of the current study, differing from previous studies. This idea was based on research in the related field of cognitive neuropsychology (Tanemura, 2006) and a brain imaging study of L2 word perception (Ishikawa & Ishikawa, 2008). In theory, a rhyme judgment task would facilitate phonological processing, while a synonym judgment task would encourage semantic processing.

The experiment was conducted in two phases—a study phase and a test phase. The former consisted of a rhyme or a synonym judgment task, and the latter a vocal repetition task (all the words used in this experiment are shown in Appendix A to D). The materials used for the study phase comprised of two sets of 48 English words in pairs (see Appendices A and B): Task 1 (rhyme judgment task): V1+V2+V3 and Task 2 (synonym judgment task): V4+V5+V6. For the repetition task in the test phase, two sets of 48 English words were used (see Appendices A, B, C, and D): Task 3 (a vocal repetition task after the rhyme judgment task): V1+V7, Task 5 (a vocal repetition task after the synonym judgment task): V4+V9. The words from V7 and V9 were not presented in the study phase (unrepeated words or new words).

All the words were selected from the English words familiarity database of Japanese EFL learners (Yokokawa, 2006, 2009). Approximately 2,000 university and junior college students were asked to rate the degree of familiarity of each word (roughly 3,000 in total) on a seven-point scale and the mean familiarity of the database was 4.72. Because it was critical to use well-known words for priming experiments (Stark & McClelland, 2000), only words with high familiarity were utilized in this experiment. Word frequency was controlled using the British National Corpus (BNC), and syllable numbers and durations were also considered. The words used in Task 1 and 2 were, on average, 2.37 syllables long, with a mean familiarity of 5.94 and a mean frequency of 26.84 occurrences per thousand words based on BNC. A one-way analysis of variance (ANOVA) was conducted to confirm that there were no differences between the words used in Task 1 and 2 in terms of familiarity ($F[1, 190] = 0.01$, $p > .05$); word frequency ($F[1, 190] = 0.24$, $p > .05$); syllable number ($F[1, 190] = 0.06$, $p > .05$).

The words used in Task3 and 5 were, on average, 2.38 and 2.33 syllables long respectively, with a mean familiarity of 5.93 and 5.92 respectively, and a mean frequency of 21.09 and 29.54 occurrences per thousand words based on the BNC. A one-way analysis of variance (ANOVA) was conducted to confirm that there were no differences between the words used in repetition tasks in terms of familiarity ($F[3, 92] = 0.01$, $p > .05$); word frequency ($F[3, 92] = 0.76$, $p > .05$); syllable number ($F[3, 92] = 0.03$, $p > .05$); and duration ($F[3, 92] = 0.001$, $p > .05$). In addition, the first phonemes of words in each list began with the same phoneme in order to counterbalance the effect of phoneme recognition through a microphone.

The selected English words were recorded using the speech synthesis software, Globalvoice English Version 2 (PENTAX). As previous studies indicate, L2 learners are in general overly-dependent on the minute, context-specific information of spoken L2 words compared to L1 words (Bradlow, Pisoni, Akahane-Yamada & Tohkura, 1997; Goldinger, 1996; Trofimovich, 2005); therefore, the same voice (a female voice called "KATE") was used in both phases. All words were checked prior to their usage in the experiment. "KATE" clearly pronounced each word from the first phoneme to the last.

3.1.3 Procedure

Each participant was tested individually in a quiet location using a personal computer. It took approximately 30 minutes for each participant to complete the experiment, including a 5-minute break. All participants had the option to take a rest between tasks if they felt tired. All words were presented as auditory input via speakers without visual aids. Word presentation was controlled by the *SuperLab 4.5* experimental laboratory software (Cedrus), which showed each word randomly and recorded the correctness of participants' responses.

The experiment began with instructions (in Japanese for Japanese EFL students and in English for native English-speaking students) presented on the screen (Japanese instructions were shown in Figure 2 to 4). In the study phase, the participants listened to 54 pairs of words (48 pairs for the study, as well as 3 pairs for primacy fillers and 3 pairs for recency fillers, in order to reduce the serial position effects of participants' short-term memory).

After the auditory presentation of each pair of words, participants were asked to press the correct key immediately. For the rhyme judgment task, they were asked to judge whether each

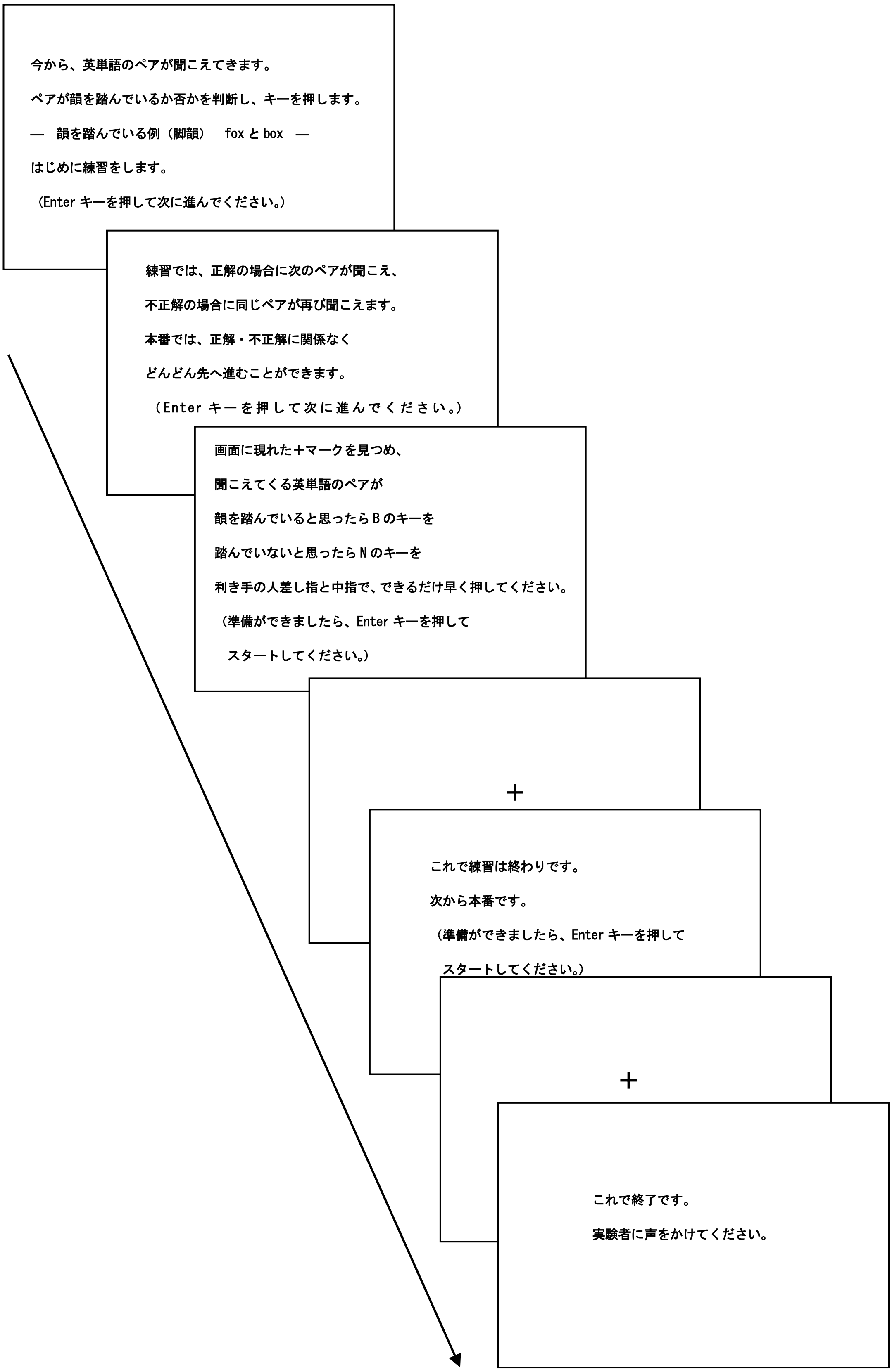

Figure 2. The Japanese instructions for the rhyme judgment task on a computer screen.

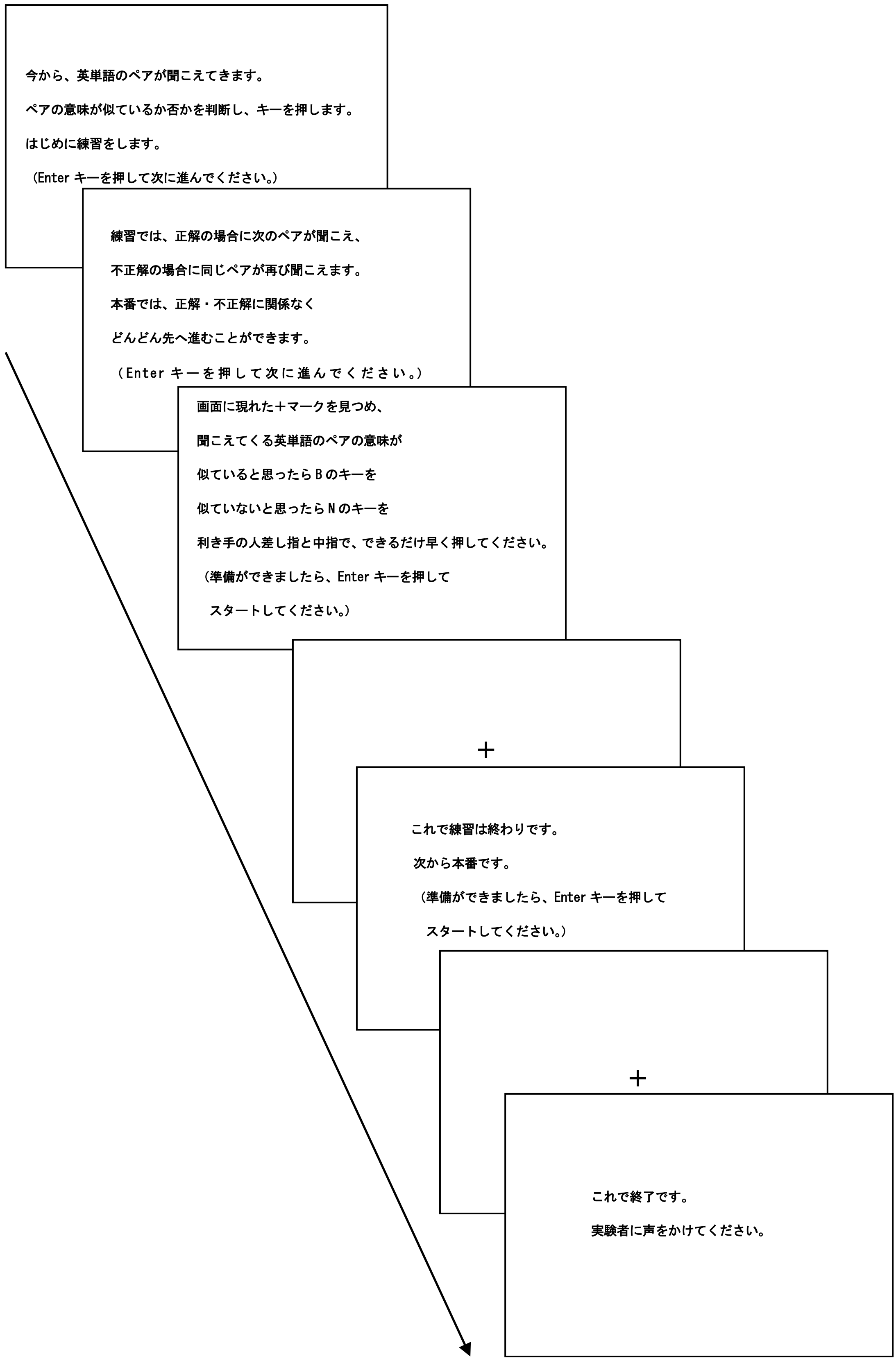

Figure 3. The Japanese instructions for the synonym judgment task on a computer screen.

pair rhymed (Japanese instructions were shown in Figure 2) and for the synonym judgment task, they were asked to judge whether each pair had a similar meaning (Japanese instructions were shown in Figure 3). Following ten practice exercises, participants could start the test at any time by pressing the start key. Feedback was only given for the practice exercises. As soon as they pressed the key, the next pair was presented. After each task they did some simple arithmetic problems called *Hyakumasu Keisan*[17] (Kageyama, 2004) for approximately three minutes on a sheet of paper as a distractor task to erase their short-term memory.

In the test phase, the participants listened to 54 words (48 words for the test, 3 words for primacy fillers and 3 words for recency fillers) presented with a 2000-millisecond inter stimulus interval and were asked to repeat the words aloud as accurately and as rapidly as possible after each study phase (Japanese instructions were shown in Figure 4). Following eight practice exercises, participants could start at any time they wanted to begin each 48-word repetition task by pressing the start key.

The order of tasks (rhyme judgment, synonym judgment) was counterbalanced across the participants in order to eliminate any task order influence.

The responses of participants were recorded using an IC recorder (SONY ICD-SX67) with a condenser microphone (SONY ECM-DS70P).

[17]Hyakumasu Keisan (Hundred-Square Calculations) is a math-drill worksheet that involves addition, subtraction, and other calculations performed on a 10-by-10 grid.

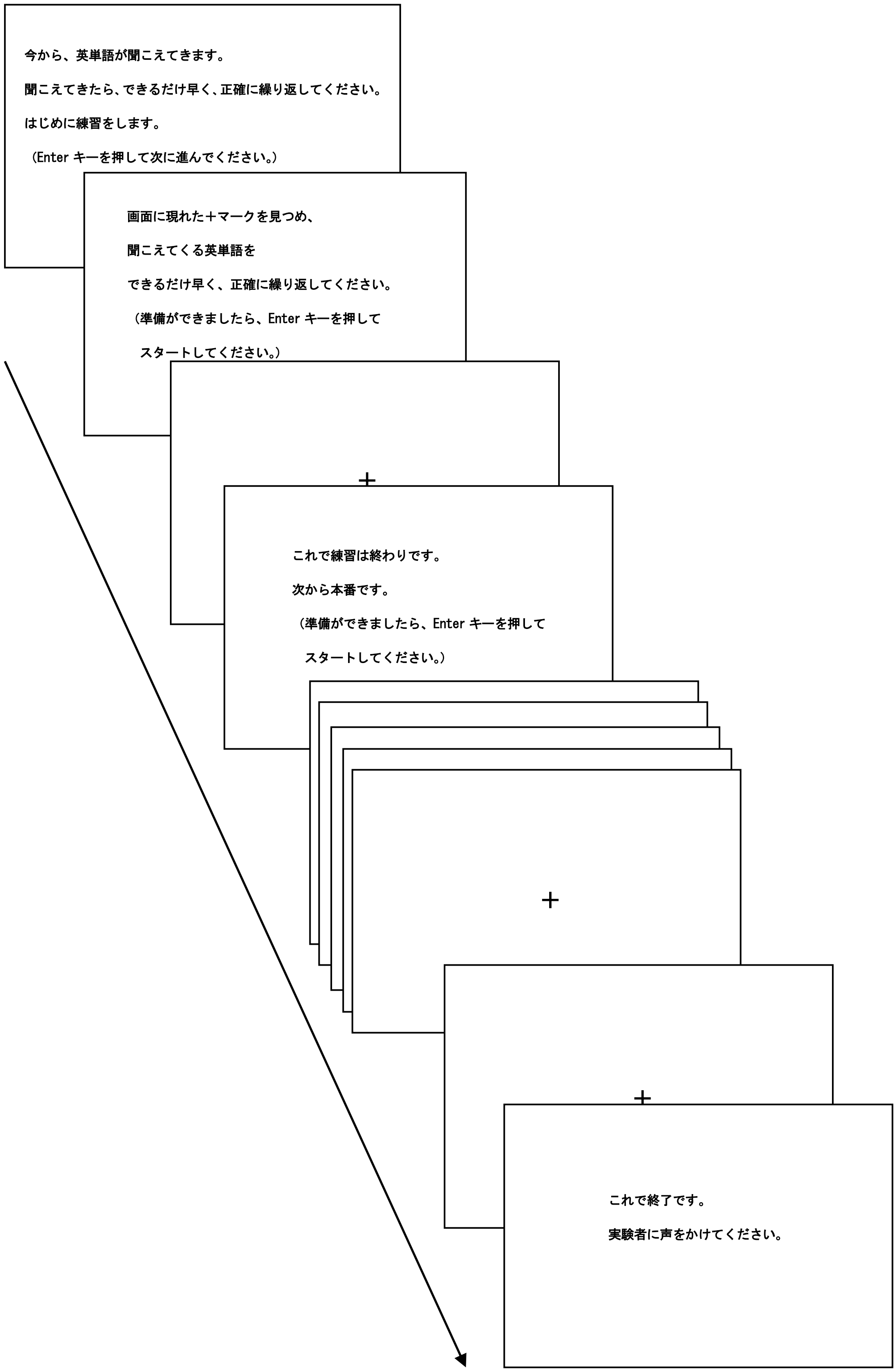

Figure 4. The Japanese instructions for the vocal repetition task on a computer screen.

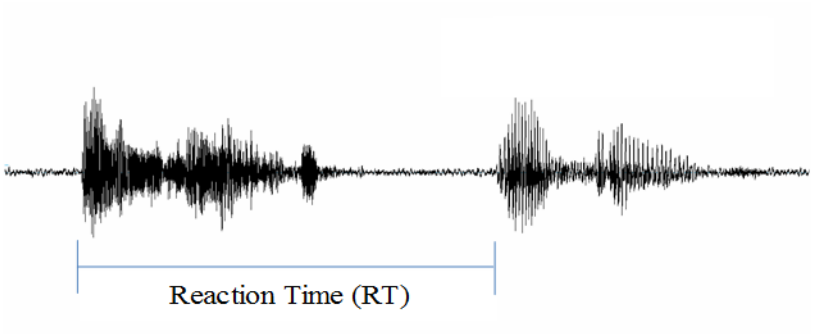

Figure 5. An illustration of RT measure used in a repetition task. The two waveforms are displayed using speech analysis software Praat and represent the stimulus word *restaurant* spoken by a female model sound (left) and repeated by a male participant (right). RT was measured from the onset of a model sound to the onset of the repetition.

3.1.4 Data Analyses

The following section describes the variables used, data preparation and the statistical analyses conducted.

RT was the dependent variable for this experiment, defined as the length of time between the onset of the model sound and the onset of the participants' response in milliseconds (ms) (see Figure 5).

Previous studies measured the RT as the time between the offset of the stimulus and the onset of the participants' response because the measurement was believed to be insensitive to differences in how swiftly stimulus words are pronounced (Trofimovich, 2008). However, in this experiment, the utterance speed of stimuli was controlled with the speech synthesis software. Moreover, some participants repeated words almost simultaneously. Therefore, the above mentioned definition of RT was deemed more appropriate (McDonough & Trofimovich, 2009, p. 43).

Praat software was used to measure RTs (Boersma & Weenink, 2011). All of the incorrect responses were excluded from analyses. Following the definitions of errors in previous studies (Trofimovich, 2005, 2008; Trofimovich & Gatbonton, 2006), an error was defined as the mispronunciation of a word. Moreover, because almost every participant substituted /l/ for /r/, it can be assumed that their intention was more likely to pronounce /r/ even if it was a mispronounced /l/. This particular mispronunciation was not counted as an error according to Trofimovich (2008, p. 317).[18]

To measure the reliability of error identification, a native English speaker (an English teacher at a university in the Kansai area) was asked to serve as a rater of the participants' repetition data. The degree of agreement in the identification of errors between the experimenter and rater was 97.25%. The incorrect data were 5.28% and 1.61% for the Japanese EFL students and the native English speaker students, respectively. Any response that was two standard deviations (*SD*) away from each participant's mean was replaced with the sum of the mean and 2 *SD*s (4.83% for Japanese EFL learners and 4.22% for native English speakers).

The RT data were then run through an analysis of variance (ANOVA). The alpha level for significance was set at .05 for all statistical analyses reported below. In accordance with previous studies, the priming effect was determined by whether there was a significant difference in the RTs for previously encountered vocabulary during the study phase and new vocabulary presented during the test phase.

[18]The participants in Trofimovich (2008) were native Chinese speakers, but the same mispronunciation could be seen in Japanese EFL learners.

3.2 Results of Experiment 1

3.2.1. RT Data

A mixed design 2 (Group) × 2 (Priming) × 2 (Type of Processing) ANOVA was applied with Group (Japanese EFL learners or native English speakers) as a between-subjects factor, while Priming (repeated or unrepeated words) and Type of Processing (form-based or meaning-based processing in the study phase) represented the within-subjects factors.

Table 1 summarizes the mean RTs of the vocal repetition task for the Japanese EFL learners and the native English speakers after the two judgment tasks (rhyme or synonym). Regarding the RT analysis, the main effect of Group was $F(1, 86) = 29.36$, $p < .01$, $\eta_p^2 = .25$ while the main effect of Priming was $F(1, 86) = 72.67$, $p < .01$, $\eta_p^2 = .46$. The interaction between Group and Priming was also significant, $F(1, 86) = 6.32$, $p < .05$, $\eta_p^2 = .07$. No other effect or interaction reached a significant level. A simple main effect test, which was performed to test the interaction effect and results, revealed that both groups (Japanese EFL learners: $F[1, 86] = 60.93$, $p < .01$, $\eta_p^2 = .01$; native English speakers: $F[1, 86] = 18.06$, $p < .01$, $\eta_p^2 = .01$) processed repeated words significantly faster than unrepeated words. This means that both groups showed priming effects. Next, the simple main effect on the groups was tested using a separate error term and the results revealed that the native English speakers were able to repeat words significantly faster than the Japanese EFL learners for both repeated ($F[1, 86] = 27.48$, $p < .01$, $\eta_p^2 = .24$) and unrepeated ($F[1, 86] = 31.01$, $p < .01$, $\eta_p^2 = .26$) words.

Table 1
Mean Reaction Times (ms)

		Rhyme		Synonym	
		M	*SD*	*M*	*SD*
Japanese EFL learners	Unrepeated	722.24	156.67	734.39	154.83
	Repeated	702.67	152.55	708.89	153.83
Native English speakers	Unrepeated	561.28	138.26	560.59	139.11
	Repeated	548.56	139.97	548.77	135.09

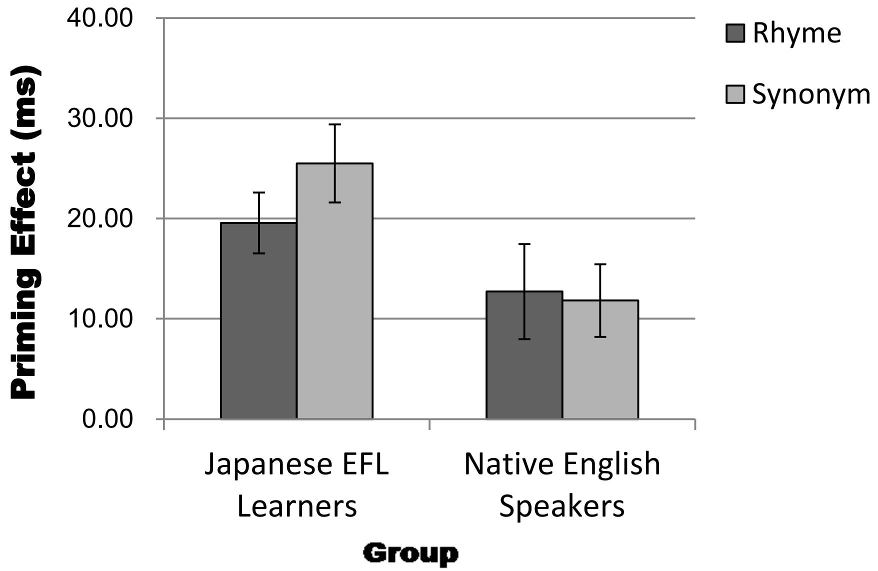

Figure 6. Mean priming effects (milliseconds). The vertical lines indicate standard error.

RT analysis of the judgment tasks used a mixed design 2 (Group) × 2 (Type of Processing) ANOVA, with Group (Japanese EFL students or native English speaker students) as a between-subject factor, while Type of Processing (the rhyme judgment task or the synonym judgment task) represented the within-subject factors. The main effect of Group was $F(1, 86) = 17.74$, $p < .01$, $\eta_p^2 = .17$, while the main effect of Type of Processing was $F(1, 86) = 65.87$, $p < .01$, $\eta_p^2 = .43$. The interaction between Group and Priming was not statistically significant. The results were statistically significant, showing that both groups judged the rhyme judgment task faster than the synonym judgment task. There was also a statistically significant difference indicating that the native English speakers were able to judge tasks faster than the Japanese EFL learners.

3.2.2. Priming Effect

The analysis of RTs revealed that both groups processed repeated words significantly faster than unrepeated words, which means that both groups showed priming effects. The priming effect of each participant was calculated (RTs for repeated words were subtracted from RTs for unrepeated words) and the mean priming effects of both groups were plotted on a graph, shown in Figure 6. To examine group differences under the Types of Processing separately, a one-way ANOVA with Group as a between-participant variable was conducted. The main effect of Group was not significant under the form-based processing condition, but the priming effect was significantly greater for the Japanese EFL learners compared to the native English speakers under the meaning-based processing condition ($F = 6.09$, $p < .05$, $\eta_p^2 = .07$).

The results indicate that auditory word priming can be seen in the word processing of both the L1 and L2 groups, although a larger priming effect could be observed in the L2 groups when meaning-based processing preceded the repetition task.

3.3 Discussion of Experiment 1

Based on previous studies, two research questions and hypotheses were suggested. First, it was hypothesized that auditory word priming can be seen in the speech processing of both Japanese EFL learners and native English speakers. The results supported the hypothesis since both Japanese EFL learners and native English speakers showed evidence of the priming effect, although native English speakers reacted significantly faster for both repeated and unrepeated words. These results support the claim that auditory word priming is in fact a mechanism that aids spoken-word processing of both L1 and L2 learners.

The second research question investigated the influence of different types of processing during the encoding stage. It was hypothesized that semantic processing during the encoding stage diminishes the priming effect of Japanese EFL learners compared to the effect on native English speakers. In contrast, the L2 learners' priming effect was significantly greater than that of the L1 speakers only under the semantic condition. A possible explanation for this is that semantic processing positively influences Japanese EFL learners' sensitivity to phonological information in comparison to that of native English speakers.

With the complete PRS (perceptual representation system), native English speakers were able to access the meaning of the words as soon as they heard the sounds, and because the process was automatic, the processing orientation (semantic or nonsemantic tasks in the study phase) did not affect their word processing. On the other hand, with an incomplete PRS, Japanese EFL learners may not be able to fully and automatically access the semantic features of words when they heard the sounds of the words. After processing the meaning of the words, they could easily access (if not fully) and use phonological information sufficiently to process the spoken words at the perceptual level. Although Japanese EFL learners also showed insensitivity to encoding manipulation, the gap of the decrease in RTs for repeated words in the two groups might show some type of learning effect. The phenomenon that we have seen here may be a process regarding how learners proceduralize L2 decoding (skill learning related to procedural memory), and the meaning-based processing might have an important implication in terms of learning potential.

3.4 Conclusion of Experiment 1

Before concluding, the limitations of the present study need to be addressed. First, this study followed the methods of previous research in order to identify errors. Other forms of mispronunciation should be considered in more detail in future investigation. Second, it is still unclear whether the priming effect in Japanese EFL learners is influenced by the learners' proficiency in English. As mentioned earlier, semantic tasks had a negative effect on "inexperienced" L2 learners (Kirsner & Dunn, 1985; Trofimovich & Gatbonton, 2006). Future studies should examine the potential influence of proficiency levels on the priming effect in EFL students. In addition, further empirical research (e.g., the effects of other influential factors such as contextual details or amount of L2 exposure) will be necessary to provide a better understanding of L2 auditory word priming.

The first implication of this study is that auditory word priming appears to be part of spoken-word processing for learners. Japanese EFL learners typically add phonological features to words whose meanings they already understand. Repeated exposure to the input, even if L2 learners already understand the meanings, reduces RTs, indicating that L2 learners may be acquiring perceptual skills through experience. This priming experiment also seems to suggest the transition of skill learning into automatization driven by spoken input.

The second implication of this study is that because semantic processing positively influences Japanese EFL learners' sensitivity to phonological information, the encoding process, which occurs at the time of initial exposure to spoken words, should be carefully considered for Japanese EFL learners. Such consideration can result in greater learning gains because the role of L2 learners' input in a learning context might have profound significance.

Chapter 4
Auditory Word Priming Effect in Talker Variability

4.1 Experiment 2

4.1.1. Participants

Participants in this study were 40 Japanese learners of English (undergraduate and graduate students). Table 2 shows data on the participants gathered from the questionnaire (Appendix E).

Table 2
English Learning Background of Participants

	M	*SD*
The Oxford Quick Placement Test	30.95	7.14
The Oxford Placement Test (Listening Test)	74.67	8.35
Age starting English study	11.41	1.72
Years of formal English education	11.92	1.33
Years of residence in English-speaking countries	0.20	0.96
Age	20.90	2.67
Self-ratings[a]	4.51	1.22
Listening	4.56	1.69
Speaking	3.92	1.38
Reading	4.97	1.39
Writing	4.56	1.48

Note. *SD* = standard deviation. [a]Ratings scored on a 10-point scale with 1 = minimum proficiency and 10 = near-native proficiency. *N* = 40.

4.1.2. Materials

Vocabulary groups, revised from Experiment 1, were created based on controlled familiarity of spoken words (Yokokawa et al. 2009; F [7, 136] = 0.33, p = .94, η_p^2 = .02 [Spoken]), familiarity of written words (Yokokawa et al. 2006; F [7, 136] = 0.81, p = .58, η_p^2 = .04), frequency (British National Corpus or BNC; F [7, 136] = 0.67, p = .70, η_p^2 = .03), number of syllables (F [7, 136] = 0.06, p = .99, η_p^2 = .003), duration (F [7, 136] = 0.06, p = .99, η_p^2 = .003), and initial consonants (initial consonants of six out of 36 pairs were uncontrolled)(see Appendices F to I).

The researcher requested assistance from two native English speakers (one male and one female, citizens of the United States who were both working as English teachers in Japanese universities). Their voices were monaurally digitally recorded in a sound booth using an IC recorder (SONY ICD-SX67) with a microphone (SONY ECM-DS70P) at a sampling rate of 48Hz and quantization at 16 bits. The following method was used to set the speed of speech, a known major factor from previous studies that influences intelligibility and comprehensibility. The English teachers were provided with sounds for the vocabulary group created using the Globalvoice English Professional version 2.0.1 (HOYA) and were asked to repeat each word twice while being conscious of speed. All words were checked by the researcher and if the speech rates were different, Praat was used to match them. The researcher also ensured similar volumes for the speech using Praat (mean 73.0 dB).

4.1.3. Procedure

The experimental procedure was the same in Experiment 1, with one exception. When using the new vocabulary groups added in this study, voice was changed from the study to the test phase (Female → Male, Male → Female), as seen in Appendices F to I.

The researcher also investigated the effect of attention, whether the focus was on the sound (rhyme judgment task) or the meaning during the study phase (synonym judgment task) using 18 pairs of words for each task, and of voice variability in the test phase. Two evaluators, including the experimenter, selected errors from the RT data (the concordance rate of error evaluations was 94.66%). Incorrect data, 5.10% of the results, were excluded from the analysis.

4.1.4. Data Analyses

The experimental procedure was the same in Experiment 1. The RT was used as the dependent variable in the current experiment and an analysis of variance (ANOVA) was conducted to investigate the priming effects.

4.2 Results of Experiment 2

Table 3 and Figure 7 show the RT data and the priming effect. RT data of the repetition task were submitted to 2 (Task) × 3 (Voice Variability) repeated measures ANOVA, with Task (rhyme judgment task or synonym judgment task in the study phase) and Voice Variability (repeated words in the same voice or different voice in the test phase, or unrepeated words) as within-subjects factors. Post-hoc comparisons were conducted using the procedure from Benjamini and Hochberg (1995). This analysis revealed a significant interaction between Task and Voice, $F(2, 78) = 30.42$, $p < .001$, $\eta_p^2 = .44$, *power* = 1.00.

The simple main effect of Task was significant for the same voice, $F(1, 39) = 8.16$, *adjusted* $p = .02$, $\eta_p^2 = .17$, but not for different voices, $F(1, 39) = 1.34$, *adjusted* $p = .32$, $\eta_p^2 = .03$, nor for unrepeated words, $F(1, 39) = 0.19$, *adjusted* $p = .66$, $\eta_p^2 = .005$. The participants could repeat the words significantly faster when rhyme judgement was conducted in the study phase and the words were presented in the same voice.

The simple main effect of Voice Variability was significant for both the rhyme judgment task, $F(2, 78) = 50.83$, *adjusted* $p < .001$, $\eta_p^2 = .57$, and the synonym judgment task, $F(2, 78) = 4.87$, *adjusted* $p = .02$, $\eta_p^2 = .11$. Under the rhyme judgement task condition, the post-hoc test of Voice showed that the RT data of the same voice and a different voice, as well as the same voice and unrepeated words, were significantly different (*adjusted ps* < .001, < .001, respectively). However, a different voice and unrepeated words did not show statistically significant differences (*adjusted* $p = .58$). Under the synonym judgement task condition, the post-hoc test of Voice Variability revealed that the RT data of a different voice and unrepeated words were significantly different (*adjusted* $p = .02$), while the same voice and a different voice, as well as the same voice and unrepeated words, did not show statistically significant difference (*adjusted ps* = .28, = .06, respectively). The rhyme judgment task in the study phase revealed a priming effect only for words repeated in the same voice. On the other hand, the synonym judgment task conducted in the study phase indicated a priming effect only for words repeated in a different voice. Importantly, there were no statistically significant differences between the RT data of the same and different voices.

Table 3
Mean Reaction Times (ms)

	Rhyme		Synonym	
	M	*SD*	*M*	*SD*
Unrepeated	836.44	106.70	840.26	104.50
Repeated				
Same voice	806.70	112.70	832.53	109.53
Different voice	838.29	105.93	828.07	107.60

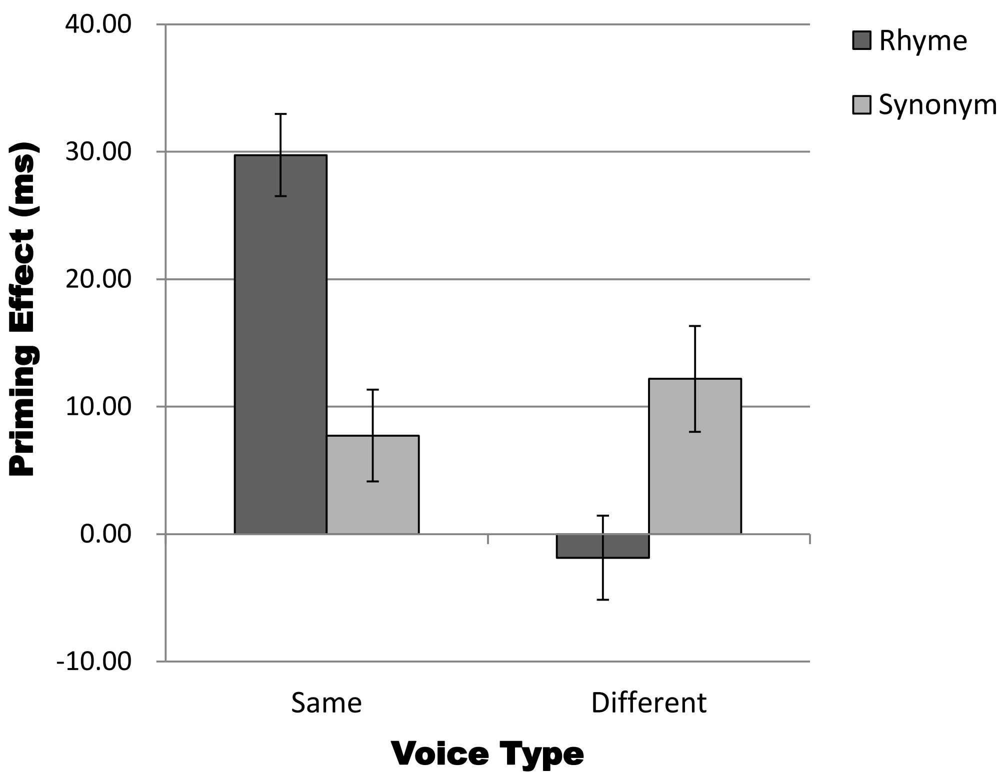

Figure 7. Mean priming effects (ms). The vertical lines indicate standard error.

4.3 Discussion of Experiment 2

The results of the experiment only partially supported Hypothesis 1 (L2 priming effect will decrease with speaker variability). When participants listened to vocabulary words and focused on sound, the effect of speaker variability was statistically significant and a negative priming effect was seen. This appears to be a reasonable result considering the fact that most learners of English in Japan learn in environments with little exposure to spoken English and that previous studies have had comparable results. However, when the participants focused on meaning, there were no statistically significant differences in RTs, with or without speaker variability. Furthermore, contrary to Hypothesis 2 (if attention is focused on meaning while listening to vocabulary words in the presence of speaker variability, L2 priming effect will decrease), when participants focused on meaning with speaker variability, a statistically significant priming effect was seen, whereas with no speaker variability, there was a trend toward statistical significance (*adjusted* p = .06).

One might conclude that the reason for these results is that because this experiment had less variation in speech than previous studies (six speakers in the previous studies compared with two in the present study), the participants may have grown accustomed to the speakers' voices, resulting in a priming effect. However, if this were the case, the same results would have occurred when the participants focused on sound. Further, a close examination of the mean RTs for each item showed that there was no particular one to which the participants had responded with notable speed. Considering the fact that no statistically significant differences in

RTs were seen with a focus on meaning, regardless of whether or not the speaker remained the same, the likelihood is high that the reason must lie in the content of the tasks.

In the previous studies, the meaning-focused task asked participants to rate how fun the words were. This study used a synonym judging task asking participants to determine whether two consecutive words they hear had similar meanings or not. One can imagine that there may have been some L1 interference. While there is also a high likelihood that L1 translation was involved in order to rate pleasantness of a word, comparing two words requires more elaboration, which may have strengthened the related memory traces. However, that alone would not suffice to explain why a priming effect was seen, regardless of speaker variability. The experimenter would like to propose the possibility that, as shown in L1 speech perception development research, attention to paralinguistic information (i.e., the emotions, age, sex and dialect of the speaker) may have been inhibited because it was not part of the semantic information essential to process spoken language. Because the cognitive load related to the meaning-focused task in this study was greater than that in the previous studies, the allocation of more cognitive resources to process semantic information may have reduced attention to paralinguistic information. As a result, participants may have been less susceptible to any effects from variation in it. Further, if participants were less susceptible to the acoustic aspects of speech, there is a possibility that top-down speech perception occurred during the test period.

4.4 Conclusion of Experiment 2

This study had some limitations, namely the lack of a wide variety of voices and inability to observe the longitudinal effects of paralinguistic information. This is something that the researcher hopes to address through future research.

If we look only at the priming effect when participants focused on the sounds of words with speaker variability, one could suggest that Japanese EFL learners were unable to independently process paralinguistic information from linguistic information. However, the perceptual learning effect seen in priming experiments is overwhelmingly high when speakers are the same, suggesting the potential presence of a speaker adaptation mechanism. If a cognitive database of L2 phonological information could be built through the accumulation of such examples, it would suggest that the *Exemplar Model* would more easily explain these phenomena. The results of this experiment indicate the possibility that attention to L2 paralinguistic information is inhibited in situations where it is necessary to allocate large amounts of cognitive resources to semantic information. This is a phenomenon seen during the development of L1 speech perception in children and it implies the existence of a common system across L1 and L2 for the allocation of the majority of cognitive resources to the essential part of spoken language, which is meaning.

Chapter 5
Auditory Word Priming Effect in Natural Human Speech and Synthetic Speech

5.1 Experiment 3

5.1.1. Participants

The Participants in this study were 80 Japanese learners of English (undergraduate and graduate students). Forty of the participants were drawn from the same pool as in Experiment 2. They were divided into two equal groups. One group of the participants participated in the experiment that used synthetic speech and the other participated in the experiment that used recorded natural human speech (Experiment 2). The participants in these two experiments did not overlap. A summary of the participants' characteristics is presented in Table 4. Based on the correct answer rate of the Oxford Quick Placement Test (Oxford University Press et al., 2001) (60 points maximum) and the Oxford Placement Test (Listening Test) (Allan, 2004) (100 points maximum), there was no significant difference in the proficiency of these two groups, $F(1, 78) = 1.84$, $p = .17$, $\eta_p^2 = .02$.

5.1.2. Materials

Vocabulary groups partly overlapped with the vocabulary group used in Experiment 2 (see Appendices F and G). They were created based on controlled familiarity (Yokokawa et al. 2006, 2009), $F(3, 68) = 0.19$, $p = .90$, $\eta_p^2 = .008$ (Spoken) and $F(3, 68) =$

Table 4
English Learning Background of Participants

	Natural (n = 40)		Synthetic (n = 40)	
	M	*SD*	*M*	*SD*
The Oxford Quick Placement Test	30.95	7.14	30.05	5.41
The Oxford Placement Test (Listening Test)	74.67	8.35	71.13	5.60
Age starting English study	11.41	1.72	10.85	2.40
Years of formal English education	11.92	1.33	11.98	1.05
Years of residence in English-speaking countries	0.20	0.96	0.18	0.78
Age	20.90	2.67	19.08	0.80
Self-ratings[a]	4.51	1.22	4.94	1.09
Listening	4.56	1.69	4.65	1.51
Speaking	3.92	1.38	4.35	1.51
Reading	4.97	1.39	5.38	1.39
Writing	4.56	1.48	4.78	1.56

Note. *SD* = standard deviation. [a]Ratings scored on a 10-point scale with 1 = minimum proficiency and 10 = near-native proficiency.

0.60, p = .61, η_p^2 = .03 (Written), frequency (British National Corpus or BNC), $F(3, 68)$ = 0.86, p = .46, η_p^2 = .04, number of syllables, $F(3, 68)$ = 0.06, p = .98, η_p^2 = .003, duration, $F(3, 68)$ = 0.03, p = .99, η_p^2 = .001, and initial consonants (the initial consonants of three pairs were uncontrolled) using synthesized TTS speech or natural human speech from native English speakers.

The Globalvoice English Professional version 2.0.1 (HOYA) was used as the speech synthesis software with "KATE" as the female voice and "PAUL" as the male voice. As stated in the previous chapter, the voices of two native English speakers were recorded for use as natural human speech (see Chapter 5 for details).

5.1.3. Procedure

The experiment was conducted in the same way as Experiment 2.

5.1.4. Data Analyses

The same data analyses as in Experiment 2 were conducted. The researcher considered whether the focus was on the sound or the meaning during the study phase and investigated how using synthetic speech was different from using recorded natural human speech. In addition, a proficiency-based analysis was added. Following the example set by previous studies (Trofimovich, 2008), two evaluators, including the researcher of this paper, selected repetitive data that were deemed errors (the concordance rate of error evaluations was 96.88%). The incorrect data, 4.51% of the data, was excluded from the analysis. The RT data that was two standard deviations (*SD*) away from each participant's mean was substituted by the sum of the mean and 2*SD*s.

5.2 Results of Experiment 3

5.2.1. Learning Effect when using Natural Human Speech and Synthetic Speech

Table 5 shows the mean RTs of the repetition task when using natural human speech and synthetic speech. A mixed design 2 (Speech) × 2 (Task) × 2 (Repetition) ANOVA was conducted with Speech (recorded natural human speech or synthetic speech) as a between-subject factor, while Task (rhyme judgment task or synonym judgment task in the study phase) and Repetition (whether the vocabulary was heard once before during the study phase) as within-subject factors. The results of the three-way ANOVA showed a significant three-way interaction, $F(1, 78) = 18.58$, $p < .001$, $\eta_p^2 = .19$. The simple interaction effect was investigated using a mixed design 2 (Speech) × 2 (Repetition) ANOVA for each Task. There was a significant interaction between Speech and Repetition when the rhyme judgment task was conducted during the study phase, $F(1, 78) = 11.24$, $p = .001$, $\eta_p^2 = .13$. The results of the simple main effect test revealed a repetition effect for both natural human speech, $F(1, 78) = 76.56$, $p < .001$, $\eta_p^2 = .50$, and synthetic speech, $F(1, 78) = 16.07$, $p < .001$, $\eta_p^2 = .17$. There was also a significant interaction between Speech and Repetition when the synonym judgment task was conducted during the study phase, $F(1, 78) = 9.45$, $p = .003$, $\eta_p^2 = .11$. The results of the simple main effect test revealed a repetition effect for synthetic speech, $F(1, 78) = 41.05$, $p < .001$, $\eta_p^2 = .34$, but not for

Table 5
Mean Reaction Times (ms) of the Repetition Task

	Natural				Synthetic			
	Rhyme		Synonym		Rhyme		Synonym	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unrepeated	836.44	106.70	840.26	104.50	824.40	148.81	835.65	145.27
Repeated	806.70	112.70	832.53	109.53	810.60	142.87	811.75	149.73

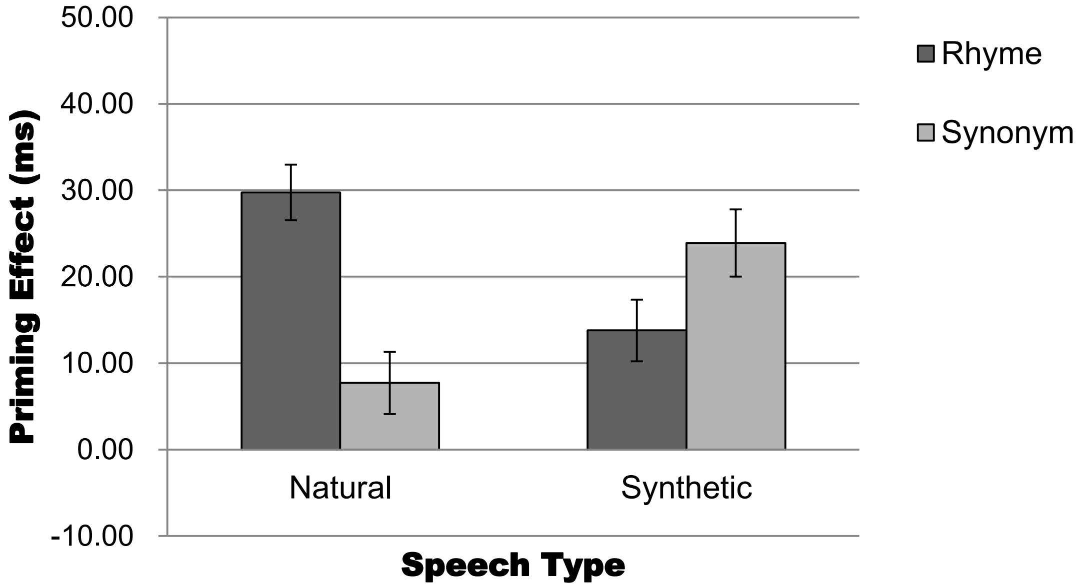

Figure 8. Mean priming effects (ms). The vertical lines indicate standard error.

natural human speech, $F(1, 78) = 4.24$, $p = .09$, $\eta_p^2 = .05$.

The results show a positive priming effect both when a real human voice was used and when synthetic speech was used. However, the statistical results and Figure 8 show that the size of the priming effect changed for the combination of Speech and Task. The perceptual learning effect was greater with recorded human speech when participants focused on the sound of the vocabulary while it was greater with synthetic speech when participants focused on the meaning.

5.2.2. Proficiency-Based Analysis

To measure the effects of proficiency, the researcher classified these learners into three groups based on the results of the proficiency tests, discarding the middle group and comparing the upper- and lower-proficiency groups. The data of the upper- and lower-proficiency groups is shown in Table 6 (natural human speech) and Table 7 (synthetic speech). Significant differences are revealed in the scores for both the Oxford Quick Placement Test and the Oxford Placement Test (Listening Test) at each proficiency level.

Table 6

Two Proficiency Groups and Their English Learning Background When Using Natural Human Speech

	Lower (n = 12)		Upper (n = 12)		F	p	η_p2
	M	SD	M	SD			
The Oxford Quick Placement Test	27.17	3.49	33.83	6.55	9.69	.005	.306
The Oxford Placement Test (Listening Test)	68.91	6.01	73.42	3.63	4.84	.039	.187
Age starting English study	9.92	3.34	11.75	1.14	3.23	.086	.128
Years of formal English education	11.92	1.62	12.08	0.51	0.12	.738	.005
Years of residence in English-speaking countries	0.59	1.37	0.01	0.03	2.14	.158	.089
Age	18.83	0.58	19.33	0.89	2.68	.116	.108
Self-ratings[a]	4.73	1.40	5.25	1.02	1.09	.308	.047
Listening	4.83	1.85	4.50	1.45	0.24	.628	.011
Speaking	4.58	1.62	4.50	1.62	0.02	.901	.001
Reading	4.83	1.59	6.50	1.00	9.48	.006	.301
Writing	4.67	1.78	5.50	0.90	2.10	.162	.087

Note. SD = standard deviation. [a]Ratings scored on a 10-point scale with 1 = minimum proficiency and 10 = near-native proficiency.

Table 7

Two Proficiency Groups and Their English Learning Background When Using Synthetic Speech

	Lower (n = 12)		Upper (n = 12)		F	p	$\eta_p 2$
	M	SD	M	SD			
The Oxford Quick Placement Test	24.58	2.68	38.08	4.85	71.21	< .001	.764
The Oxford Placement Test (Listening Test)	68.00	7.41	76.33	7.36	7.63	.011	.258
Age starting English study	11.42	1.24	11.58	2.11	0.06	.816	.003
Years of formal English education	12.00	0.85	12.33	0.65	1.16	.294	.050
Years of residence in English-speaking countries	0.00	0.01	0.14	0.28	3.00	.097	.120
Age	20.17	0.83	22.33	4.19	3.09	.093	.123
Self-ratings[a]	4.04	1.20	4.90	1.44	2.50	.128	.102
Listening	3.58	1.51	5.08	1.78	4.96	.036	.184
Speaking	3.92	1.31	4.00	1.81	0.02	.898	.001
Reading	4.42	1.44	5.67	1.30	4.96	.037	.184
Writing	4.25	1.48	4.83	1.47	0.94	.344	.041

Note. SD = standard deviation. [a]Ratings scored on a 10-point scale with 1 = minimum proficiency and 10 = near-native proficiency.

5.2.2.1. Effect of proficiency of learners when using natural human speech

Tables 8 shows the mean RTs of the repetition task when using natural human speech. A mixed design 2 (Proficiency) × 2 (Task) × 2 (Repetition) ANOVA was conducted with Proficiency (the upper- or lower-proficiency) as a between-subject factor, while Task (rhyme judgment task or synonym judgment task in the study phase) and Repetition (whether the vocabulary was heard once before during the study phase) as within-subject factors. Using recorded natural human speech, the results showed a significant main effect of Proficiency, $F(1, 22) = 5.47$ $p = .03$, $\eta_p^2 = .20$, and a significant interaction between Task and Repetition, $F(1, 22) = 19.42$, $p < .001$, $\eta_p^2 = .47$. The results of the simple main effect test showed a significant main effect only when the rhyme judgment task was conducted during the study phase, $F(1, 22) = 52.68$, $p < .001$, $\eta_p^2 = .71$, and not when the synonym judgment task was conducted during the study phase, $F(1, 22) = 2.27$, $p = .15$, $\eta_p^2 = .09$. Figure 9 shows that though there was a perceptual learning effect if participants focused the perceptual dimension of words, the same effect was not present if they focused on the meaning of the words, regardless of proficiency when the tasks used natural human speech.

Table 8
Mean Reaction Times (ms) of the Repetition Task of Two Proficiency Groups When Using Natural Human Speech

	Lower (*n* = 12)				Upper (*n* = 12)			
	Rhyme		Synonym		Rhyme		Synonym	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unrepeated	837.09	93.40	862.22	66.85	777.74	92.54	762.13	103.45
Repeated	810.79	87.62	859.54	70.04	741.95	100.75	751.49	118.44

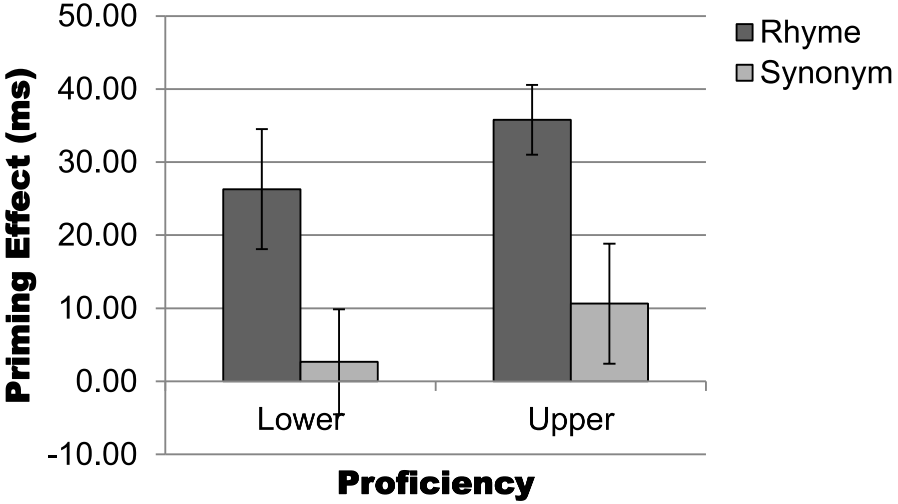

Figure 9. Mean priming effects (ms) of two proficiency groups when using natural human speech. The vertical lines indicate standard error.

5.2.2.2. Effect of proficiency of learners when using synthetic speech

Tables 9 shows the average RT of the repetition task when using synthetic speech. The same three-way ANOVA was conducted and the results showed significant main effects for Proficiency, $F(1, 22) = 22.90$, $p < .001$, $\eta_p^2 = .51$, and Repetition, $F(1, 22) = 38.25$, $p < .001$, $\eta_p^2 = .61$.

To verify hypotheses 2 and 3, a mixed 2 (Task) × 2 (Repetition) ANOVA was conducted on the data of the lower and upper proficiency groups. The results of the lower proficiency group showed a significant main effect of Repetition, $F(1, 11) = 24.58$, $p < .001$, $\eta_p^2 = .69$. The results of the upper group showed a significant interaction between Task and Repetition, $F(1, 11) = 6.74$, $p = .02$, $\eta_p^2 = .38$. The results of the simple main effect test showed a repetition effect for the synonym judgment tasks during the study phase, $F(1, 11) = 15.72$, $p = .009$, $\eta_p^2 = .59$, but not for the rhyme judgment tasks during the study phase, $F(1, 11) = 2.23$, $p = .33$, $\eta_p^2 = .17$.

The rhyme judgment task was meant to facilitate perceptual learning because the task required attention to a perceptual dimension of words. However, the statistical results and Figure 10 revealed a smaller perceptual learning effect for upper proficiency learners compared to that of lower proficiency learners. Contrarily, for the synonym judgment task in this study, which required attention to a higher-level dimension, a semantic dimension of words, a learning effect was apparent in the data for both upper and lower proficiency learners. However, the priming effect can be seen when proficiency is low, regardless of the task in

Table 9

Mean Reaction Times (ms) of the Repetition Task of Two Proficiency Groups When Using Synthetic Speech

	Lower (*n* = 12)				Upper (*n* = 12)			
	Rhyme		Synonym		Rhyme		Synonym	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Unrepeated	931.50	121.54	943.88	112.07	718.02	107.15	717.39	119.36
Repeated	912.49	107.17	928.98	125.82	710.98	109.27	688.82	124.31

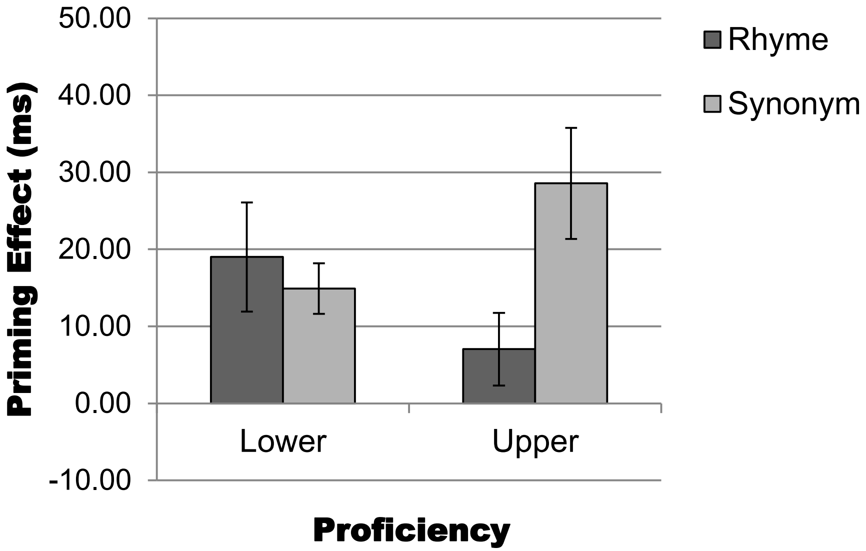

Figure 10. Mean priming effects (ms) of two proficiency groups whenusing synthetic speech. The vertical lines indicate standard error.

the study phase.

5.3 Discussion of Experiment 3

The researcher suggests that if synthesizing TTS speech software is effective in perceptual learning training, it can be used to freely create speech to raise the proficiency of listening skills of Japanese students learning English as a second language. The researcher therefore examined how the auditory learning effect from synthesized TTS differs from that of natural human speech based on three hypotheses in line with three research questions.

5.3.1. Natural Human Speech versus Synthetic Speech

There is a high possibility that the results of the priming effect from natural human speech and synthetic speech support hypothesis 1 (the perceptual learning effect will be greater with natural human speech than with synthetic speech). When participants focused more on the perceptual dimension, the results coincided with the results of previous studies on intelligibility and comprehensibility of synthetic speech processed in L1 or L2. In other words, hypothesis 1 is supported only when participants focused more on the perceptual dimension. Therefore, task differences in the study phase should be considered regarding hypothesis 2.

5.3.2. Task Differences in the Study Phase

Next, this study verified that hypothesis 2 (the perceptual learning effect will decrease when using natural human speech and focusing on meaning) is in accordance with previous studies. This result confirms the complexity of the components of human speech. In addition, since this effect is unrelated to proficiency levels, it suggests that even Japanese learners of English with a higher language proficiency can have a processing delay in the perceptual stages of learning when communicating in real time, which requires a focus on meaning. This effect is even more pronounced in learners with lower language proficiency, suggesting a greater necessity for perceptual learning training.

5.3.3. The Effect of Proficiency Levels on Research Outcomes

Finally, considering the proficiency-based analyses, as in hypothesis 3 (the perceptual learning effect will be greater using natural human speech but decrease using synthetic speech for learners with high proficiency levels), this study confirmed that the perceptual learning effect with synthetic speech was lower in learners with higher proficiency when focusing on the sound, but higher when focusing on the meaning.

At first glance, this may seem to contradict the previous L2 priming studies. However, the previous priming studies used natural human speech, whereas this study also includes synthetic speech. Since the components of synthetic speech are more controlled, such as steady reading speed and regular segmentation as previously mentioned, there are some unnatural features (i.e., constant reading speed and regular segmentation) in the speech. Because of the unnatural features, the perceptual learning effects of synthetic speech when focusing on sound might be diminished for learners with an upper proficiency level. If it were difficult for them to become accustomed to the synthetic speech because of the unnaturalness, the mean RTs of the repetition task when using synthetic speech should be longer than the mean RTs of the repetition task with natural human speech (see the mean RTs of Table 5 & 6). However, the data suggest that they easily become accustomed to synthetic speech, in contrast with lower proficiency level learners, and respond faster to the task. If the task is too easy, there will be no priming effect because there is no learning gain, as we saw in this study.[19]

On the other hand, we can see the perceptual learning effect when focusing on meaning. The logical inference is that the semantic process facilitates the perceptual learning of students with upper proficiency levels if they do not have any problems with the stages of phonetic perception (i.e., if they can easily capture the sound of the word).

Interestingly, the priming data from the lower-proficiency group reveals that some sort of perceptual learning was facilitated regardless of the tasks in the study phase. Considering the trends revealed in previous studies that low proficiency learners prefer synthetic speech (Hirai & O'ki, 2011), this result suggests that using synthetic speech to facilitate perceptual learning for learners in the early stages of learning may be more successful. In light of the results of this study, synthetic speech may be able to play a larger role in environments where there are fewer chances of encountering L2 naturally outside of the classroom as long as the purpose, proficiency, and task combinations are taken into

---

[19]Experiment 1 in Chapter 3 also verified it because native English speakers showed smaller priming effects compared to that of Japanese EFL learners.

consideration.

Though further evidence is necessary, these results suggest the need to incorporate synthesized TTS into second-language learning while taking into consideration that the perceptual learning effect will vary based on the proficiency level of the students.

5.4 Conclusion of Experiment 3

Experiment 3 investigated auditory word priming to demonstrate the applicability of using synthetic TTS in perceptual learning training. The results of the instant study suggest that applying synthetic TTS to perceptual learning training may be effective among Japanese students with lower proficiency levels in English provided they are placed in an environment where there are few daily interactions with L2. The lower proficiency learners demonstrated a higher possibility that synthetic TTS could be a positive contribution to perceptual learning, especially for learners in the early stages of L2 development. In the future, it will be possible to create a rich learning environment for better language acquisition by incorporating this technique into training, such as shadowing.

Furthermore, synthesized TTS shows great potential for contributing to research activities that require experiments using L2 speech to control various components. In such cases, it will be possible to conduct experiments more precisely by considering that the size of the perceptual priming effect differs based on the learner's proficiency.

Finally, the challenges of this study should be addressed.

Since this study used only one version of speech synthesizing software, there is a need for future research using a number of different software packages. Furthermore, the researcher recognizes that English language classrooms often have non-native as well as native teachers; however, to this date there have been no auditory priming studies using native and non-native speech. This is something the researcher hopes to address further in the future. In addition, since this study was on auditory word priming experiment to verify spoken word recognition, the researcher is of the opinion that adding evidence in phrasal or sentence units may increase the potency of these results.

There is a high possibility that synthetic TTS will play a part in building a rich learning environment in English language classrooms, as well as outside the classrooms, in the future. If synthetic TTS is incorporated into English language learning where the level of language proficiency is part of the criteria for using it, then synthetic TTS will likely become an exceptionally convenient and effective tool.

Chapter 6
Effects of Auditory Word Repetition

6.1 Experiment 4

6.1.1. Participants

Forty Japanese undergraduate students (17 men and 23 women) at a university located in Nishinomiya City participated in this experiment. All participants were drawn from the same pool as in Experiment 1. All participants volunteered for the present research. The participants were aged between 18 and 22 and enrolled in different faculties. Twenty-six participants reported the scores of the Test of English for International Communication (the TOEIC® Listening and Reading test), ranging from 280 to 940 (*M* = 540.00, *SD* = 150.30). All participants were living in Japan, and for them English was a foreign language. None of them had spent more than three months in an English-speaking country. All participants reported normal hearing and vision at the time of this experiment.

As proficiency varied among participants, variables least likely to pose a cognitive load on L2 learners were chosen. Based on the results of Experiments 1 through 3, this experiment utilized a single voice (no variability) for synthesized speech.

6.1.2. Materials

Words used in this experiment partially overlapped those of Experiment 1. The study phase consisted of a rhyme or a synonym judgment task, and the test phase included a vocal repetition task or a vocal repetition task after subvocal repetition.

All words used in the current experiment were separated into 10 vocabulary groups (see V1 to V10 in Appendices A to D). Two sets of 48 English words in pairs were used (see Appendices A and B): Task 1 (rhyme judgment task): V1+V2+V3 and Task 2 (synonym judgment task): V4+V5+V6. For the repetition task in the test phase, 4 sets of 48 English words were used (see Appendices A, B, C, and D): Task 3 (a vocal repetition task after the rhyme judgment task): V1+V7, Task 4 (a vocal repetition task after subvocal repetition conducted after the rhyme judgment task): V2+V8, Task 5 (a vocal repetition task after the synonym judgment task): V4+V9, and Task 6 (a vocal repetition task after subvocal repetition conducted after the synonym judgment task): V5+V10. Words from V7, V8, V9, and V10 were not presented in the study phase (unrepeated words or novel words).

All the words were selected from the English words familiarity database of Japanese EFL learners (Yokokawa, 2006, 2009). The words used in Task 1 and 2 were, on average, 2.37 syllables long, with a mean familiarity of 5.94 and a mean frequency of 26.84 occurrences per thousand words based on the British National Corpus (BNC). The words in each task did not differ significantly in terms of word frequency, word familiarity, or syllable number.

The words used in Task3, 4, 5, and 6 were, on average, 2.37 syllables long, with a mean familiarity of 5.93, and a mean frequency of 24.96 occurrences per thousand words based on the BNC. The words in each task did not differ significantly in terms of word frequency, word familiarity, syllable number, and duration.

As in Experiment 1, the first phonemes of words in each list began with the same phoneme in order to counterbalance the effect of phoneme recognition through a microphone.

6.1.3. Procedure

The experiment was conducted in the same way as Experiment 1, with some exceptions. It took approximately 60 minutes for participants to complete the experiment (see Figure 13).

While the same procedure was conducted during the study phase, the participants were asked to repeat the words aloud or in their minds, as accurately and as rapidly as possible during the test phase (instructions were shown in Figures 11 and 12.). The number of repetitions in this study was set to four in accordance with Terasawa, Yoshida, and Ohta (2008). For the vocal repetition task, participants were asked to repeat aloud 48 words four times. For the vocal repetition task after subvocal repetition, they were asked to do one vocal repetition after three subvocal ones.

The order of tasks (rhyme judgment, synonym judgment, vocal repetition, and vocal repetition after subvocal repetition) was counterbalanced across the participants in order to eliminate any influence of task order (see the caption of Figure 13).
In order to measure the reliability of error identification, a Japanese woman teaching English at a high school was asked to be an independent rater. The degree of agreement in error identification between the researcher and rater was 96.11%. Incorrect repetitions were excluded (8.39% of all responses) from the RTs and duration data. In addition, in order to eliminate

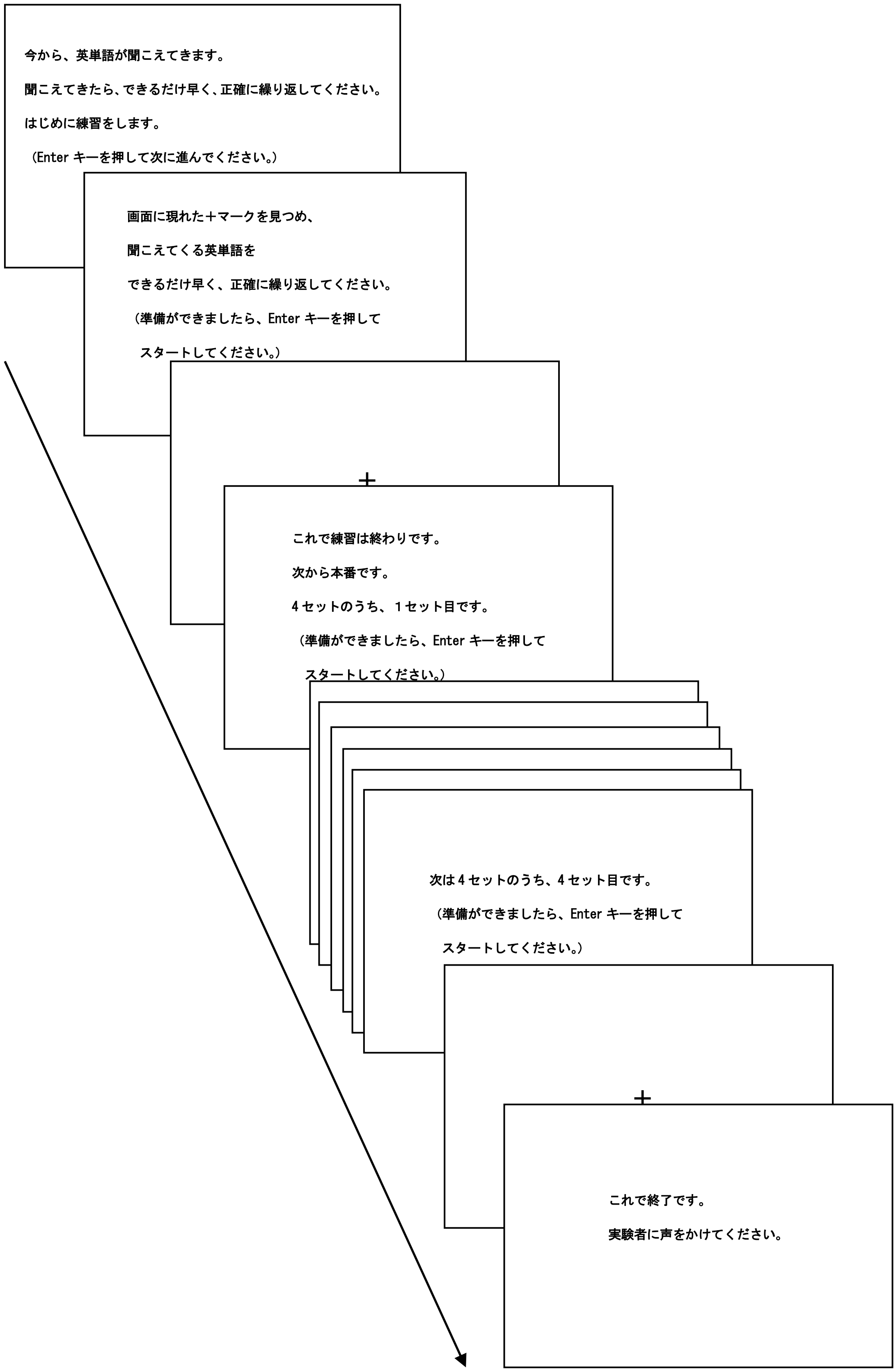

Figure 11. The instructions for the vocal repetition task on a computer screen.

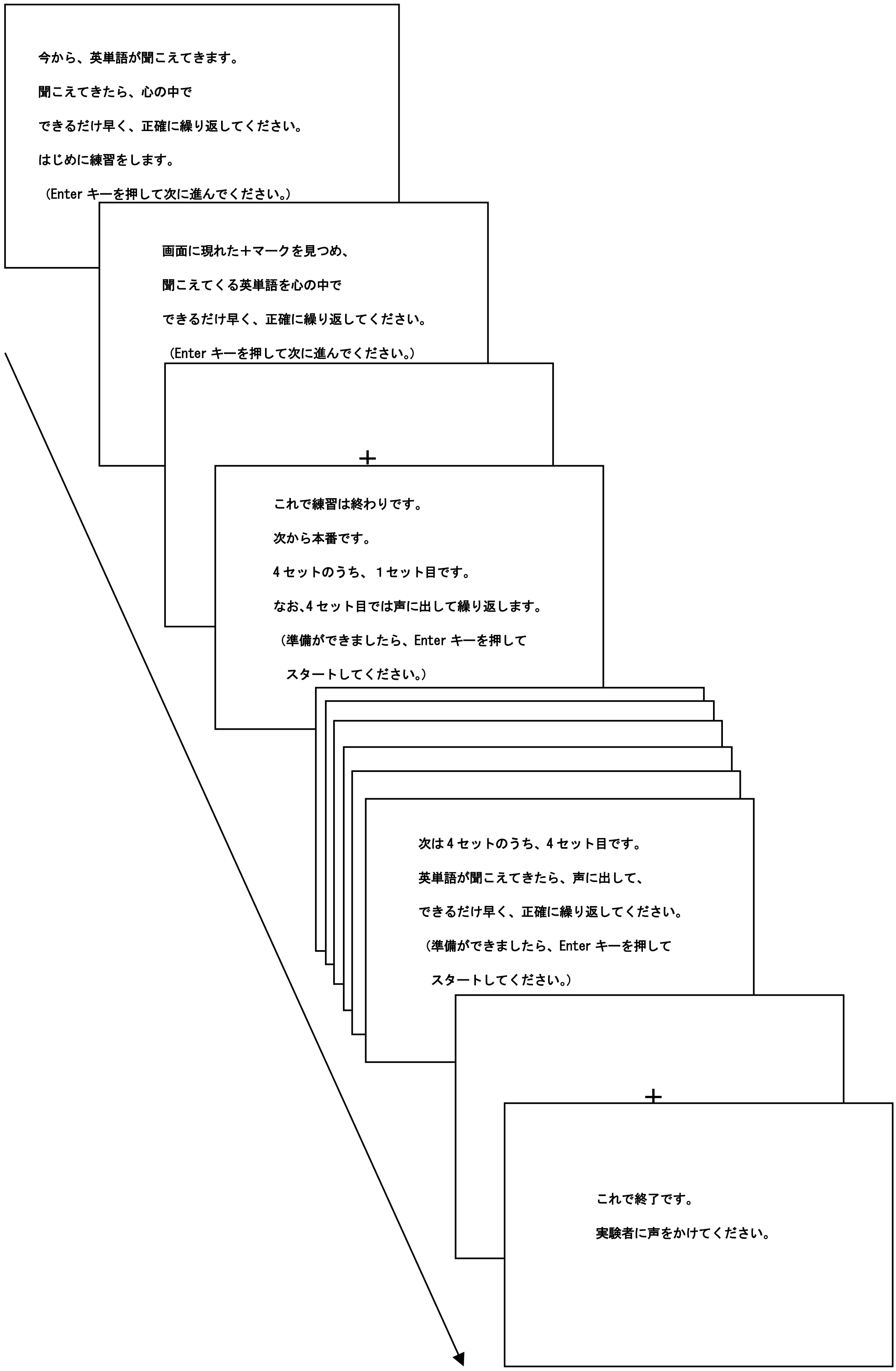

Figure 12. The instructions for the vocal repetition task after subvocal repetition on a computer screen.

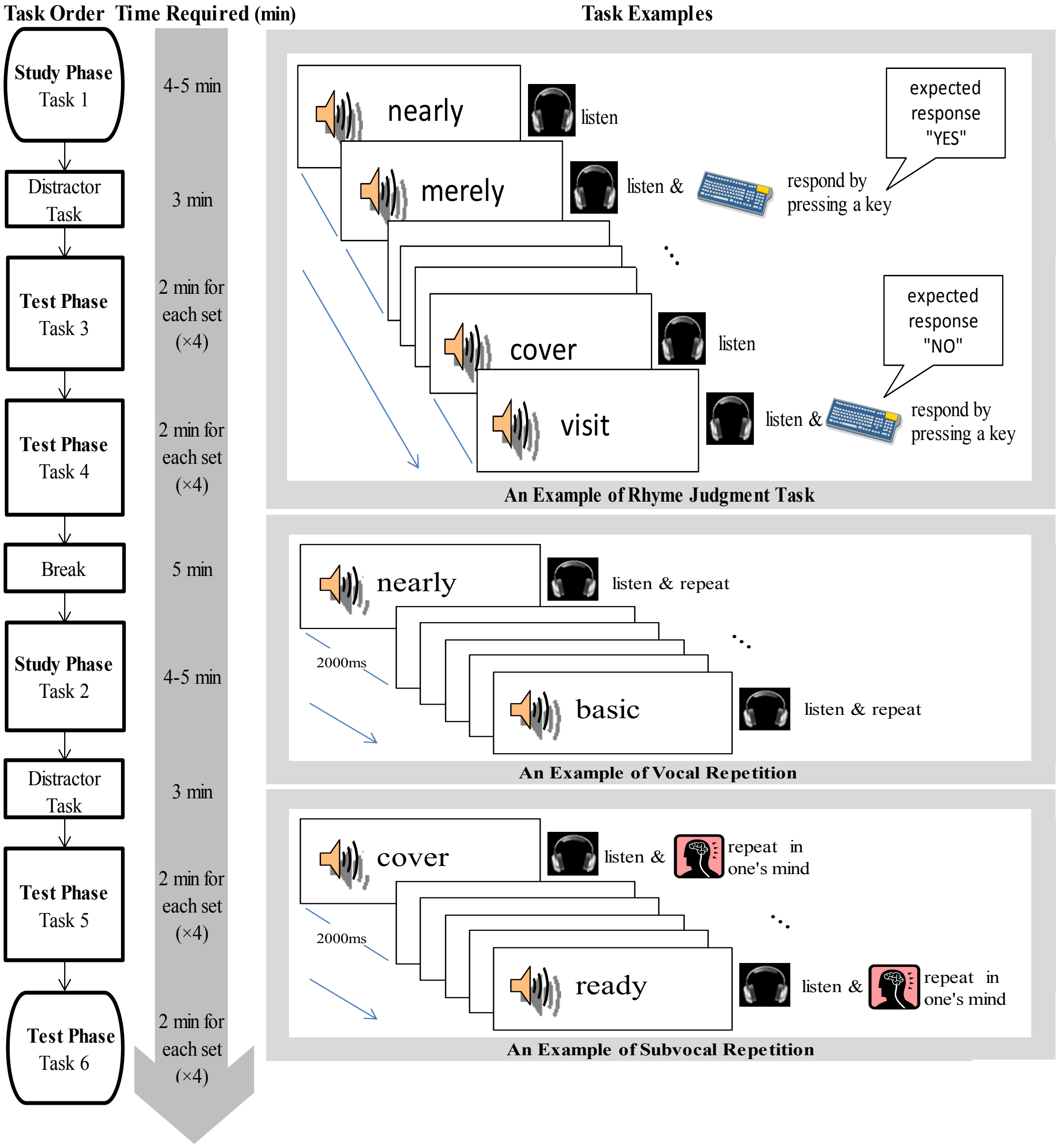

Figure 13. An example of the task orders, time required for each task, and task examples. Task 1 = rhyme judgment task; Task 2 = synonym judgment task; Task 3 = a vocal repetition task (four times) after rhyme judgment task; Task 4 = a vocal repetition task (once) after subvocal repetition (three times) conducted after rhyme judgment task; Task 5 = a vocal repetition task (four times) after synonym judgment task; Task 6 = a vocal repetition task (once) after subvocal repetition (three times) conducted after synonym judgment task. The order of tasks was counterbalanced across the participants in order to eliminate any influence of task order. There were eight patterns for the task orders: Task 1-3-4-2-5-6 (as shown in the illustration), Task 1-4-3-2-6-5, Task 1-4-3-2-5-6, Task 1-3-4-2-6-5, Task 2-5-6-1-3-4, Task 2-6-5-1-4-3, Task 2-6-5-1-3-4, Task 2-5-6-1-4-3. Five people participated in each pattern.

outliers, the RTs and duration data below and above mean responses±2 standard deviation (*SD*) of each participant were discarded (4.28% of all correct responses). The recorded response of the participants—11,520 words in total—was transferred to a computer for analyses.

6.1.4. Data Analyses

RT data were analyzed in the same way as in Experiment 1. RT data and error rates were analyzed with ANOVA.

6.2 Results of Experiment 4

The effects of the number of repetitions and the processing orientation are explained in the first subsection, while the effects of the two different repetition methods and the processing orientation are presented in the next subsection. The results are presented in the following order: RTs and error rates of the two dependent variables.

6.2.1. The Effects of Number of Repetitions and Processing Orientation

6.2.1.1. RT data

RT data of vocal repetition were analyzed with a 2×3 repeated measures ANOVA with the number of repetitions (the first repetition and fourth repetition) and the three conditions of the processing orientation in the study phase (rhyme condition: the rhyme judgment task was conducted in the study phase; synonym condition: the synonym judgment task was conducted in the study phase; novel condition: no judgment task in the study phase) as within-subjects factors.

This analysis revealed a statistically significant main effect for the number of vocal repetitions, $F(1, 39) = 50.02$, $p = .00$, $\eta_p^2 = .56$; however, there was no statistically significant main effect for processing orientation, nor was there an interaction between the number of vocal repetitions and processing orientation (see Table 10 and Figure 14). Thus, RTs decreased as the number of repetitions increased regardless of the judgment task in the study phase.

Table 10

Mean Reaction Times (ms) as a Function of the Number of Repetitions and Processing Orientation

	1st		4th	
	M	*SD*	*M*	*SD*
Rhyme	720.84	142.30	604.99	185.59
Synonym	714.22	161.28	623.26	190.67
Novel	739.16	140.90	622.94	183.87

Note. $N = 40$

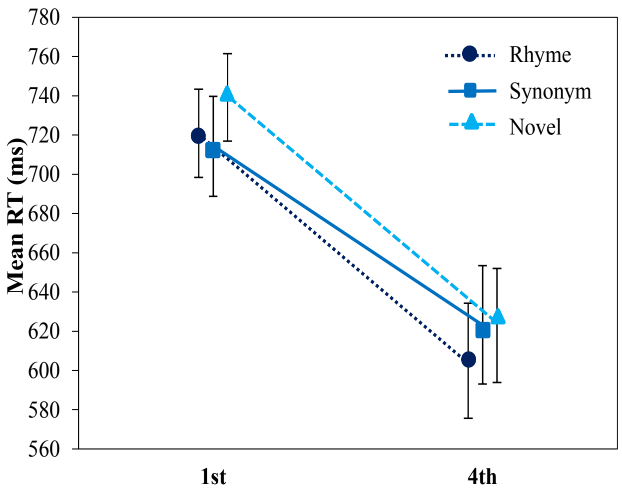

Figure 14. Mean Reaction Times (ms) as a function of the number of repetitions and processing orientation. The vertical lines indicate standard error.

6.2.1.2. Error Rate Data

As above, error rates of vocal repetition were analyzed with a 2×3 repeated measures ANOVA with number of repetitions (first, fourth) and the three conditions of the processing orientation in the study phase (rhyme, synonym, novel) as within-subjects factors. The analysis revealed a statistically significant main effect for processing orientation, $F(2, 78) = 13.99$, $p < .001$, $\eta_p^2 = .26$, but there was no statistically significant main effect for number of repetitions, nor was there an interaction between the number of repetitions and processing orientation (see Table 11 and Figure 15). Bonferroni tests revealed that the error rate showed a statistically significant decrease for the synonym condition compared with the rhyme condition ($p < .001$), and the synonym condition compared with the novel condition ($p = .002$). However, the difference between the rhyme and novel conditions did not reach statistical significance. In sum, error rates were not affected by the number of repetitions, but appear to be affected by the difference in the processing orientation. In fact, the mean rate of making the same errors for the first and fourth repetition was 61.46% for the rhyme judgment task, 48.15% for the synonym judgment task, and 61.25% for novel words.

6.2.2. The Effects of the Repetition Method and Processing Orientation

6.2.2.1. RT Data

RT data of the fourth vocal repetition after three vocal or subvocal repetitions were analyzed with a 2×3 repeated measures

Table 11

Mean Error Rates (%) as a Function of The Number of Repetitions and Processing Orientation

	1st		4th	
	M	*SD*	*M*	*SD*
Rhyme	5.00	3.05	4.75	3.36
Synonym	2.82	2.94	2.29	2.35
Novel	4.17	2.34	4.14	2.28

Note. N = 40.

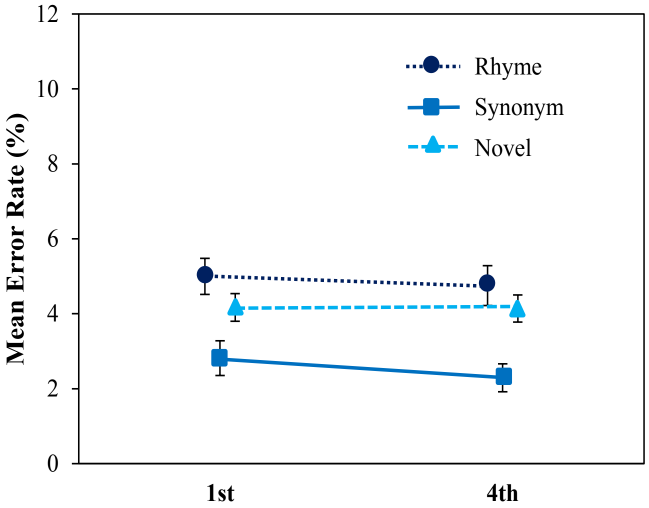

Figure 15. Mean error rates (%) as a function of the number of repetitions and processing orientation. The vertical lines indicate standard error.

ANOVA with repetition method (vocal repetitions, vocal repetitions after subvocal repetitions) and the three conditions of the processing orientation in the study phase (rhyme, synonym, novel) as within-subjects factors. The main effect of both the repetition method, $F(1, 39) = 36.60$, $p < .001$, $\eta_p^2 = .48$, and the processing orientation, $F(2, 78) = 5.98$, $p = .02$, $\eta_p^2 = .13$, showed statistical significance but there was no statistically significant interaction between the repetition method and the processing orientation (see

Table 12 and Figure 16). Bonferroni tests revealed a statistically significant difference between the rhyme and novel conditions (p = .002), but not between the synonym and novel or the rhyme and synonym conditions. The mean RT was shorter for the rhyme condition than for the novel condition. In sum, vocal repetition and rhyme condition appear to have accelerated the RTs.

Table 12

Mean Reaction Times (ms) as a Function of Repetition Method and Processing Orientation

	Vocal		Subvocal	
	M	*SD*	*M*	*SD*
Rhyme	604.99	185.59	662.68	175.97
Synonym	623.26	190.67	692.85	178.74
Novel	622.95	183.87	690.46	171.10

Note. $N = 40$. Subvocal = vocal repetitions after subvocal repetitions.

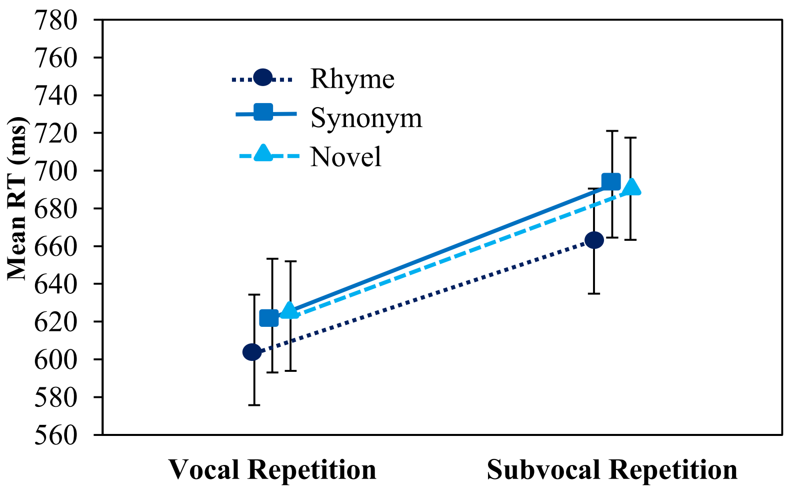

Figure 16. Mean Reaction Times (ms) as a function of repetition method and processing orientation. The vertical lines indicate standard error.

6.2.2.2. Error Rate Data

As above, the data of the fourth vocal repetition after three vocal or subvocal repetitions were analyzed with a 2×3 repeated measures ANOVA with repetition method (vocal repetitions, vocal repetitions after subvocal repetitions) and the three conditions of the processing orientation in the study phase (rhyme, synonym, novel) as within-subjects factors. The main effects of the repetition method, $F(1, 39) = 13.75$, $p = .001$, $\eta_p^2 = .26$, and processing orientation, $F(2, 78) = 36.92$, $p < .001$, $\eta_p^2 = .49$, showed statistical significance. The interaction between the repetition method and processing orientation was also statistically significant, $F(2, 78) = 17.19$, $p < .001$, $\eta^2 = .31$ (see Table 13 and Figure 17). Bonferroni tests revealed that the error rate was statistically significantly lower for the synonym condition compared with the rhyme condition ($p = .001$), the rhyme condition with the novel condition ($p < .001$), and the synonym condition with the novel condition ($p < .001$). Consequently, the synonym judgment task in the study phase appears to have resulted in decreased errors for the vocal repetition task.

Table 13
Mean Error Rates (%) as a Function of Repetition Method and Processing Orientation

	Vocal		Subvocal	
	M	*SD*	*M*	*SD*
Rhyme	4.75	3.36	3.35	2.84
Synonym	2.29	2.35	3.02	2.45
Novel	4.14	2.28	8.98	6.65

Note. $N = 40$. Subvocal = vocal repetitions after subvocal repetitions.

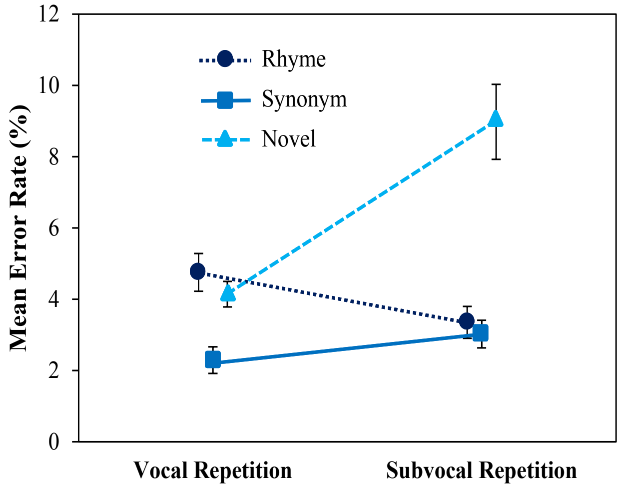

Figure 17. Mean error rates (%) as a function of repetition method and processing orientation. The vertical lines indicate standard error.

6.3 Discussion of Experiment 4

Detailed analyses of the key research findings are presented with reference to each of the research questions and hypotheses.

6.3.1. Effects of the Number of Repetitions

It was hypothesized that RTs and error rates would decrease as the number of repetitions increased. However, the results show that the number of repetitions had a statistically significant effect on the RTs but not on the error rate. Since repetition is considered to be a fluency building task, improvement of accuracy is likely to be one of the ultimate goals. The present study investigated word processing and participants were asked to repeat each word by simply relying on their phonetic perception of it. However, the results indicate that frequent exposure to vocabulary does not improve accuracy in Japanese EFL learners.

To sum up, the number of repetitions seems to contribute to faster responses for L2 words; however, there is no indication that word-based repetition improves or affects accuracy in Japanese EFL learners. In addition, no interaction was seen between the other factors in this study. These results seem to suggest that, at least as far as Japanese EFL learners are concerned, word repetition tasks may be familiarity-building rather than a fluency-building. Building familiarity is important as well from the point of view of memory research because it affects word memory retrieval depicted in the following sections.

6.3.2. Effects of the Repetition Method

The second research question investigated the effects of repetition method—vocal and subvocal repetitions. It was hypothesized that RTs and error rates would be lower with vocal repetitions compared with vocal repetitions after subvocal repetitions. As hypothesized, it was revealed that vocal repetitions lowered both RTs and error rates compared with subvocal repetitions. A possible explanation for the ineffectiveness of subvocal repetitions compared with vocal repetition is that participants used their own phonological representation of the words during subvocal repetition, which may in fact differ from the model sounds. Moreover, as noted in previous studies, vocal repetition provides an opportunity for auditory self-perception (Baker & Trofimovich, 2006), resulting in self-correction.

It is important to note that there was a statistically significant interaction between the repetition method and processing orientation for error analysis. The combination of the synonym judgment task in the study phase and the vocal repetition task in the test phase appears to be an efficient way to reduce errors.

Although a more detailed analysis is required to draw concrete conclusions, this study found that vocal repetitions appear to have enabled participants to produce each word more accurately and quickly. In addition, combining repetition with semantic tasks seemed to aid learners in improving accuracy, making this a potentially efficient repetition method.

6.3.3. Effects of Processing Orientation

The third research question is concerned with the effect of processing orientation. It was hypothesized that a rhyme judgment task would produce lower RTs and error rates compared with a synonym judgment task, based on the priming paradigm.

No statistically significant differences were found among RTs between the three conditions—rhyme, synonym, and novel—both for the first and fourth repetitions. Moreover, the fourth repetition data for the vocal repetition and vocal repetition after subvocal repetitions methods showed a statistically significant difference between rhyme and novel conditions, but did not find a statistically significant difference between the rhyme and synonym conditions. This illustrates that the rhyme condition appears to reduce RTs, while the difference in processing orientation does not seem to have any influence on RTs. In addition, the number of repetitions and the repetition method appear to have a larger influence on RTs.

Contrary to the hypothesis, the synonym judgment task showed a lower error rate in repetition tasks. This suggests that semantic processing may produce more accurate responses. Moreover, as stated before, the combination of vocal repetitions and semantic processing appears to be particularly efficient.

6.4 Conclusion of Experiment 4

Several limitations must be pointed out. First of all, participants in this study were limited to undergraduate students at one university and their number was relatively small. As such, we must avoid over-generalizing until these results can be repeated with a bigger and more varied sample. In addition, processing orientation is a method (solely based on theory) for priming experiments. This means that regardless of the experimenter's intention, high-proficiency-level participants may in fact focus their attention on other factors (e.g. meaning instead of phonological character, and vice versa). Therefore, at this point it cannot be confidently said that processing details during one's learning period are identical for all participants. Although the present study followed the experimental methods of preceding studies, a more sophisticated experiment method will be required in the future. To measure the effects of subvocal repetition, participants were asked to repeat words aloud once in the subvocal repetition experiment. Therefore, it cannot be definitively said that the results showed pure subvocal repetition effects, which is another limitation of the present methodology.

The current study focused on analyzing two quantitative categories of data—RTs and error rates—concerning word processing; however, additional data, including qualitative data, must also be considered in future studies, as other factors are also likely to be involved in word processing.

Despite these limitations, the analyses of auditory word repetition in the present study provide a basic picture of the relationship of factors associated with repetition. Since repetition is considered to be a fluency-building task, repetition tasks should facilitate the desired effects: fast and accurate responses. However, this experiment showed that an increased number of repetitions appears to contribute only to swifter responses and has little effect on the accuracy of word-based repetitions. This suggests that repetition for Japanese EFL learners may be simply a familiarity-building task. However, vocal repetition seems to aid learners in producing each word more accurately and quickly, and if it followed after a semantic task, it appears to be more efficient. In other words, the two factors of repetition analyzed here (repetition method and processing orientation) seem to have a mutually complementary relationship, while the number of repetitions appears to be independent. The number of repetitions can then be interpreted as the quantity of speech and the repetition method and processing orientation as the quality of speech. The quantitative factors and qualitative factors of speech seem to have separate influences on L2 speech processing.

The conclusion that we can draw from these results is that it may be more effective for Japanese EFL learners to use semantic processing before repetition, while the repetitions themselves should be more frequent and vocalized. The overall findings consistently underscore the importance of well-planned repetition tasks for Japanese EFL learners and the importance of repetition in L2 learning at the perceptual level. Further empirical studies, including more fundamental research into psycholinguistics, will lead to a fuller understanding of the mechanisms of speech processing in EFL learners.

6.5 Experiment 5

6.5.1. Participants

Forty native speakers (NS) of English who were students from overseas and 40 nonnative speakers (NNS) who were Japanese EFL graduate and undergraduate students of universities in the Kansai area participated in this experiment. There is a partial overlap with the participants of Experiment 1. The participants were individually tested in a quiet room for about 60 to 70 min.

6.5.2. Materials

All the words used at the encoding stage were shown in Appendices A, B, C and D used in Experiment 4, and some words used at the retrieval stage were selected from these vocabulary lists, where word familiarity (Yokokawa, 2006, 2009), frequency (BNC), syllable number, duration and the first phonemes of words were taken into account (see Appendix J & K). A one-way ANOVA was conducted to confirm that there were no differences among the vocabulary groups in terms of familiarity (spoken), $F(4, 75) = 0.02$, $p = .99$, $\eta_p^2 = .001$; familiarity (written), $F(4, 75) = 0.27$, $p = .90$, $\eta_p^2 = .01$; word frenquency, $F(4, 75) = 0.58$, $p = .68$, $\eta_p^2 = .02$; syllable number, $F(4, 75) \fallingdotseq 0$, $p = 1.00$, $\eta_p^2 \fallingdotseq 0$; and duration, $F(4, 75) = 1.21$, $p = .31$, $\eta_p^2 = .06$.

Words for the recognition task were recorded using the speech synthesis software, Globalvoice English Version 2 (PENTAX), and presented using personal computers via speakers without visual aids. The Super-Lab Experimental Laboratory Software 4.0 (Cedrus) was used to present the words randomly and record the responses of participants.

6.5.3. Procedure

Figure 18 shows the procedure at the encoding and retrieval stages. The number of repetition was set to four based on previous research. All tasks were self-paced. During the encoding stage, participants were asked to do a rhyme judgment task (judge whether each pair rhymed or not) or synonym judgment task (judge whether each pair had a similar meaning) and then repeat words (which include the studied words in the judgment tasks and new words) aloud or in their minds. The order of tasks was counterbalanced across the participants (for further details of repetition tasks, see Experiment 4). At the retrieval stage, participants were asked to do a recognition task (participants did not anticipate a memory test at the encoding stage). In the task, a total of 80 words were presented, with 40 words selected from those presented at the encoding stage (Vocabulary Group 1 to 4 in Figure 18 and Appendix J) and 40 new words (Vocabulary Group 5 in Figure 18 and Appendix K). The participants were asked to judge whether or not they heard the word in the previous experiment by pressing the correct key as soon as possible (instructions were shown in Figure 19).

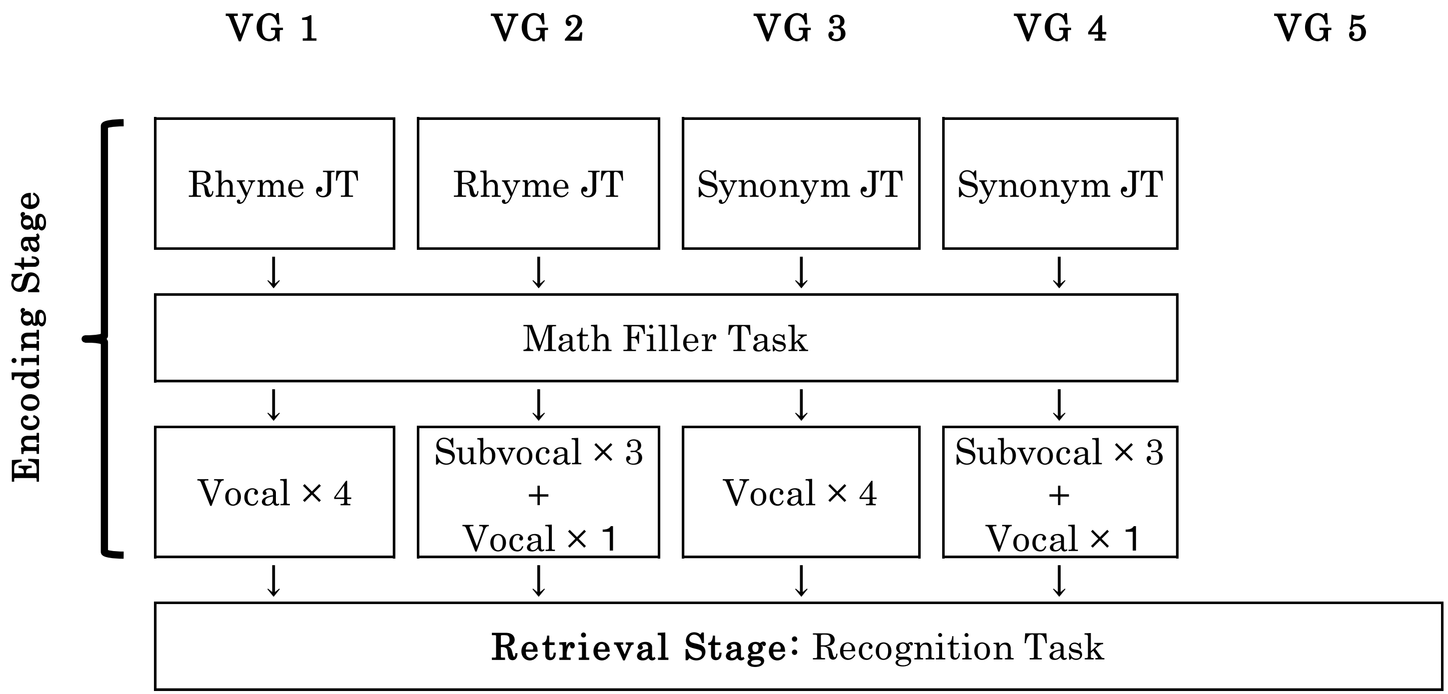

Figure 18. Procedure at the encoding and retrieval stages. VG = vocabulary group; JT = judgment task; Vocal = vocal repetitions; Subvocal = subvocal repetitions.

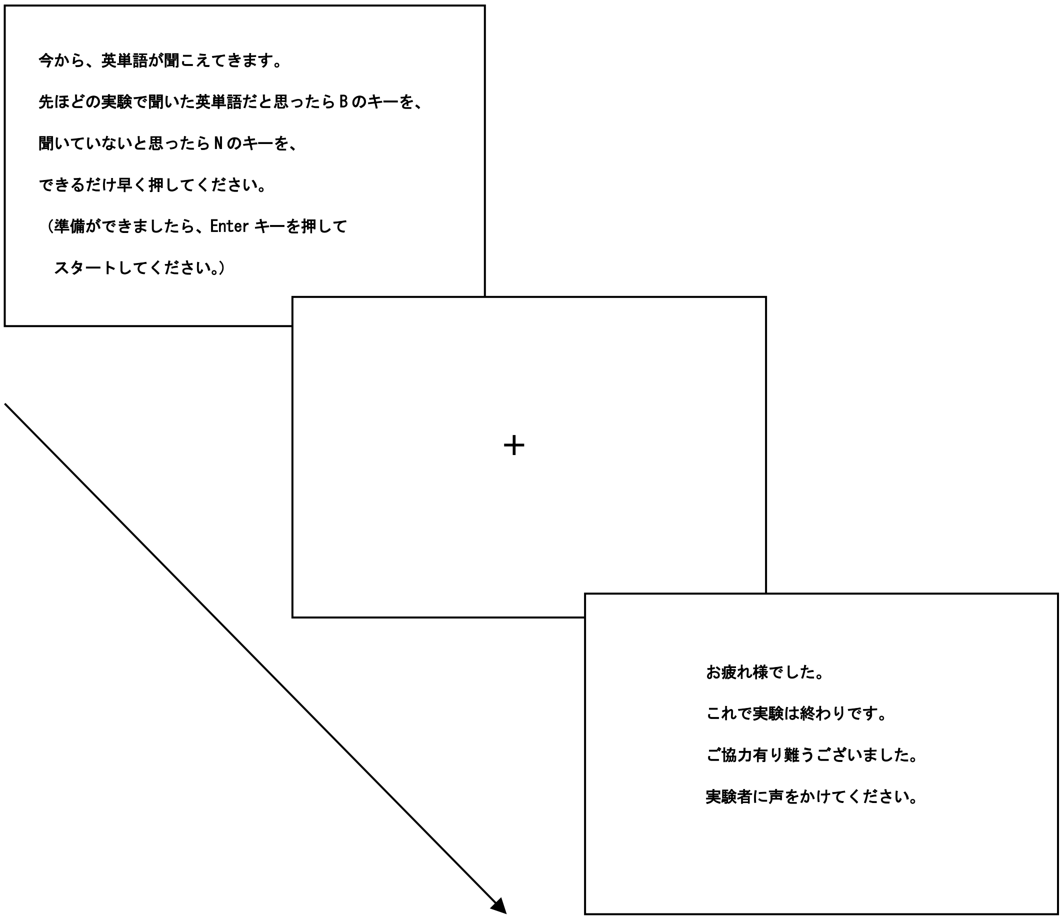

Figure 19. The instructions for the recognition task on a computer screen.

6.5.4. Data Analyses

The RT data of correct answers ('YES' for VG 1 to 4, 'NO' for VG 5) and error rates of the recognition task were computed. RT data 2.5 standard deviations (*SDs*) away from each participant's mean was replaced with the sum of the mean and 2.5 *SD*s (2.08% of the data). Significance level was set at .05.

6.6 Results of Experiment 5

6.6.1. Error Rate

Table 14 and Figure 20 show the mean error rates of the recognition test. A mixed design 2 (Group) × 5 (Encoding Factors) ANOVA was conducted with Group (native or nonnative) as a between-subject factor and Encoding Factors (see VG1 to VG5 of Figure 1) as within-subject factors. Post-hoc comparisons were conducted using the procedure from Benjamini and Hochberg (1995). The analysis revealed statistically significant main effects of Group, $F(1, 78) = 8.52$, $p = .005$, $\eta_p^2 = .10$, *power* = 1.00, and Encoding Factors, $F(4, 312) = 6.59$, $p < .001$, $\eta_p^2 = .08$, *power* = 1.00, and there was no statistically significant interaction between Group and Encoding Factors, $F(4, 312) = 1.41$, $p = .23$, $\eta_p^2 = .02$, *power* = 0.63. The post-hoc test of Encoding Factors showed that the error rates of VG1 and VG2, VG1 and VG3, VG2 and VG3, VG2 and VG4 and VG2 and VG5 (*adjusted ps* = .02, = .046, < .001, = .006, = .03, respectively) had statistically significant differences, while other pairs did not reach statistical significance: VG1 and VG4, VG1 and VG5, VG3 and VG4, VG3 and VG5, VG4 and VG5 (*adjusted ps* = .58, = .88, = .11, = .10, = .75).

Deep processing (semantic processing) significantly affected

Table 14
Mean Error Rates (%) as a Function of Encoding Factors

		VG1	VG2	VG3	VG4	VG5
	n	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*
NS	40	28.25 (2.58)	31.50 (3.45)	20.75 (3.29)	21.25 (2.71)	23.88 (2.20)
NNS	40	15.50 (2.06)	26.75 (3.15)	12.25 (2.71)	19.50 (3.18)	19.13 (1.61)

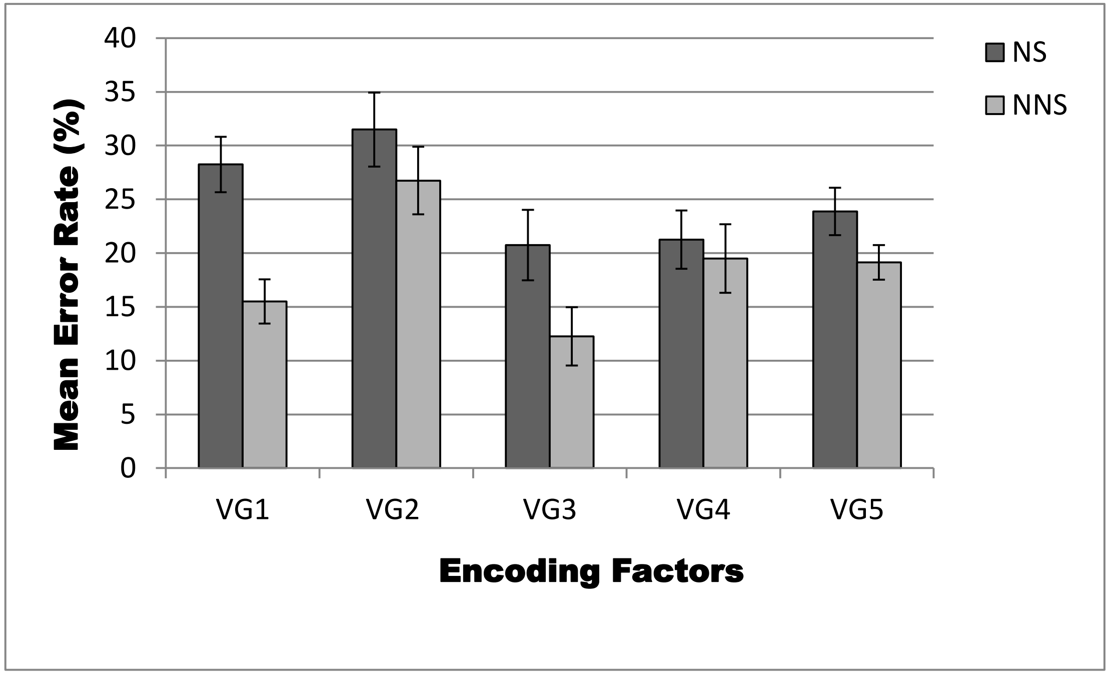

Figure 20. Mean error rates (%) as a function of encoding factors. The vertical lines indicate standard error.

accuracy, while with shallow processing (nonsemantic processing), vocal repetition significantly influenced both NS and NNS. In addition, it is interesting to note that NS made more mistakes than NNS.

6.6.2. RT Data

Table 15 and Figure 21 display the mean RT data of the recognition test. The same statistical analyses as for error rate data were performed on the RT data. The results showed a statistically significant interaction between Group and Encoding Factors, $F(4, 312) = 2.86$, $p = .02$, $\eta_p^2 = .04$, *power* = 1.00. The simple main effect of Encoding Factors was statistically significant for NS, $F(4, 312) = 22.01$, *adjusted p* < .001, $\eta_p^2 = .22$, and for NNS, $F(4, 312) = 8.09$, *adjusted p* < .001, $\eta_p^2 = .10$. The post-hoc test showed that the RT of VG5 (where the correct answer is 'NO') was statistically significantly longer than VG1 to VG4 (where the correct answer is 'YES') for NS (*adjusted ps* < .001, = .003, < .001,< .001, respectively) and NNS (*adjusted ps* < .001, = .02, < .001, < .001, respectively). Moreover, NS, VG1 and VG2, VG2 and VG3, and VG3 and VG4 (*adjusted ps* = .04, < .001, < .001, respectively) showed statistically significant differences, while other pairs did not show statistical significance: VG1 and VG3, VG1 and VG4, and VG2 and VG4 (*adjusted ps* = .054, = .08, = .66, respectively). Similarly, NNS, VG1 and VG3, VG2 and VG3 (*adjusted p* = .04, for both) have statistically significant differences, while other pairs did not show statistical significance; VG1 and VG2, VG1 and VG4, VG2 and VG4, and VG3 and VG4 (*adjusted ps* = .64, = .41, = .27, = .20, respectively).

Table 15
Mean Reaction Times (ms) as a Function of Encoding Factors

		VG1	VG2	VG3	VG4	VG5
	n	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*	*M (SE)*
NS	40	1131 (31.82)	1206 (46.19)	1071 (27.60)	1190 (36.71)	1310 (35.94)
NNS	40	1167 (36.97)	1178 (37.86)	1118 (38.62)	1146 (33.79)	1262 (33.91)

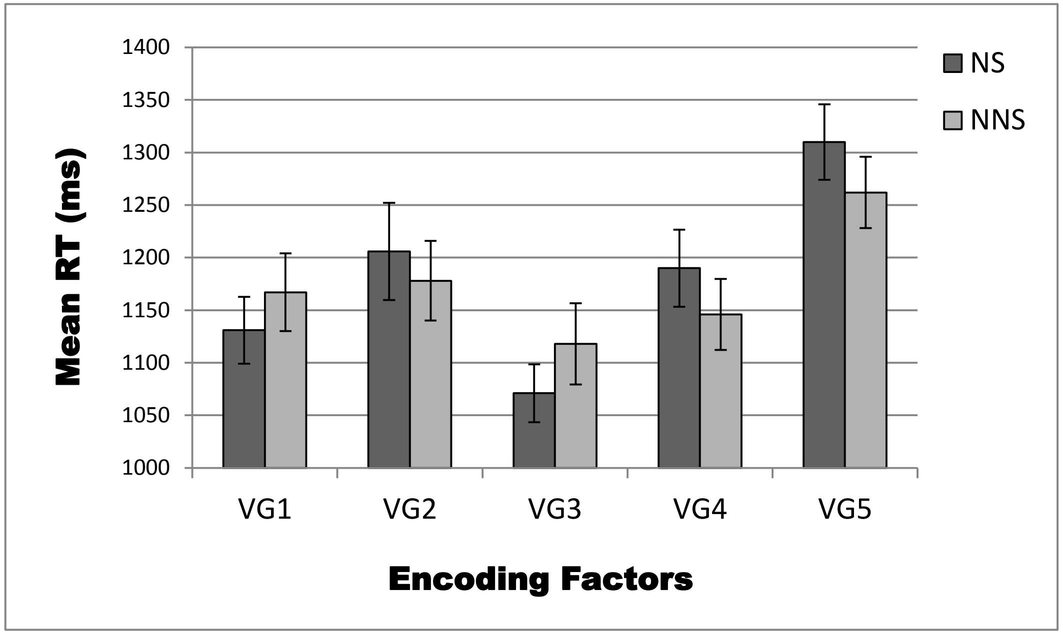

Figure 21. Mean Reaction Times (ms) as a function of encoding factors. The vertical lines indicate standard error.

Consequently, 'NO' answers were statistically significantly faster than 'YES' answers. Deep processing seemed to affect the speed of recognition for NS since the effect of deep processing is approaches statistical significance (e.g.,VG1 and VG3). More importantly, vocal repetition produced statistically significantly lower RTs regardless of previous processing. On the other hand, vocal repetition produced statistically significantly lower RTs for NNS only after deep processing.

6.7 Discussion of Experiment 5

The present study analyzed the effects of encoding factors on vocabulary learning using recognition tasks.

Error rate results support the first hypothesis (deep processing at the encoding stage will reduce the error rate in the recognition task) for both NS and NNS. In accordance with a series of previous studies (LOP framework, TAP principle and the encoding specificity principle), the level of processing affected the accuracy of word memory retrieval in L1 and L2 and the effect was consistent. As stated before, word information could be elaborated through deep or semantic processing because it could be integrated with existing knowledge in one's mental lexicon.[20] Since the L1 mental lexicon usually has a lot of information, it should be superior in word memory retrieval. However, the results did not support this idea. Taking the distribution of the error data (see *SEs* of Table 14) of NS and NNS into account, it may

[20]The mental lexicon is a construct where words are organized. Linguists and psycholinguists use the word to refer to individuals' words representations.

be assumed that the complexity of word information in the L1 mental lexicon hinders accurate word retrieval.

RT data results partially support the second hypothesis (repetition will accelerate the recognition of words and above all, vocal repetition will enhance recollection of words). Both vocal and subvocal repetitions (VG1 to VG4) seem to enhance speed in the recognition task compared with new words (VG5). However, the effects of vocal repetition seem to be different between NS and NNS. The RT data of NS showed that the speed of recognition was enhanced by the number of overt rehearsals regardless of the level of processing, but this does not apply to NNS. Overt rehearsals showed advantages only after semantic processing for NNS. Furthermore, based on error rate results, vocal repetition after shallow processing could aid accurate retrieval. The effects of vocal repetition seemed to be supplemental for accurate word retrieval for both NS and NNS.

In summary, deep processing can enhance accuracy but not retrieval speed for NS, while vocal repetition appears to facilitate the skills of consulting a dictionary. The results coincide with hypotheses derived from the dual-process models, and the existence of familiarity-based and recollection-based retrievals can be inferred. On the other hand, deep processing appears to enhance retrieval accuracy for NNS, but vocal repetition does not by itself appear to accelerate retrieval. The possible explanation is that the role of deep processing for NNS is qualitatively different from that of NS. Most word representations of the L2 mental lexicon are considered to be vulnerable. Elaboration of L2 words by deep or semantic processing is essentially equal to elaboration in L1, because L1 and L2 should be interconnected. The use of L1 might compensate for the unstable L2 word representations. After elaboration, the target word clearly manifests in the L2, making access easier and allowing overt rehearsal to be effective as a retrieval cue. Therefore, it may be more efficient for L2 learners to elaborate the word representation using semantic processing, such as L1 translation or the use of context, before enhancing the skills of consulting the dictionary. Moreover, it should be noted that the supplemental effects of repetition are from the viewpoint of word recognition. Considering that repetition after shallow processing increased the accuracy rate, there is a high probability that the learners' gains through perceptual fluency building may include factors that cannot be measured by this memory test.

6.8 Conclusion of Experiment 5

This study investigated the effect of encoding factors on L2 word memory retrieval. The results of this study emphasize the importance of meaning-based processing before conducting fluency building tasks at the encoding stage for L2 word retrieval.

Limitations of this study could be seen in the procedure of the experiment. Some methods are available to measure two processes, familiarity and recollection, such as the task-dissociation method.[21] This study could be further supported with the use of such methods. Moreover, this study measured only short term vocabulary retention; however, in order to address educational gains, future studies should also investigate long term

---

[21] Refer to Jacoby (1991) and Jacoby, Toth, and Yonelinas (1993).

vocabulary retention of EFL learners.

Chapter 7
General Discussion

Experiment 1 was an auditory priming experiment conducted on Japanese learners of English (non-native speakers: NNS) and native speakers (NS) of English. As in previous research, the priming effect was observed in both groups. These results supported the existence of a universal mechanism involved in the learning and acquisition of language based on implicit memory. However, unlike in previous studies, the priming effect that followed semantic processing in NNSs was statistically significantly higher. Since the extent of processing effects exhibited by NSs following semantic and nonsemantic processing was the same, this experiment can be considered to be valid. This study demonstrated that approaches based on the levels of processing (LOP) framework may be effective for improving speech perception ability in L2 acquisition. However, considering that NNSs were more easily affected by perceptual information, unlike in previous research, it might be necessary to consider the influence of using synthetic speech in experiments. Moreover, the influence of the learner's proficiency level remains a topic for future analyses.

In Experiment 2, the effects of changing the speaker (= paralinguistic information) were examined through priming experiments that used natural human voices. Similar to previous research, it was found that NNSs could not process paralinguistic information independent of linguistic information, and that the perceptual learning effect was statistically significantly higher in cases where the same voice was used. However, if it is necessary to allocate a significant amount of cognitive resources to semantic information, attention to paralinguistic information may be reduced, which suggests the existence of a common mechanism between languages to devote as large an amount of cognitive resources as possible to the core of spoken language (i.e., meaning).

Experiment 3 was a priming experiment using synthesized speech and natural human speech with NNSs, where analyses were performed according to their proficiency levels. The results supported the conclusions of Experiment 1. In addition, when using natural voices, the same results as in previous research were shown, confirming that sensitivity to sensory information, which is a feature of implicit memory, can also be seen in perceptual information processing in L2. In Experiment 1, the fact that the post-semantic processing priming effect was statistically significantly higher in NNSs than in NSs turned out to be a reflection of the perceptual learning effect of the group with a higher proficiency level. Moreover, when the proficiency level of the participants was low, the priming effect was high even when synthetic speech was used, and it was found that a certain perceptual learning effect could be expected.

The results of all three experiments (Experiment 1 ~ 3) show that the characteristics of implicit memory indicated in previous research, such as the direction of focus, differences in speech (synthesized or natural), and the speaker, had a significant influence on the priming effect for NNSs. Moreover, even with the same direction of focus, the results varied for different combinations, that is, different speech and speakers. Since it is unlikely that abstraction occurs immediately after one exposure, these results seem to point to the existence of exemplars in implicit memory that have not been abstracted. It can be speculated that the reason why paralinguistic information and linguistic information can be processed independently in L1 is due to the existence of some complicated structure obtained through the accumulation of an extensive number of exemplars. Since the number of experiments on L1 in this study was limited, the researcher will limit the conclusion of the present study to supporting EBM as a language model for encouraging implicit memory usage in L2 learning.

Experiments 4 and 5 were unlike the first three. Perceptual learning training through repetition was conducted to examine the effects of perceptual learning over time, changes in learner's perceptual processing, and the degree of retention in memory. The results of Experiment 4 showed an improvement in perceptual processing after a brief amount of training. When synthesized speech was used, that is, when the acoustic cognitive load was not very high, the results showed that perceptual learning was deepened by vocal repetition following semantic processing. Moreover, from the viewpoint of learning effects, the quantity of (the number of repetition) speech independently increased learner responses and seemed to build word familiarity, while the quality of speech input (the repetition method and processing orientation) had a complementary relationship.

Experiment 5 mimicked an actual L2 vocabulary learning situation and measured the degree of retention using recognition tasks – a type of explicit memory task. Similarly to the results of experiment 4, the differences between NNSs and NSs suggested that for NNSs a combination of independent variables and vocalization after semantic processing increases the retention in memory. This could be due to the phonological instability of L2 word representation. It is important to note that the results of online and offline repetition experiments suggested that overt rehearsal can be efficient only after elaboration of L2 word representation.

Chapter 8
Conclusion

8.1 Summary of Key Findings and Pedagogical Implications

The ability to derive what one has learned in L2 education without having to consciously recall it from long-term memory (that is, using implicit memory) is indispensable in verbal communication. The results of the experiments in this study show that Japanese EFL learners are greatly affected by changes in perceptual information. Hence, one goal for learners should be the formation of robust representations that are not significantly affected by perceptual information, as in L1. Experiment 3 demonstrated that, the group with high proficiency levels was able to respond quicker and showed greater learning effects. When this is applied to EBM and interpreted accordingly, high proficiency learners retained more exemplars in their implicit memory than low proficiency learners; therefore, their processing efficiency can be considered to be high. These results indicate that in order to retain vocabulary information that can be processed quickly without being affected by various perceptual factors at the level of implicit memory, the number of exemplars should be increased. As mentioned above, since previous research has shown that implicit memory is not easily influenced by the passage of time or aging, it can be assumed that increasing the number of exemplars for L2 learners is important regardless of age. Even if the results are not immediately apparent in the form of explicit tests or grades, if exemplars are accumulated in implicit memory, they will likely have a large long-term influence on L2 learning. Considering that the perceptual learning effect disappears when L2 learners, especially lower proficiency learners, concentrate on meaning in high cognitive load situations (i.e., using natural human speech, see Figure 9), ensuring that learners are allowed some time to listen without comprehending and are provided with some variations of auditory input appears to be essential. English education in primary schools, in particular, could be made easier from the junior high school level by focusing on increasing the variations and quantity of speech input, that is, by increasing the number of exemplars. Moreover, Experiments 4 and 5 demonstrated the importance of combining vocal repetition and elaboration tasks, which can strengthen the ability to process perceptual information.

8.2 Limitation of the Study and Further Research

One limitation of this study is that the focus was only on the processing of vocabulary. This was based on the assumption that words are phonetically important units of verbal recall for Japanese EFL learners and following previous research in the field. Thus, only the applicability of EBM was verified through the present experiments, while that of UBM remains unverified. The validity of UBM should also be verified through future research to clarify whether the hypothesis that our linguistic knowledge consists of abstracted 'schemas' is accurate or not. Moreover, perceptual learning training in Experiments 4 and 5 was conducted using methods that were considered to put the lowest cognitive load on Japanese EFL learners; i.e., using synthetic speech with no voice variability. Comparative studies implementing training using vocal variations, including changes in speakers, are required to confirm the arguments presented here.

Perception is likely to be the basis of all cognitive processing. The researcher further argues that the present study succeeded in demonstrating that the perceptual dimension of implicit memory is also closely related to L2 learning. While it is interesting to think that the accumulation of each trivial exemplar is what constitutes the foundation of human knowledge, it is the researcher's opinion that the weight of responsibility regarding choosing the kind of exemplars to be provided lies with the educators.

Relevant Research

Experiment 1

Matsuda, N. (2013). Second-language speech processing: Auditory word priming in Japanese EFL learners and native English speakers, *Journal of the Japan Society for Speech Sciences, 14*, 43–62.

Experiment 3

Matsuda, N. (2017). Evidence of the effects of text-to-speech synthetic speech to improve second language learning, *JACET Journal, 61*, 149–164.

Experiment 4

Matsuda, N. (2012). Effects of auditory word repetition on speech processing of Japanese EFL learners, *Language Education & Technology, 49*, 143–172.

Experiment 5

Matsuda, N. (2017). Effects of encoding factors on word memory retrieval: Differences between native English speakers and Japanese EFL learners. In Y. Maruhashi, M. Hidaka, & M. Nishiyama (Eds.), Collected essays on comparative studies: Bridges between cultures (pp. 108–123). Tokyo, Japan: Eikosha.

All the above publications were revised and edited for the study.

References

Allan, D. (2004). *Oxford placement test 2.* Oxford, England: Oxford University Press.

Allen, J. P. B. (1983). A three-level curriculum model for second language education. *The Canadian Modern Language Review 40,* 23–43.

Ames, H., & Grossberg, S. (2008). Speaker normalization using cortical strip maps: A neural model for steady-state vowel categorization. *The Journal of the Acoustical Society of America, 124*, 3918–3936.

Anderson, J. R. (1976). *Language, memory and thought*. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Erlbaum.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 239–289). San Francisco, CA: Freeman.

Axmear, E., Reichle, J., Alamsaputra, M., Kohnert, K., Drager, K., & Sellnow, K. (2005). Synthesized speech intelligibility in sentences: A comparison of monolingual English-speaking and bilingual children. *Language, Speech, and Hearing Services in Schools, 36*, 244–250.

Azuma, J. (2010). Impact of TTS technology on foreign language teaching: New horizons of multimedia teaching material development. *Ryutsu Kagaku Daigaku Kyoiku Koudoka Suishin Center Kiyou, 6*, 1–11.

Bäckman, L., & Nilsson, L. G. (1984). Aging effects in free recall: An exception to the rule. *Human Learning: Journal of Practical Research & Applications, 3*, 53–69.

Bäckman, L., & Nilsson, L. G. (1985). Prerequisites for lack of age differences in memory performance. *Experimental Aging Research, 11*, 67–73.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The psychology of learning and motivation* (Vol.8, pp. 47–89). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *International Review of Applied Linguistics in Language Teaching, 44*, 231–250.

Bassili, J. N., Smith, M. C., & MacLeod, C. M. (1989). Auditory and visual word-stem completion: Separating data-driven and conceptually driven processes. *The Quarterly Journal of Experimental Psychology Section A, 41*, 439–453.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B, 57*, 289–300.

Boersma P., & Weenink D. J. M. (2011). Praat: doing phonetics by computer (Version 5.2.13) [Computer program]. Retrieved from http://www.praat.org/

Bower, G. H., & Winzenz, D. (1970). Comparison of associative learning strategies. *Psychonomic Science, 20*, 119–120.

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

Bregman, M. R., & Creel, S. C. (2014). Gradient language dominance affects talker learning. *Cognition, 130*, 85–95.

Chevalier, N. (2015). The development of executive function: Toward more optimal coordination of control with age. *Child Development Perspectives, 9*, 239–244.

Chung, C., & Light, L. L. (2009). Effects of age and study repetition on plurality discrimination. *Aging, Neuropsychology, and Cognition*, *16*, 446–460.

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. *Journal of Memory and Language, 39*, 523–542.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 521–533.

Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., & Gilchrist, A. L. (2010). Seven-year-olds allocate attention like adults unless working memory is overloaded. *Developmental science, 13*(1), 120–133.

Craik, F. I. (2002). Levels of processing: Past, present … and future? *Memory*, *10*, 305–318. doi:10.1080/09658210244000135

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684.

Craik, F. I., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge, UK: Cambridge University Press.

Delaney, P. F., Reder, L. M., Staszewski, J. J., & Ritter, F. E. (1998). The strategy-specific nature of improvement: The power law applies by strategy within task. *Psychological Science, 9*, 1–7.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review, 13*, 1–21.

Dobbins, I. G., Kroll, N. E., & Yonelinas, A. P. (2004). Dissociating familiarity from recollection using rote rehearsal. *Memory and Cognition, 32*, 932–944.

Drager, D. R. K., Reichle, J., & Pinkoski. C. (2010). Synthesized speech output and children: A scoping review. *American Journal of Speech-Language Pathology, 19*, 259–273.

Educational Testing Service (2015). Test and score data summary for TOEFL internet-based and paper-based tests (January 2015-December 2015 test data). Retrieved from https://www.ets.org/s/toefl/pdf/94227_unlweb.pdf#search=%27educational+testing+service+test+data+2016%27

Field, J. (2008). *Listening in the Language Classroom.* Cambridge: Cambridge University Press.

Friend, M., & Bryant, J. B. (2000). A developmental lexical bias in the interpretation of discrepant messages. *Merrill-Palmer Quarterly, 46*(2), 342–369.

Gahl, S., & Yu, A. C. (2006). Introduction to the special issue on exemplar-based models in linguistics. *The linguistic review, 23*, 213–216.

Gardiner, J. M., & Richardson-Klavehn, A. (2000) Remembering and Knowing. In E. Tulving (Ed.), *The Oxford handbook of memory* (pp. 229–244). New York, NY: Oxford University Press.

Globalvoice English Professional (Version 2.0.1) [Computer software]. Tokyo, Japan: Hoya.

Goh, C. M. (2000). A cognitive perspective on language learners' listening comprehention problems. *System, 28*, 55–75.

Goldinger, S.D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echo of echoes? An episodic theory of lexical access. *Psychological Review. 105*, 251–279.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 501–518.

Grant, S. C., & Logan, G. D. (1993). The loss of repetition priming and automaticity over time as a function of degree of initial learning. *Memory & Cognition, 21*, 611–618.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist, 39*, 1372–1388.

Hasher, L., Zacks, R. T., Rose, K. C., & Sanft, H. (1987). Truly incidental encoding frequency information. *The American Journal of Psychology, 100*, 69–91.

Hawkins, S. (2010). Phonetic variation as communicative system: Perception of the particular and the abstract. *Laboratory Phonology, 10*, 479–510.

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically-rich approach to speech understanding. *Italian Journal of Linguistics, 13*, 99–188.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*, 411–428.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes and familiarity and recall. *Journal of Memory and Language, 33*, 1–18.

Hirai, A., & O'ki. T. (2011). Comprehensibility and Naturalness of Text-To-Speech Synthetic Materials for EFL Listeners. *JACET Journal, 53*, 1–17.

Houston, D. M., & Jusczyk, P. W. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1570–1582.

Hu, Z., Liu, H., & Zhang, J. X. (2010). Effects of material emotional valance on the time course of massive repetition priming. *Journal of Psycholinguistic Research, 39*, 199–211. doi: 10.1007/s10936-009-9135-1

Ikeda, S. & Haryu, E. (2016). *Hatsuwa karano kanjyouhandan ni okeru rekishikarubaiasu: sono hattatsuteki kijyo wo megutte* [Young children's judgments of speaker affect: The developmental mechanism of the appearance and disappearance of lexical bias]. *Ninchi Kagaku* [Cognitive Studies]*, 23*(1), 22–36.

Ishikawa, S., & Ishikawa, Y. (2008). L2 proficiency and word perception: An fMRI-based study. *ARELE, 19*, 131–140.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30*, 513–541.

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language, 35*, 32–52.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 122*, 139–154.

Jerger, S., Martin, R. C., & Pirozzolo, F. J. (1988). A developmental study of the auditory Stroop effect. *Brain and Language, 35*(1), 86–104.

JET Programme. (2014). Introduction. Official Homepage of the Japan Exchange and Teaching Programme. Retrieved from www.jetprogramme.org/e/introduction/index.html

Johnson, K. (2005). Speaker Normalization in speech perception. In D.B. Pisoni & R. Remez (Eds.), *The handbook of speech perception* (pp. 363–389). Oxford: Blackwell Publishers.

Johnson, K. (2006). Resonance in an exemplar-based lexicon: The emergence of social identity and phonology. *Journal of Phonetics, 34*, 485–499.

Jones, C., Berry, L., & Stevens, C. (2007). Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners. *Computer Speech and Language, 21*, 641–651.

Jusczyk, P. W., Pisoni, D. B., & Mullennix, J. (1992). Some consequences of stimulus variability on speech processing by 2-month-old infants. *Cognition, 43*, 253–291.

Kadota, S. (2007). *Shadouingu to ondoku no kagaku* [The science of shadowing and oral reading]. Tokyo, Japan: Cosmopier.

Kadota, S. (2015). *Shadouingu ondoku to eigo komyunikeshon* [The science of shadowing, oral reading and English communication]. Tokyo, Japan: Cosmopier.

Kageyama, H. (2004). *Kageyama Hideo no nou ikiiki otonano keisan purinto* [The sheet of counting problems for adults by Hideo Kageyama for activation of your brain]. Tokyo: Shogakukan.

Kashiwagi, H., Kang, M., & Ohtsuki, K. (2008). Current status and future prospects of application of synthetic speech in foreign language learning, *Journal of the School of Language and Communication Kobe University, 5*, 10–19.

Kataoka, H., & Ito, M. (2013). A Comparative Study on Reading Aloud: Instructions by Text-To-Speech Synthesis Sounds and a High School Japanese English Teacher. *The JACEC Bulletin, 22*(1), 39–54.

Kato, K., & Kakei, K (1988). *Onseichikaku ni okeru washa eno tekiousei no kentou* [Listener adaptability to individual speaker differences in monosyllabic speech perception]. *Nihon Onkyougaku Kaishi* [Acoustical Science and Technology]*, 44*, 180–186.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 701–722.

Kemmer, S., & Barlow, M. (2000). Introduction: A usage-based conception of language. In M. Barlow &. S. Kemmer (Eds.), *Usage-based models of language* (pp. vii-xxviii). Stanford: CSLI Publications.

Kingston, J. (2003). Learning foreign vowels. *Language and Speech, 46*(2–3), 295–348.

Kirsner, K., & Dunn, J. (1985). The perceptual record: A common factor in repetition priming and attribute retention? In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 547–565). Hillsdale, NJ: Erlbaum.

Koul, R., & Clapsaddle, K. C. (2006). Effects of repeated listening experiences on the perception of synthetic speech by individuals with mild-to-moderate disabilities. *Augmentative and Alternative Communication, 22*, 112–122.

Kuroda, K. (2010). *Choujyoukioku shoukougun no rirontekigani* [Theoretical implications of the hyperthymestic syndrome]. *Nihon ninchikagakukai dai 27 kai taikai ronbunshu* [Proceedings of the 27th Annual meeting of the Japanese Cognitive Science Society], 789–792.

Lado, R. (1964). Language teaching: A scientific approach. New York: McGraw-Hill.

Langacker, R. W. (1987). *Foundations of cognitive grammar, vol. 1, Theoretical prerequisites.* Stanford, CA: Stanford University Press.

Langacker, R. W. (2000). A dynamic usage-based model. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of language* (pp. 1– 63). Stanford: CSLI Publications.

Langacker, R. W. (2009). A dynamic view of usage and language acquisition. *Cognitive Linguistics, 20*, 627–640.

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 25–42). Ann Arbor, Michigan: University of Michigan Press.

Light, L. L., Prull, M. W., La Voie, D. J., & Healy, M. R. (2000). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). Oxford, UK: Oxford University Press.

Littlewood, W. (1981). Communicative language teaching. Cambridge: Cambridge University Press.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*, 1242–1255.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*, 492–527.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 319–331.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*, 252–271.

McDonough, K., & Trofimovich, P. (2009). *Using priming methods in second language research*. New York, NY: Routledge.

McLaughlin, B. (1987). *Theories of second-language learning.* London: Edward Arnold.

McNaughton, D., Fallon, K., Tod, J., Weiner, F., & Neisworth, J. (1994). Effect of repeated listening experiences on the intelligibility of synthesized speech. *Augmentative and Alternative Communication, 10*, 161–168.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *Journal of Verbal Learning and Verbal Behavior, 2*, 1–21.

Morris, C., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16*, 519–533.

Nosofsky, R. M. (1991). Tests of an exemplar model for relating perceptual classification and recognition memory. *Journal of Experimental Psychology: Human Perception & Performance, 17*, 3–27.

Ohta, N. (2011). *Kiokukenkyu no rekishi to genjyou* [The history of memory research and the present condition]. In N. Ohta & Y. Itsukushima (Eds.), *Gendai no ninchishinrigaku 2 kioku to nichijyou* [Modern Cognitive psychology 2 memory and daily life] (pp. 2–25). Kyoto, Japan: Kitaojishobou.

Ohta, N. (2016). *Ninchishinrigaku to kyouiku: Kiokukenkyu wo chushin ni* [Cognitive psychology and education: With a focus on memory research]. In N. Ohta & Y. Sakuma (Eds.), *Eigokyouikugaku to Ninchishinrigaku no kurosupointo: Shougakkou kara daigaku madeno eigogakushu wo kangaeru* [The overlap of English language education and cognitive psychology: considering English study from elementary school to college] (pp. 4–22). Kyoto, Japan: Kitaojishobou.

Onishi, K. H., Chambers, K. E., & Fisher, C. (2002). Learning phonotactic constraints from brief auditory experience. *Cognition, 83*, B13–23.

Osaka, M. (2002). *Nou no memochou waking memori* [Working memory: The sketchpad in the brain]. Tokyo: Shinyosha.

Oxford University Press., University of Cambridge., & Association of Language Testers in Europe. (2001). *Quick placement test.* Oxford, England: Oxford University Press.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pilotti, M., Bergman, E. T., Gallo, D. A., Sommers, M., & Roediger, H. L., III (2000). Direct comparison of auditory implicit memory tests. *Psychonomic Bulletin & Review, 7*, 347–353.

Pilotti, M., & Beyer, T. (2002). Perceptual and lexical components of auditory repetition priming in young and older adults. *Memory and Cognition, 30*, 226–236.

Pilotti, M., Gallo, D. A. & Roediger, H. L., III (2000). Effects of hearing words, imaging hearing words, and reading on auditory implicit and explicit memory tests. *Memory & Cognition, 28*, 1406–1418.

Pinkoski-Ball, C., Reichle, J., & Munson, B. (2012). Synthesized speech intelligibility and early preschool-age children: comparing accuracy for single-word repetition with repeated exposure. *American Journal of Speech-Language Pathology, 21*, 293–301.

Pisoni, D. B. (1997). Some thoughts on normalization in speech production. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing*, (pp. 9–32). San Diego, CA: Academic Press.

Port, R. (2007). How are words stored in memory? Beyond phones and phonemes. *New Ideas in Psychology, 25*, 143–170.

Reynolds, M., & Jefferson, L. (1999). Natural and synthetic speech comprehension: Comparison of children from two age groups. *Augmentative and Alternative Communication, 15*, 174–182.

Rickard, T. (1997). Bending the power law: A CMPL theory of strategy shifts and the automatization of cognitive skills. *Journal of Experimental Psychology: General, 126*, 288–311.

Roediger, H. L., & McDermott, K. B. (1993). Implicit memory in normal human subjects. *Handbook of neuropsychology, 8*, 63–63.

Rost, G. C., & McMurray, B. (2009). Speaker variability augments phonological processing in early word learning. *Developmental science, 12*, 339–349.

Rost, G. C., & McMurray, B. (2010). Finding the signal by adding noise: The role of noncontrastive phonetic variability in early word learning. *Infancy, 15*, 608–635.

Rugg, M. D., & Yonelinas, A. P. (2003). Human recognition memory: A cognitive neuroscience perspective. *Trends in Cognitive Sciences, 7*, 313–319.

Salasoo, A., Shiffrin, R. M., & Feustel, T. C. (1985). Building permanent memory codes: Codification and repetition effects in word identification. *Journal of Experimental Psychology: General, 114*, 50–77.

Schacter, D.L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 915–930.

Schacter, D. L., & Tulving, E. (1994). What are the memory systems of 1994? In D. L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 1–38). Cambridge, MA: MIT Press.

Schneider, W., & Chein, J. (2003). Controlled & automatic processing: Behavior, theory, and biological mechanisms. *Cognitive Science, 27*, 525–559.

Sheffert, S. M. (1998). Voice-specificity effects on auditory word priming. *Memory and Cognition, 26*, 591–598.

Skinner, E. I., & Fernandes, M. A. (2007). Neural correlates of recollection and familiarity: A review of neuroimaging and patient data. *Neuropsychologia, 45*, 2163–2179.

Sloman, S. A., Hayman, C. A., Ohta, N., Law, J., & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 223–239.

Stark, C. E. L., & McClelland, J. L. (2000). Repetition priming of words, pseudowords, and nonwords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 945–972.

Stenberg, G., Hellman, J., Johansson, M., & Rosén, I. (2009). Familiarity or conceptual priming: Event-related potentials in name recognition. *Journal of Cognitive Neuroscience, 21*, 447–460.

Sugiura, K., & Hori, T. (2012). Auditory priming effect in Japanese learners of English: An investigation using a repetition task of spoken words. *KATE Journal, 26*, 39–51.

Takeuchi, O. (Ed.). (2000). *Ninchiteki apurochi niyoru gaikokugokyoiku* [A cognitive approach to language teaching]. Tokyo: Shohakusha.

Tamai, K. (2005). *Risuningu shidouhou toshiteno shadoingu no kouka ni kansuru kenkyu* [A study on the effectiveness of shadowing as an instructional method of listening]. Tokyo, Japan: Kazama Shobou.

Tanemura, J. (2006). *Shitsugoshou chiryou ni okeru ninchishinkeishinrigakuteki houhou* [Cognitive neuropsychological approaches in aphasia therapy]. *Koujinoukinou Kenkyu* [Higher Brain Function Research]*, 26*, 1–7.

Terasawa, T. (2016). *Senzaikioku to gakushu no jissenteki kenkyu* [Implicit memory and practical studies of learning]. In N. Ohta & Y. Sakuma (Eds.), *Eigokyouikugaku to Ninchishinrigaku no kurosupointo: Shougakkou kara daigaku madeno eigogakushu wo kangaeru* [The overlap of English language education and cognitive psychology: considering English study from elementary school to college] (pp. 37–55). Kyoto, Japan: Kitaojishobou.

Terasawa, T., Yoshida, T., & Ohta, N. (2008). *Eitangogakushu ni okeru jikaku dekinai gakushudankai no kenshutsu* [Unconscious stage of learning in the acquisition of a second language: Experimental designs in long-term everyday learning situations]. *Kyouikushinrigaku Kenkyu* [*Japanese Journal of Educational Psychology*], *56*, 510–522.

Torikai, K., Tamai, K., Someya, Y., Tanaka, M., Tsuruta, C., & Nishimura, T. (2003). *Hajimeteno shadouingu* [Shadowing for the first time]. Tokyo: Gakushu Kenkyusha.

Trofimovich, P. (2005). Spoken-word processing in native and second languages: An investigation of auditory word priming. *Applied Psycholinguistics, 26*, 479–504. doi:10.1017.S0142716405050265

Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research, 37*, 309–329. doi: 10.1007/s10936-008-9069-z

Trofimovich, P., & Gatbonton, E. (2006). Repetition and focus on form in processing L2 Spanish words: Implications for pronunciation instruction. *Modern Language Journal, 90*, 519–535.

Tulving, E. (1979). Relation between encoding specificity and levels of processing. In L. S. Cermak & F. I. Craik (Eds.), *Levels of processing in human memory* (pp. 405–428). Hillsdale, NJ: Erlbaum.

Tulving, E. (1987). Multiple memory systems and consciousness. *Human neurobiology, 6*(2), 67–80.

Tulving, E. (1995). Organization of memory: Quo vadis? In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (pp. 839–847). Cambridge, MA: MIT Press.

Tulving, E., & Schacter, D. L. (1990). Priming and human memory systems. *Science, 19*, 301–306.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 336–342.

Tulving, E., & Thomson, D. M., (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Venkatagiri, H. S. (2005). Phoneme intelligibility of four text-to-speech products to nonnative speakers of English in noise. *International Journal of Speech Technology, 8*, 313–321.

Waidergoren, S., Segalowicz, J. & Gilboa, A. (2012). Semantic memory recognition is supported by intrinsic recollection-like processes: “The butcher on the bus” revisited. *Neuropsychologia, 50*, 3573–3587.

Walsh, D. A. & Jenkins, J, J. (1973). Effects of orienting tasks on free recall in incidental learning: “Difficulty,” “Effort,” and “Process” explanations. *Journal of Verbal Learning and Verbal Behavior, 12*, 481–488.

Warrington, E. K., & Weiskrantz, L. (1974). The effect of prior learning on subsequent retention in amnesic patients. *Neuropsychologia, 12*, 419–428.

Watanabe, S., Iwaki, T., Kaneysu, T., & Miki, K. (2006). Corpus-based Text-to-speech and its application. *Oki Technical Review, 206*(73-2), 62–65.

Westerberg, C. E., Paller, K. A., Weintraub, S., Mesulam, M. M., Holdstock, J. S., Mayes, A. R., & Reber, P. J. (2006). When memory does not fail: Familiarity-based recognition in mild cognitive impairment and Alzheimer’s disease. *Neuropsychology, 20*, 193–205.

Wixted, J. T. (2007). Dual-process theory and signal-detection theory of recognition memory. *Psychological Review, 114*, 152–176.

Woutersen, M., Cox, A., Weltens, B., & de Bot, K. (1994). Lexical aspects of standard dialect bilingualism. *Applied Psycholinguistics, 15*, 447–473.

Woutersen, M., de Bot, K., & Weltens, B. (1995). The bilingual lexicon: Modality effects in processing. *Journal of Psycholinguistic Research, 24*, 289–298.

Yamamoto, H. & Haryu, E. (2016). *Youji no tangogakushu ni okeru akusento patan riyou no hattatsu katei* [Japanese children's learning of homophones with different accentual patterns]. *Ninchi Kagaku* [Cognitive Studies]*, 23*(1), 22–36.

Yokokawa, H., Ikemura, D., Kadota, S., Sakai, M., Shimamoto, T., Tanimura, M., Nakanishi, Y., …Yoshimura, M. (2006). *Kyouikukenkyu no tameno dainigengo detabesu: Nihonjin eigogakushusha no eitangoshinmitsudo mojihen* [The data base of a second language for teaching and research: Written English word familiarity of Japanese EFL learners]. Tokyo, Japan: Kuroshioshuppan.

Yokokawa, H., Ikemura, D., Kadota, S., Satoi, H., Shimamoto, T., Sugiura, K., Tanimura, M., …Yoshimura, M. (2009). *Kyouikukenkyu no tameno dainigengo detabesu: Nihonjin eigogakushusha no eitangoshinmitsudo onseihe*n [The data base of a second language for teaching and research: Spoken English word familiarity of Japanese EFL learners]. Tokyo, Japan: Kuroshioshuppan.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.

Yow, W. Q., & Markman, E. M. (2011). Bilingualism and children's use of paralinguistic cues to interpret emotion in speech. *Bilingualism: Language and Cognition, 14*, 562–569.

Appendix A

Pair Words for Task 1

Group	Word	Familiarity	Frequency (BNC)	Syllable number	Group	Pair word	Familiarity	Frequency (BNC)	Syllable number	Correct response
V1	again	6.10	59829	2	V3	campaign	5.44	11841	2	Yes
V1	below	5.13	14335	2	V3	although	5.26	43635	2	Yes
V1	contrast	5.49	8172	2	V3	final	5.69	16468	2	No
V1	design	5.27	26375	2	V3	advance	5.78	8087	2	No
V1	expect	5.55	27221	2	V3	respect	6.38	12627	2	Yes
V1	feature	6.37	17219	2	V3	teacher	6.82	19744	2	Yes
V1	husband	5.60	12263	2	V3	section	5.90	23188	2	No
V1	listen	6.34	12080	2	V3	order	6.13	45595	2	No
V1	myself	6.08	12444	2	V3	yourself	6.50	10746	2	Yes
V1	nearly	5.28	11484	2	V3	merely	5.20	7596	2	Yes
V1	open	6.13	46095	2	V3	special	6.71	22040	2	No
V1	project	6.34	21648	2	V3	correct	6.32	7711	2	Yes
V1	present	6.40	36806	2	V3	extent	5.16	10071	2	Yes
V1	recent	5.79	15812	2	V3	future	6.19	24055	2	No
V1	sentence	5.54	10127	2	V3	people	6.09	125430	2	No
V1	water	6.70	35767	2	V3	daughter	4.98	11522	2	Yes
V1	addition	5.97	10664	3	V3	condition	5.38	23742	3	Yes
V1	character	5.34	12511	3	V3	family	5.43	42773	3	No
V1	energy	5.94	13083	3	V3	manager	6.29	19636	3	No
V1	however	6.25	60498	3	V3	yesterday	6.15	19459	3	No
V1	influence	5.21	15130	3	V3	government	5.53	66894	3	No
V1	probably	6.05	27303	3	V3	anyway	6.62	12232	3	No
V1	remember	6.79	26748	3	V3	November	6.15	9400	3	Yes
V1	information	6.68	38656	4	V3	population	6.21	14664	4	Yes
V2	about	6.41	197115	2	V3	without	6.35	45867	2	Yes
V2	believe	5.40	34603	2	V3	escape	5.72	7509	2	No
V2	cover	6.11	24698	2	V3	visit	5.69	22091	2	No
V2	copy	6.59	10562	2	V3	century	6.17	23259	2	No
V2	depend	5.65	10125	2	V3	attend	6.41	8801	2	Yes
V2	either	5.16	27766	2	V3	neither	5.32	8245	2	Yes
V2	express	6.00	12519	2	V3	success	5.00	14330	2	Yes
V2	heavy	5.11	10439	2	V3	female	5.51	10090	2	No
V2	into	5.45	163469	2	V3	after	5.50	116794	2	No
V2	language	6.29	22117	2	V3	action	6.48	26894	2	No
V2	mother	6.60	27784	2	V3	other	6.36	185308	2	Yes
V2	okay	6.40	12190	2	V3	today	6.88	25775	2	Yes
V2	reduce	5.32	17226	2	V3	produce	6.36	30295	2	Yes
V2	science	5.93	12644	2	V3	favor	5.47	9308	2	No
V2	table	6.22	23092	2	V3	able	5.44	30454	2	Yes
V2	wonder	6.17	14375	2	V3	under	6.44	61925	2	Yes
V2	arrangement	5.25	9054	3	V3	performance	5.42	14620	3	No
V2	collection	6.67	9639	3	V3	protection	5.55	8025	3	Yes
V2	difficult	6.21	22033	3	V3	national	6.73	37231	3	No
V2	exactly	6.46	10729	3	V3	tomorrow	6.35	9243	3	No
V2	opinion	5.21	9213	3	V3	behavior	5.05	12853	3	No
V2	position	5.48	28071	3	V3	original	6.13	11610	3	No
V2	September	6.60	10400	3	V3	December	6.30	9400	3	Yes
V2	application	5.79	16281	4	V3	education	6.13	25987	4	Yes
Average		5.93	27258.63	2.38			5.94	28438.96	2.38	

Appendix B

Pair Words for Task 2

Group	Word	Familiarity	Frequency (BNC)	Syllable number	Group	Pair word	Familiarity	Frequency (BNC)	Syllable number	Correct response
V4	control	6.13	38281	2	V6	manage	6.13	14894	2	Yes
V4	concept	5.93	9093	2	V6	idea	6.65	32798	3	Yes
V4	decide	5.10	24380	2	V6	exist	5.32	11515	2	No
V4	ever	6.55	27195	2	V6	sometimes	5.90	20517	2	No
V4	effect	5.83	34881	2	V6	result	4.93	42171	2	Yes
V4	father	6.33	23216	2	V6	value	5.92	26887	2	No
V4	image	6.53	11024	2	V6	picture	6.60	17023	2	Yes
V4	measure	6.07	17443	2	V6	standard	6.13	21744	2	Yes
V4	over	5.73	135170	2	V6	above	6.17	25747	2	Yes
V4	person	5.68	28981	2	V6	human	5.98	21620	2	Yes
V4	publish	5.07	12242	2	V6	release	4.94	12851	2	Yes
V4	police	5.97	27508	2	V6	inside	6.20	14094	2	No
V4	paper	6.71	23694	2	V6	nothing	6.54	34064	2	No
V4	response	5.20	14627	2	V6	answer	6.09	22736	2	Yes
V4	technique	5.94	10548	2	V6	hotel	5.64	11683	2	No
V4	window	6.78	19340	2	V6	spirit	6.08	8384	2	No
V4	computer	6.40	16976	3	V6	experience	6.68	29191	4	No
V4	exercise	6.52	12721	3	V6	training	5.33	13503	2	Yes
V4	history	6.11	20064	3	V6	interest	6.20	39629	3	No
V4	official	5.49	15931	3	V6	public	5.69	38394	2	Yes
V4	possible	5.59	34178	3	V6	natural	6.11	14304	3	No
V4	realize	6.05	15575	3	V6	understand	6.56	24252	3	Yes
V4	separate	5.66	12159	3	V6	physical	5.31	9569	3	No
V4	ability	5.20	10468	4	V6	society	5.13	28150	4	No
V5	approach	6.17	23763	2	V6	access	5.62	11488	2	Yes
V5	agree	6.31	23497	2	V6	explain	6.15	19218	2	No
V5	better	6.27	15626	2	V6	always	6.45	46228	2	No
V5	country	6.37	48177	2	V6	nation	5.75	8508	2	Yes
V5	degree	5.22	12996	2	V6	July	5.37	11900	2	No
V5	exchange	6.16	11054	2	V6	forget	5.66	12353	2	No
V5	happy	6.42	12854	2	V6	sorry	6.38	11453	2	No
V5	include	5.50	34753	2	V6	appear	5.77	30595	2	No
V5	major	6.04	23629	2	V6	great	5.85	64369	1	Yes
V5	message	6.47	8938	2	V6	software	5.73	9134	2	No
V5	number	6.84	60607	2	V6	system	6.57	61912	2	No
V5	only	6.10	152903	2	V6	alone	5.08	13350	2	Yes
V5	promise	6.06	10432	2	V6	office	5.85	29943	2	No
V5	subject	5.03	32392	2	V6	issue	5.18	35021	2	Yes
V5	supply	6.20	16892	2	V6	provide	5.48	47923	2	Yes
V5	woman	5.05	63087	2	V6	lady	5.13	9739	2	Yes
V5	argument	5.35	12125	3	V6	discussion	6.36	11315	3	Yes
V5	advantage	5.76	10285	3	V6	benefit	5.73	19513	3	Yes
V5	another	5.38	60182	3	V6	different	6.13	48373	3	Yes
V5	document	6.54	10498	3	V6	newspaper	6.73	8544	3	No
V5	holiday	6.30	9731	3	V6	telephone	6.20	9403	3	No
V5	interview	5.38	9008	3	V6	question	6.85	43178	2	Yes
V5	occasion	5.75	9152	3	V6	development	5.57	37386	4	No
V5	actually	5.73	25990	4	V6	really	6.62	48062	3	Yes
Average		5.94	26963.88	2.38			5.93	24679.75	2.33	

Appendix C

Unrepeated Words for Task 3 (V7) and Task 4 (V8)

Group	Unrepeated word	Familiarity	Frequency (BNC)	Syllable number	Duration (ms)	Group	Unrepeated word	Familiarity	Frequency (BNC)	Syllable number	Duration (ms)
V7	address	5.09	11984	2	745.90	V8	attack	5.94	16549	2	694.05
V7	allow	5.40	33687	2	708.60	V8	arrive	6.51	14093	2	700.99
V7	basic	6.04	10860	2	604.18	V8	body	5.00	32231	2	645.46
V7	couple	5.64	15330	2	522.13	V8	culture	6.07	10196	2	666.29
V7	career	5.36	9441	2	723.50	V8	finish	6.77	13902	2	596.88
V7	easy	6.57	21480	2	512.68	V8	morning	6.68	21845	2	645.46
V7	island	6.45	7649	2	604.18	V8	market	5.45	36905	2	596.88
V7	money	6.72	37892	2	560.05	V8	nature	6.16	18223	2	666.29
V7	movement	5.94	17880	2	708.60	V8	problem	5.96	56483	2	638.52
V7	power	5.08	38824	2	589.26	V8	profit	5.02	11944	2	645.46
V7	party	6.51	52979	2	566.88	V8	prevent	5.64	10286	2	638.52
V7	report	5.06	51517	2	671.31	V8	practice	6.25	24019	2	784.27
V7	Sunday	6.51	10100	2	753.36	V8	ready	6.31	10110	2	569.12
V7	suppose	5.63	14482	2	835.41	V8	surface	5.11	10361	2	749.57
V7	target	6.00	10110	2	611.64	V8	support	5.87	40248	2	714.87
V7	welcome	6.67	9570	2	675.59	V8	worker	5.23	18247	2	596.88
V7	anyone	6.30	14956	3	663.85	V8	animal	6.16	15250	3	617.70
V7	area	6.17	58449	3	589.26	V8	attention	6.34	13968	3	784.27
V7	difference	5.15	19138	3	790.65	V8	everything	6.27	18675	3	742.63
V7	expression	5.18	8756	3	835.41	V8	hospital	5.91	16898	3	728.75
V7	October	5.61	10600	3	775.73	V8	imagine	6.64	8300	3	687.11
V7	restaurant	6.82	5100	3	709.83	V8	reaction	5.73	7565	3	839.80
V7	serious	6.00	12232	3	835.41	V8	radio	6.50	9066	3	756.51
V7	activity	6.48	23105	4	745.90	V8	interested	5.25	7605	4	700.99
Average		5.93	21088.38	2.38	680.80			5.95	18457.04	2.38	683.64

Appendix D

Unrepeated Words for Task 5 (V9) and Task 6 (V10)

Group	Unrepeated word	Familiarity	Frequency (BNC)	Syllable number	Duration (ms)	Group	Unrepeated word	Familiarity	Frequency (BNC)	Syllable number	Duration (ms)
V9	across	5.23	25202	2	806.64	V10	around	5.82	45286	2	760.18
V9	account	5.31	25390	2	701.43	V10	begin	5.42	43740	2	597.75
V9	announce	6.09	12582	2	764.56	V10	contact	5.46	13867	2	753.68
V9	become	5.25	67219	2	687.40	V10	county	5.36	9745	2	565.26
V9	brother	5.87	11757	2	540.10	V10	follow	5.60	46145	2	558.77
V9	council	5.78	34496	2	715.46	V10	leader	6.30	15903	2	623.74
V9	even	5.61	90473	2	568.16	V10	music	6.98	15024	2	552.27
V9	happen	6.08	32075	2	610.24	V10	maybe	5.70	10472	2	636.73
V9	little	5.70	63383	2	512.04	V10	process	6.16	30120	2	792.67
V9	machine	6.09	13518	2	715.46	V10	prepare	5.20	14961	2	636.73
V9	never	6.50	55899	2	561.14	V10	program	6.48	32068	2	753.68
V9	real	5.27	22982	2	596.21	V10	research	6.30	27663	2	714.70
V9	summer	6.71	11563	2	603.23	V10	receive	5.86	24111	2	688.71
V9	sample	5.16	8182	2	715.46	V10	station	6.31	12328	2	721.20
V9	suggest	5.71	28665	2	932.90	V10	travel	6.01	12288	2	649.73
V9	worry	6.19	9006	2	589.20	V10	weather	6.54	42042	2	493.79
V9	anything	6.62	28321	3	701.43	V10	company	5.20	57754	3	597.75
V9	already	6.24	34292	3	785.60	V10	everyone	6.71	13337	3	695.21
V9	accident	6.36	8374	3	736.50	V10	encourage	5.63	10664	3	930.45
V9	doctor	5.62	13684	2	610.24	V10	production	6.00	15837	3	701.71
V9	example	6.87	43402	3	694.41	V10	recover	5.99	4932	3	662.72
V9	important	5.56	39265	3	708.44	V10	recognize	5.57	15203	3	892.21
V9	relation	5.82	19628	3	764.56	V10	Saturday	5.93	8700	3	753.68
V9	interesting	6.49	9624	4	715.46	V10	independent	5.55	8968	4	799.80
Average		5.92	29540.92	2.33	680.68			5.92	22131.58	2.38	688.88

Appendix E

Questionnaire for Participants

アンケート

名前 ＿＿＿＿＿＿＿＿＿＿＿＿

1. 性別・及び年齢を教えてください。

男性・女性　　(どちらかに○をつけてください)

満 ＿＿＿＿＿ 歳

2. 英語を勉強し始めたのは何歳ですか？

満 ＿＿＿＿＿ 歳

3. 学校で英語を勉強し始めたのは何歳ですか？

満 ＿＿＿＿＿ 歳

4. 海外での生活経験はありますか？

はい・いいえ　　(どちらかに○をつけてください)

5. 4 で「はい」と答えた方にお聞きします。滞在された国名と期間を教えてください。

国名（複数回答可）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

期間（複数回答可）＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

6. ご自分の英語の習熟度を 10 段階で評価してください。
（1：ほとんど○○できない　～　10：英語母語話者並みに○○できる）

リスニング ＿＿＿＿＿　　スピーキング ＿＿＿＿＿

リーディング ＿＿＿＿＿　　ライティング ＿＿＿＿＿

7. 実用英語技能検定（英検）、TOEIC や TOEFL などの英語のテストを受けたことがある方はテストの合否または点数を教えてください。（複数回答可）

受験した年（西暦）＿＿＿＿＿ 年に ＿＿＿＿＿＿＿＿＿ を受験

合否または点数 ＿＿＿＿＿ （点）

4. 実験の感想などがありましたら、お願いいたします。

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿＿

以上です。本研究へのご理解とご協力、本当に有難うございました。

Appendix F

Words for Repetition Task after Rhyme Judgment Task (Same Voice)

Word	Voice (Study)	Voice (Test)	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
			Repeated Word				
contrast	Female	Female	5.49	5.11	8172	2	857.78
energy	Female	Female	5.94	5.42	13083	3	671.31
husband	Female	Female	5.60	5.08	12263	2	634.01
influence	Female	Female	5.21	5.23	15130	3	790.65
myself	Female	Female	6.08	5.91	12444	2	611.64
open	Female	Female	6.13	6.77	46095	2	566.88
remember	Female	Female	6.79	6.07	26748	3	663.85
sentence	Female	Female	5.54	5.89	10127	2	880.16
water	Female	Female	6.70	6.77	35767	2	604.18
addition	Male	Male	5.97	4.70	10664	3	693.69
character	Male	Male	5.34	6.21	12511	3	677.27
design	Male	Male	5.27	5.75	26375	2	708.60
feature	Male	Male	6.37	5.28	17219	2	559.42
however	Male	Male	6.25	5.73	60498	3	589.26
listen	Male	Male	6.34	6.17	12080	2	618.66
present	Male	Male	6.40	5.85	36806	2	686.23
probably	Male	Male	6.05	5.63	27303	3	716.06
recent	Male	Male	5.79	5.25	15812	2	648.93
Average			5.96	5.71	22172.06	2.39	676.59
			Unrepeated Word				
address	Female	Female	5.09	6.41	11984	2	745.90
basic	Female	Female	6.04	6.05	10860	2	604.18
career	Female	Female	5.36	5.00	9441	2	723.50
difference	Female	Female	5.15	6.48	19138	3	790.65
island	Female	Female	6.45	5.85	7649	2	604.18
movement	Female	Female	5.94	5.05	17880	2	708.60
October	Female	Female	5.61	5.88	10600	3	775.73
party	Female	Female	6.51	6.48	52979	2	566.88
welcome	Female	Female	6.67	6.30	9570	2	675.59
anyone	Male	Male	6.30	5.68	14956	3	555.22
area	Male	Male	6.17	6.00	58449	3	589.26
couple	Male	Male	5.64	5.54	15330	2	522.13
easy	Male	Male	6.57	6.80	21480	2	512.68
expression	Male	Male	5.18	4.80	8756	3	835.41
money	Male	Male	6.72	6.61	37892	2	560.05
restaurant	Male	Male	6.82	5.90	5100	3	709.83
Sunday	Male	Male	6.51	6.53	10100	2	753.36
suppose	Male	Male	5.63	5.25	14482	2	835.41
Average			6.02	5.92	18702.56	2.33	670.47

Appendix G

Words for Repetition Task after Synonym Judgment Task (Same Voice)

Word	Voice (Study)	Voice (Test)	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
			Repeated Word				
control	Female	Female	6.13	6.49	38281	2	673.37
decide	Female	Female	5.10	5.31	24380	2	694.41
father	Female	Female	6.33	6.29	23216	2	652.33
history	Female	Female	6.11	6.13	20064	3	596.21
measure	Female	Female	6.07	4.48	17443	2	645.31
official	Female	Female	5.49	6.22	15931	3	638.30
paper	Female	Female	6.71	6.50	23694	2	575.17
response	Female	Female	5.20	5.08	14627	2	848.73
window	Female	Female	6.78	6.18	19340	2	631.29
computer	Male	Male	6.40	6.80	16976	3	687.40
concept	Male	Male	5.93	5.18	9093	2	743.51
ever	Male	Male	6.55	6.16	27195	2	512.04
exercise	Male	Male	6.52	5.78	12721	3	960.96
image	Male	Male	6.53	6.54	11024	2	603.23
over	Male	Male	5.73	6.55	135170	2	568.16
police	Male	Male	5.97	6.27	27508	2	708.44
possible	Male	Male	5.59	6.27	34178	3	687.40
realize	Male	Male	6.05	5.50	15575	3	813.66
Average			6.07	5.99	27023.11	2.33	680.00
			Unrepeated Word				
accident	Female	Female	6.36	6.02	8374	3	736.50
brother	Female	Female	5.87	6.15	11757	2	540.10
council	Female	Female	5.78	3.07	34496	2	715.46
doctor	Female	Female	5.62	6.38	13684	2	610.24
even	Female	Female	5.61	6.14	90473	2	568.16
happen	Female	Female	6.08	6.02	32075	2	610.24
relation	Female	Female	5.82	5.17	19628	3	764.56
sample	Female	Female	5.16	5.92	8182	2	715.46
worry	Female	Female	6.19	6.16	9006	2	589.20
already	Male	Male	6.24	5.73	34292	3	785.60
anything	Male	Male	6.62	6.31	28321	3	701.43
example	Male	Male	6.87	6.11	43402	3	694.41
important	Male	Male	5.56	6.67	39265	3	708.44
little	Male	Male	5.70	6.47	63383	2	512.04
machine	Male	Male	6.09	6.37	13518	2	715.46
real	Male	Male	5.27	6.15	22982	2	596.21
suggest	Male	Male	5.71	5.30	28665	2	932.90
summer	Male	Male	6.71	6.73	11563	2	603.23
Average			5.96	5.94	28503.67	2.33	672.20

Appendix H

Words for Repetition Task after Rhyme Judgment Task (Different Voice)

Word	Voice (Study)	Voice (Test)	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
			Repeated Word				
copy	Female	Male	6.59	5.79	10562	2	589.94
difficult	Female	Male	6.21	6.63	22033	3	714.87
either	Female	Male	5.16	5.39	27766	2	548.30
into	Female	Male	5.45	5.82	163469	2	610.76
mother	Female	Male	6.60	6.60	27784	2	610.76
opinion	Female	Male	5.21	5.58	9213	3	694.05
reduce	Female	Male	5.32	4.61	17226	2	687.11
science	Female	Male	5.93	5.74	12644	2	812.03
wonder	Female	Male	6.17	5.87	14375	2	596.88
collection	Male	Female	6.67	5.91	9639	3	749.57
cover	Male	Female	6.11	5.75	24698	2	513.59
depend	Male	Female	5.65	5.34	10125	2	652.40
exactly	Male	Female	6.46	5.43	10729	3	791.21
heavy	Male	Female	5.11	5.92	10439	2	471.95
language	Male	Female	6.29	6.35	22117	2	909.20
okay	Male	Female	6.40	4.98	12190	2	569.12
position	Male	Female	5.48	6.27	28071	3	659.34
arrangement	Male	Female	5.25	4.24	9054	3	867.56
Average			5.89	5.68	24563.00	2.33	669.37
			Unrepeated Word				
animal	Female	Male	6.16	6.25	15250	3	617.70
attention	Female	Male	6.34	5.57	13968	3	784.27
body	Female	Male	5.00	6.35	32231	2	645.46
finish	Female	Male	6.77	6.53	13902	2	596.88
hospital	Female	Male	5.91	6.15	16898	3	728.75
market	Female	Male	5.45	6.11	36905	2	596.88
practice	Female	Male	6.25	5.92	24019	2	784.27
reaction	Female	Male	5.73	5.76	7565	3	839.80
support	Female	Male	5.87	6.13	40248	2	714.87
culture	Male	Female	6.07	6.55	10196	2	666.29
everything	Male	Female	6.27	6.44	18675	3	742.63
imagine	Male	Female	6.64	5.65	8300	3	687.11
morning	Male	Female	6.68	6.55	21845	2	645.46
prevent	Male	Female	5.64	4.66	10286	2	638.52
radio	Male	Female	6.50	6.38	9066	3	756.51
ready	Male	Female	6.31	5.89	10110	2	569.12
surface	Male	Female	5.11	4.79	10361	2	749.57
worker	Male	Female	5.23	5.89	18247	2	596.88
Average			6.00	5.98	17670.67	2.39	686.72

Appendix I

Words for Repetition Task after Synonym Judgment Task (Different Voice)

Word	Voice (Study)	Voice (Test)	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
			Repeated Word				
advantage	Female	Male	5.76	5.40	10285	3	877.13
degree	Female	Male	5.22	4.43	12996	2	604.25
happy	Female	Male	6.42	6.75	12854	2	545.77
holiday	Female	Male	6.30	6.50	9731	3	656.23
interview	Female	Male	5.38	6.30	9008	3	708.20
major	Female	Male	6.04	6.23	23629	2	643.23
message	Female	Male	6.47	6.58	8938	2	773.18
subject	Female	Male	5.03	6.25	32392	2	708.20
woman	Female	Male	5.05	6.45	63087	2	578.26
another	Male	Female	5.38	6.40	60182	3	643.23
argument	Male	Female	5.35	3.98	12125	3	688.71
better	Male	Female	6.27	6.47	15626	2	558.77
country	Male	Female	6.37	6.35	48177	2	649.73
document	Male	Female	6.54	5.29	10498	3	773.18
include	Male	Female	5.50	5.37	34753	2	727.70
occasion	Male	Female	5.75	4.22	9152	3	734.19
only	Male	Female	6.10	6.62	152903	2	558.77
supply	Male	Female	6.20	5.50	16892	2	695.21
Average			5.84	5.84	30179.33	2.39	673.55
			Unrepeated Word				
company	Female	Male	5.20	6.30	57754	3	597.75
contact	Female	Male	5.46	5.84	13867	2	753.68
county	Female	Male	5.36	4.52	9745	2	565.26
encourage	Female	Male	5.63	4.98	10664	3	930.45
leader	Female	Male	6.30	5.63	15903	2	623.74
maybe	Female	Male	5.70	5.67	10472	2	636.73
production	Female	Male	6.00	5.87	15837	3	701.71
receive	Female	Male	5.86	4.96	24111	2	688.71
research	Female	Male	6.30	5.72	27663	2	714.70
everyone	Male	Female	6.71	6.15	13337	3	695.21
follow	Male	Female	5.60	5.46	46145	2	558.77
music	Male	Female	6.98	6.88	15024	2	552.27
prepare	Male	Female	5.20	4.86	14961	2	636.73
program	Male	Female	6.48	6.14	32068	2	753.68
recognize	Male	Female	5.57	4.35	15203	3	892.21
recover	Male	Female	5.99	5.10	4932	3	662.72
station	Male	Female	6.31	6.75	12328	2	721.20
weather	Male	Female	6.54	5.92	42042	2	493.79
Average			5.96	5.62	21225.33	2.33	676.63

Appendix J

Words for Recognition Task (VG1, VG2, VG3, VG4)

Vocabulary Group	Word	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
VG1	again	6.10	6.42	59829	2	596.72
VG1	contrast	5.49	5.11	8172	2	857.78
VG1	design	5.27	5.75	26375	2	708.60
VG1	open	6.13	6.77	46095	2	566.88
VG1	project	6.34	6.02	21648	2	730.98
VG1	recent	5.79	5.25	15812	2	648.93
VG1	addition	5.97	4.70	10664	3	693.69
VG1	character	5.34	6.21	12511	3	677.27
VG1	energy	5.94	5.42	13083	3	671.31
VG1	remember	6.79	6.07	26748	3	663.85
VG2	about	6.41	6.47	197115	2	694.05
VG2	cover	6.11	5.75	24698	2	513.59
VG2	copy	6.59	5.79	10562	2	589.94
VG2	depend	5.65	5.34	10125	2	652.40
VG2	language	6.29	6.35	22117	2	909.20
VG2	table	6.22	6.40	23092	2	583.00
VG2	arrangement	5.25	4.24	9054	3	867.56
VG2	difficult	6.21	6.63	22033	3	714.87
VG2	opinion	5.21	5.58	9213	3	694.05
VG2	position	5.48	6.27	28071	3	659.34
VG3	control	6.13	6.49	38281	2	673.37
VG3	ever	6.55	6.16	27195	2	512.04
VG3	image	6.53	6.54	11024	2	603.23
VG3	person	5.68	6.02	28981	2	645.31
VG3	publish	5.07	5.62	12242	2	659.34
VG3	police	5.97	6.27	27508	2	708.44
VG3	history	6.11	6.13	20064	3	596.21
VG3	official	5.49	6.22	15931	3	638.30
VG3	realize	6.05	5.50	15575	3	813.66
VG3	separate	5.66	4.64	12159	3	827.69
VG4	agree	6.31	5.86	23497	2	760.18
VG4	degree	5.22	4.43	12996	2	604.25
VG4	message	6.47	6.58	8938	2	773.18
VG4	number	6.84	6.51	60607	2	565.26
VG4	only	6.10	6.62	152903	2	558.77
VG4	promise	6.06	6.24	10432	2	747.19
VG4	advantage	5.76	5.40	10285	3	877.13
VG4	another	5.38	6.40	60182	3	643.23
VG4	interview	5.38	6.30	9008	3	708.20
VG4	occasion	5.75	4.22	9152	3	734.19
Average		5.93	5.87	29099.43	2.40	683.58

Appendix K

Words for Recognition Task (VG5)

Vocabulary Group	New Word	Familiarity (Spoken)	Familiarity (Written)	Frequency (BNC)	Syllable Number	Duration
VG5	adult	5.19	6.09	8402	2	623.74
VG5	against	5.92	5.68	56208	2	707.79
VG5	among	5.48	5.10	22864	2	579.50
VG5	attempt	5.03	3.79	21750	2	667.98
VG5	because	6.43	6.40	103003	2	787.42
VG5	before	6.17	6.56	88275	2	616.09
VG5	between	5.49	5.89	91141	2	641.43
VG5	college	6.76	6.29	8375	2	703.37
VG5	enjoy	6.69	6.41	14527	2	610.47
VG5	enough	5.73	5.55	32593	2	539.69
VG5	every	5.40	6.42	40114	2	428.10
VG5	extra	5.89	5.38	8885	2	473.33
VG5	further	6.45	4.48	20138	2	522.00
VG5	kitchen	5.64	6.46	8866	2	486.61
VG5	network	5.45	6.16	8853	2	583.93
VG5	offer	5.75	5.58	36365	2	544.11
VG5	option	6.37	5.51	9189	2	561.81
VG5	partner	5.99	5.65	8605	2	557.38
VG5	pressure	5.38	4.93	14635	2	579.50
VG5	pupil	5.91	2.75	10320	2	473.33
VG5	quickly	5.62	6.00	12381	2	517.57
VG5	second	6.45	6.31	9445	2	650.28
VG5	series	5.54	4.89	14348	2	791.84
VG5	shoulder	5.68	5.41	8800	2	650.28
VG5	someone	5.83	6.15	18681	2	685.67
VG5	something	6.27	6.11	52452	2	667.98
VG5	speaker	6.36	6.24	9456	2	707.79
VG5	until	6.14	5.63	41484	2	606.05
VG5	very	6.16	6.75	123080	2	486.61
VG5	afternoon	5.88	6.02	8934	3	791.84
VG5	beautiful	6.35	6.78	8670	3	681.25
VG5	finally	5.80	5.82	13014	3	583.93
VG5	library	6.34	6.28	10356	3	667.98
VG5	otherwise	5.83	4.34	8798	3	858.20
VG5	together	6.34	6.23	30960	3	619.32
VG5	difficulty	5.46	5.69	13177	4	729.91
VG5	January	6.23	6.18	10200	4	721.06
VG5	particular	5.66	4.74	29718	4	707.79
VG5	professional	6.54	6.31	13496	4	716.64
VG5	situation	6.47	5.66	19856	4	864.66
Average		5.95	5.72	26760.35	2.40	634.85