Effects of Accelerated Speech Dictation and Shadowing on Speech Perception Ability

A Dissertation Submitted to the University of Tsukuba in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Linguistics

Toshihide O'KI

2022

Abstract

Effects of Accelerated Speech Dictation and Shadowing on Speech Perception Ability

by

Toshihide O'KI

Speech perception, defined as the process by which listeners decode spoken input phonologically and identify words, plays a crucial role in listening comprehension. Several researchers have noted that listening breakdowns in a second or foreign language are often caused by word segmentation failure derived mainly from characteristics of spoken English (e.g., phonological modification such as assimilation and elision). It is well-established in the literature that dictation and shadowing help learners develop their perception abilities, yet there has been little empirical research that attests to this assumption. In contrast, numerous studies have attempted to reveal the effectiveness of these exercises on listening comprehension ability, revealing miscellaneous results. One of the techniques to enhance language learning is to increase the cognitive load of the training. With reference to listening, this can be accomplished by accelerating the speed of aural materials. Dictation is assumed to take advantage of this kind of manipulation because, unlike shadowing, dictation is an offline task where learners perform listening and writing separately. Based on these backgrounds, the present research is aimed at revealing the effectiveness of dictation using accelerated (i.e., fast-paced) speech and shadowing for developing the speech perception abilities of Japanese English learners. The following five studies were conducted as a part of this research.

Study 1 attempted to test the hypothesis that the threshold at which comprehension becomes challenging for English language learners is at a speech rate of around 200 words per minute (wpm). Participants listened to English passages at three speech rates (i.e., 135, 175, and 215 wpm) and transcribed as many words as they could. Analyses revealed that the reproduction rates gradually decreased as the speech rate increased, thus indicating that the hypothesis was not supported. However, it also emerged that the performance of upper-level learners at the fastest rate was poorer than that of lower-level learners at the slowest rate. This suggests that English speech faster than 200 wpm is challenging for most learners, indicating that dictation using materials at this speed may be effective for developing perception ability.

Study 2, conducted as a pilot study, examined whether a one-time training session would improve learners' perception abilities. In a pre-post study design, participants completed either accelerated speech dictation or shadowing for 15 minutes, and their improvement was measured using a written reproduction task. The analysis revealed that both groups improved significantly and performed comparably in the pre- and post-tests, implying that both exercises were equally efficacious for speech perception ability. However, the improvement was more likely just a practice effect brought on by using the same material throughout the study. Another limitation was that participants' writing abilities had an impact on how well they performed on written reproduction tasks, which are skill integration tasks. Hence, a non-integrative measure needed to be developed for further study.

Studies 3 and 4 investigated the possibility of substituting an original perception task named a *word count task* for the written reproduction task. This task, specializing in assessing the word segmentation skill, does not require participants to reproduce the input linguistically but to count the number of words masked in the text. Study 3 was a preliminary study aimed at evaluating the validity and reliability of this task using short sentences, whereas Study 4 was a

follow-up investigation of its revised version using passages. Analyses of Study 4 revealed that the word count task displayed a high-reliability coefficient, slightly exceeding that of the written reproduction task. Regarding validity, the external aspect measured through a correlation analysis with the written reproduction task revealed a substantial association between the two tasks, implying that the word count task can replace the written reproduction task. However, listening strategy surveys indicated that the cognitive processes involved in both tasks were somewhat different. Therefore, speech perception ability should be measured using both tasks.

Study 5, the main investigation of this research, sought to examine the effectiveness of the two exercises when executed for a longer period than Study 2. For two months, the participants engaged in any exercise they chose and maintained a weekly journal. Improvement of their perception ability was measured through the two tasks—the written reproduction task and the word count task—in a pre-post-study design. After the post-test, the participants also responded to the same listening strategy survey as in Study 4. An overall conclusion drawn from a series of analyses indicated that the two exercises were effective for fostering speech perception abilities; however, some different effects can be anticipated between them. On the one hand, accelerated speech dictation, which directs learners' attention toward meaning and word forms, may improve their general listening abilities and expand the lexical knowledge necessary for better word recognition. On the other hand, shadowing, in which learners displayed an inclination to repeat it in their minds, could enhance perception efficiency while developing their working memory function.

The current research has provided new insights into the effectiveness of dictation and shadowing; however, future research needs to investigate four aspects. First, the effectiveness of the short-term exercises remains open to debate. Second, the final study could not examine the role of learners' proficiency because of the small sample size of each group; thus, this needs

to be considered in future research. Third, the measurement tasks used in this research were intended to measure accuracy in perception; however, efficiency is also an important aspect. Further insight into the topic may be obtained by conducting reaction time research. Finally, it is also necessary to investigate whether improvement in perception abilities will truly lead to better comprehension abilities.

Acknowledgments

First and foremost, I would like to express my deepest and most sincere gratitude to Professor Akiyo Hirai, my advisor at the University of Tsukuba, for her continuous and heartfelt support of my M.A. and Ph.D. studies. Her enthusiasm for research has always motivated me to conduct better research, and without her wise counsel, I could not have completed this research.

My sincere thanks also go to Professor Hirosada Iwasaki and Professor Yuichi Ono at the University of Tsukuba for their valuable advice on my research. I would also like to express my appreciation to Professor Yasuyuki Sakuma, a professor at Fukushima University, who served as the external member of the dissertation committee and was my teacher while I was enrolled at Fukushima University, for his constructive suggestions on this dissertation.

My deep gratitude is also extended to my colleagues in the English education course at Hakuoh University, who encouraged and supported my work. I am incredibly grateful to Professor Kyoko Miyazato, our caring leader, for her warm and heartening encouragement. I am deeply indebted to my colleague, Professor Yoshinobu Mori, a graduate of the same Ph.D. program, who generously provided valuable information to advance my research.

I would also like to send special thanks to Professor Takahiro Yamanoi at Bunkyo University, my ex-coworker at the Hakuoh Corporation since 2005, as well as my lifetime friend, for engaging me in casual conversations about our work, personal lives, and families.

Words cannot express my gratitude to my parents and parents-in-law for their continued and unconditional support. Their unwavering support and assistance to my family and me is a huge relief.

Last but not least, I would like to extend my deepest gratitude to my caring and

supportive wife Aya, and our lovely daughter Mitsuha. Your smiles and encouragement gave me great comfort and strong energy to complete this research, even during the most challenging times. My heartfelt thanks.

Toshihide O'KI

Table of Contents

Chapter 9 General Discussion and Conclusion

List of Tables

List of Figures

List of Appendices

Chapter 1

Introduction

1.1 Research Background

Listening ability in a first language (L1) develops naturally and without intentional effort if a person grows up in an ordinary language environment. This is owing to the abundance of language input that we receive from people and the world around us after birth. Even after infancy, listening remains the most important language skill because, as Rivers and Temperley (1978) state, nearly half of our language activity in L1 is devoted to listening. As L1 speakers receive this tremendous amount of input, they gradually acquire linguistic knowledge and become familiar with the phonological system of their L1. This leads to the automation of the comprehension process, making it possible for them to understand very rapid speech, as revealed by Beatty, Behnke, and Goodyear (1979) and Wingfield and Nolan (1980).

In contrast to L1, listening in a second language (L2) is not as easy for learners. Even if learners of English as a second language (ESL) spend 50% of their language activity on listening (Nunan, 1997), they often can catch some words but interpret the passage wrongly. According to Buck (2001), the listening performance of L2 learners can be affected by various factors derived from the characteristics of spoken language (e.g., phonological modification, speech rate, and non-verbal signals) and a lack of knowledge of the linguistic and socio-cultural aspects of L2. Therefore, L2 listeners face many more obstacles than L1 listeners, as Buck explains metaphorically:

> If we think of language as a window through which we look at what the speaker is saying, in the case of first-language listening, the glass is very clean and we see through it without even noticing it is there; but in the case of

> second-language listening, the glass is dirty: we can see clearly through some parts, other parts are smudged, and yet other parts are so dirty we cannot see through them at all. We are very aware of the glass because it gets in the way. (p. 50)

For learners to enhance the transparency of their "glass," they must inevitably develop their ability in *speech perception* (or just perception). According to Richards and Schmidt (2010), perception in listening means "to detect different kinds of acoustic signals" (p. 427). In well-known comprehension models (e.g., Anderson, 2005; Vandergrift & Goh, 2012), perception plays the most fundamental role in language comprehension. Hence, learners with limited perception ability fail to perform higher-level processing such as recognition of words, parsing of grammatical structures, and understanding of sentence and passage meaning. In a word, their window is very opaque. For this reason, L2 listening instruction should primarily focus on improving learners' speech perception ability.

Nevertheless, attaining this goal can be challenging for instructors who teach English as a foreign language (EFL). To improve speech perception ability, repeated exposure to the target language is essential. However, in comparison to an ESL environment, where learners can easily access English outside classrooms, an EFL environment such as Japan greatly lacks such an opportunity. Furthermore, since English is not used in daily conversation in EFL countries, it is difficult to enhance learners' motivation to study listening. In fact, Hirai, Fujita, and O'ki (2013) report that even the introduction of the Center Listening Test, a national high-stakes exam in Japan, did not provide a strong incentive for high school students to study listening harder. The national guidelines for English education in Japan (in the Courses of Study) stipulate that listening instruction be primarily aimed at fostering learners' ability to grasp the main point of an English passage; however, as stated above, accurate comprehension

is hardly achieved without good perception ability. Therefore, the role of EFL teachers is to make the best of their limited lesson time to develop learners' speech perception ability by employing effective activities.

Such activities recommended by the literature are *dictation* (e.g., Field, 2003; Morris, 1983) and *shadowing* (e.g., Kadota, 2007, 2012; Tamai, 2005). Dictation has a very long pedagogical history as a teaching technique (e.g., Tamai, 2005; Yanagihara, 1995) as well as a testing method (e.g., Henning, Gary, & Gary, 1983; Templeton, 1977). Regarding the former, there has been extensive research on its effect on listening ability (e.g., Mohammed, 2015; Suenobu, Young, Kanzaki, & Yamane, 1982; Takeuchi, 1997). As explained in the next chapter, many studies have revealed its positive effects; however, several studies have suggested that diction is more effective for other language skills than listening. In comparison, shadowing, which was originally introduced as a basic training method for simultaneous interpreters (Yashima, 1988), has a shorter pedagogical history (e.g., Tamai, 1992). Research on its effect on listening ability flourished in the 2000s in Japan, especially after Kadota (2007, 2012) published books focused on the scientific aspects of shadowing. It then started to gain worldwide recognition, probably because of an English book written by Hamada (2017). In their books, shadowing is advocated as a means to improve learners' speech perception ability.

The rationale for this effect of shadowing on speech perception ability is that shadowing, where a learner listens and speaks simultaneously, imposes an additional cognitive load on learners; thus, they can only focus their attention on sounds rather than on meanings (Kadota, 2007). On the other hand, dictation, which is an offline activity where a learner usually transcribes words *after* hearing them, may be less demanding. Speech rate is one of the factors that cause cognitive difficulty in listening (Buck, 2001). Therefore, it is speculated that using fast speech for dictation may exert a similar effect on a learner as shadowing. Despite these expectations, little or no empirical research has been done to test these hypotheses.

1.2 Organization of the Dissertation

Based on the background stated above, the current research investigated the effects of dictation and shadowing on the speech perception ability of Japanese EFL learners. To amplify the effect of dictation, *accelerated speech* (i.e., speech whose rate is manipulated to be faster) was used for training in this study. This dissertation consists of nine chapters covering five studies (see Table 1 at the end of this chapter for an overview of the research).

Chapter 2 provides an overview of the literature related to the research objective. First, the definition of perception is discussed based on several listening comprehension models. Second, the reason why perception ability is crucial for ESL/EFL listening is explained with reference to the characteristics of spoken English and their influence on speech perception. Third, comprehensive overviews of empirical studies of the effectiveness of dictation and shadowing and the influence of speech rate on listening are given, followed by a discussion of how dictation using accelerated speech can improve perception ability. The chapter finally addresses the general objective of this research (i.e., whether dictation using accelerated speech and shadowing will develop speech perception ability).

Chapter 3 reports Study 1, which explored the influence of accelerated speech on learners' perception. This was a preliminary study to investigate whether accelerated speech was cognitively more demanding than the original speech. Japanese university students listened to English speech at three different speech rates (approximately 135, 175, and 215 words per minute; wpm) and were asked to transcribe as many words as possible. Their reproduction rates were compared between the three speech rates to examine whether perception became more difficult as the speech rate increased. Moreover, error analyses were conducted to explore what kind of difficulty learners would have in perception.

Chapter 4 describes Study 2, a pilot study that investigated the effect of short-term training on the improvement of speech perception ability. Japanese university students were

randomly assigned to either the dictation group or the shadowing group and engaged in a one-shot 15-minute training. The dictation group practiced with English speech accelerated to around 200 wpm, while the shadowing group shadowed the same speech at slower rates (150/110 wpm). A pre-post study design using a written reproduction task was employed to measure perception improvement. The students also responded to a questionnaire to evaluate the effectiveness of their training.

Chapters 5 and 6 report two studies (Studies 3 and 4) that were intended to develop a new task to measure speech perception ability. The written reproduction tasks used in the previous two studies were integrated-skills tasks of listening and writing; therefore, participants' task performances were affected by their spelling knowledge. To avoid this influence, the author developed an original non-integrative test task named the *word count task*, where learners listen to English sentences or passages with blanks and answer how many words were spoken in the blanks. Study 3 was conducted to create a prototype of such a test, and its validity and reliability were compared with a written reproduction task. Based on the limitations identified in Study 3, Study 4 attempted to revise the word count task so that learners' speech perception ability could be assessed more accurately.

Chapters 7 and 8 report on the final study (Study 5), which was conducted to investigate the training effect over a longer term than Study 2 on the improvement of perception ability. Japanese university students engaged in either accelerated speech dictation or shadowing for two months and kept a weekly journal on their learning. Improvement of their perception ability was compared with that of a control group using the tasks developed in Study 4. On the last day, they responded to a listening strategy survey and gave feedback on the whole training. Chapter 7 discusses the training effect based on quantitative analyses of their performance in the perception tests and of their responses to the strategy items as in Study 5-1, while Chapter 8 provides further insight by examining the results of quantitative analyses using

text mining in Study 5-2.

Chapter 9 summarizes the findings of the five studies and draws a general conclusion on the effect of speech perception training with accelerated speech dictation and shadowing. The chapter also adduces the educational implications of incorporating these activities into listening instruction and, finally, discusses the limitations of this research.

Table 1

Overview of the Five Studies in the Present Research

Study	Goals and Research Questions (RQs)
1	**Goal: To examine the influence of accelerated speech on perception**
	RQ1-1: Does the reproduction rate of English speech decrease drastically when the speech rate exceeds 200 wpm? Is the influence of speech rate on perception more critical to lower-level learners? RQ1-2: What kinds of words are difficult to perceive when the speech rate exceeds 200 wpm?
2	**Goal: To examine the effectiveness of short-term training**
	RQ2-1: Will the training using accelerated speech dictation and shadowing improve learners' speech perception ability? RQ2-2: Will learners appreciate the effectiveness of accelerated speech dictation and shadowing?
3 4	**Goal: To develop a new task to measure speech perception ability**
	RQ3: Is a word count task valid and reliable as a measure of speech perception ability? RQ4: Is a revised version of the word count task valid and reliable as a measure of speech perception ability?
5	**Goal: To examine the effectiveness of long-term training**
	RQ5-1: Will long-term training with accelerated speech dictation and shadowing improve learners' speech perception ability? RQ5-2: Will the training develop learners' strategy use in listening? RQ6-1: What kinds of effects did the participants feel throughout the training? RQ6-2: What kinds of cognitive processes were involved in accelerated speech dictation and shadowing?

Chapter 2

Literature Review

2.1 Models of L2 Listening: What is Speech Perception?

2.1.1 Problem of Defining Speech Perception

Although the term speech perception is commonly used in the literature, there is a variation in its meaning. *Longman Dictionary of Language Teaching & Applied Linguistics* (Richards & Schmidt, 2010) defines it as "the understanding or comprehension of speech" (p. 427). In this simple definition, speech perception seems to be regarded as interpretation of a spoken message; however, it does not match the traditional view on speech perception held in well-known theories of speech perception. For example, *categorical perception* and *motor theory of speech perception* attempted to explain the mechanism of how acoustic properties of phonemes are processed in listeners' mind (see Ryals, 1996 for an overview of these theories). Similarly, the *Perceptual Assimilation Model* (Best, 1994; Best & Tyler, 2007) claims that L2 listeners tend to classify (or assimilate) unfamiliar L2 phonemes into sounds in the phonological system of their L1. The view of speech perception adopted in these theories is the narrowest one since they are concerned with perception of only phonemes.

Rost (2016), taking a similar view, defines perception as "the initial neurological response to any source of sensory stimulation, such as sound waves; auditory sensations are considered to reach perception only if they are received and processed by a cortical area in the brain" (p. 335). His definition looks more progressive than the traditional view since it stresses that perception is a neurological process controlled by brain. Moreover, its reference to the acoustic information of speech input (i.e., sound waves) makes us recall the aforementioned speech perception theories. Rost and Wilson (2013) define "perceptual processes" more simply as "meaning-oriented responses to sensory stimulation, e.g., sound waves" (p. 307). This

definition can be distinguished from the previous ones in that the expression "meaning-oriented" seems to indicate the role of top-down processing based on meaning of words or sentences. Therefore, it is not certain whether Rost considers that speech perception operates only at the phoneme level.

In contrast to these views, there are researchers who hold a wider perspective. For example, Ur (1984) introduces several activities for developing perception ability both at word-level and sentence-level. Most activities for the word-level are aimed at enhancing learners' sensitivity toward different sounds; thus, they are phoneme-based[1].On the other hand, activities for the sentence-level are focused upon developing the abilities to detect words in a sentence and analyze its prosodic features such as intonation and assimilation of consonants. For example, an activity named "Identifying Word-Divisions" requires learners to listen to a colloquial sentence like "wotcha won? (What do you want?)" and to analyze how many words were in the utterance (i.e., four). This type of exercise may request learners to have knowledge about formulaic expressions; thus, it is apparently beyond the phoneme-level.

Another example is Tatham and Morton (2006), who give a somewhat abstract explanation about perception. Below is the citation of their view, in which they claim that perception not only involves processing of acoustic information but also entails construction of meaning representation. As they state:

> For us perception is the assigning of some meaningful symbolic representation to input sensory data. The assignment is based on a characterization of possible symbolic representations held in the listener's

[1] An example activity is "How Often Did You Hear It?," in which the teacher reads aloud phrases or short sentences for learners (e.g., "a bit of cheap ribbon") and ask learners how many times the target sound (/ɪ/) appeared in the input.

> mind. Thus perception is essentially an act of interpretation since it is clear that there is no linear correlation or 'direct path' between the acoustic signal and the assigned symbolic representation. … What they (perceivers) do is interpret what is heard in a complex process of assignment, from what they already know, of symbolic representations. (p. 20)

Here, they express the limited role of acoustic information in understanding spoken language. This is probably derived from the notion that there is often an inconsistency between acoustic signal and our perception of the sound (e.g., syllables with strong intensity in the soundwave are not necessarily recognized to be prominent by listeners). This issue is called "the lack of acoustic invariance" (Ryalls, 1996, p. 41). Therefore, it is not reasonable to put too much emphasis on acoustic information when discussing the role of speech perception in listening comprehension.

The discussions made above have revealed that speech perception is an ambiguous notion and seems not to have a uniform definition. The goal of this research is to investigate the effectiveness of dictation using accelerated speech and shadowing in improving learners' ability to figure out words for listening comprehension. To conduct studies on this topic, it is necessary to clarify the role of speech perception in listening comprehension. Hence, the next section will refer to several well-known comprehension models to define speech perception for this research.

2.1.2 Listening Comprehension Models

According to Field (2008), instruction for listening had long been overlooked in the L2 classroom, and it was not until the late 1960s when listening started to receive attention from those involved in L2 teaching. Before that era, when the Audiolingual Method prevailed,

listening played a limited role only as an input to present new grammar. However, during the 1970s and 1980s, such as generative linguistics and cognitive psychology emerged, and new teaching trends derived from these schools flourished. One among them was the Natural Approach proposed by Krashen and Terrell, which emphasized the exclusive role of comprehensive input for L2 acquisition and led L2 researchers and teachers to recognize the importance of listening instruction (Sano, 1995). It was during these eras that significant models and theories related to listening were developed. The subsequent sections will introduce some of them to see what kind of role speech perception plays in listening comprehension.

The Two-Stage View. According to Buck (2001), a classical view on listening comprehension is that the listening comprehension process is composed of two stages. At the first stage, listeners decode the linguistic input into meaning while, at the second stage, listeners interpret the meaning for some kind of communicative purposes. Buck introduces several researchers who advocate this view. For example, Rivers (1966) stated that listening comprehension involves two levels of processing; *recognition* (the level at which listeners recognize the linguistic elements) and *selection* (the level at which listeners extract important information to understand the gist). Similarly, Carroll (1972) considered that comprehension would take place as the results of understanding linguistic information and interpreting it by referring to the communicative context. In these views, what is considered to occur at the first stage is similar to Ur's (1984) view on speech perception mentioned previously since all the views assume that perception can operate at the sentence level. However, those researchers do not use the term "perception" to explain their views, thus it is not clear yet what it means.

Anderson's Three-Stage Comprehension Model. Anderson's (2005) comprehension model, which was introduced in the first edition of his book in 1980, is well-known and

probably the first model that was equipped with the perception stage. His model, especially designed for L1 comprehension, consisted of the following three stages. The first stage is called perceptual processing, or *perception*, meaning the process in which "the spoken (acoustic) or written message is originally encoded" (p. 388). This definition is too concise to understand what perception refers to. Especially, Anderson does not make clear whether perception means only identification of sounds or includes that of words as well. There is an impression that Anderson uses the terms speech perception and speech recognition almost interchangeably, thus the boundary between them is unclear.

The second stage is *parsing*, referred to as "the process by which the words in the message are transformed into a mental representation of the combined meaning of the words" (p. 388). In short, parsing means to make sense of a message based primarily on incoming linguistic information. This process is similar to comprehension in that listeners make some kind of interpretation about messages of the input, but different because parsing is completed without reference to the outside-text context.

The third stage is *utilization*[2], defined as the process that "comprehenders use the mental representation of the sentence's meaning" (p. 388) or that "language comprehenders respond to the meaning of a linguistic message" (p. 465) to "go from the literal meaning of a sentence to something that will be useful" (p. 406). For example, listeners not only answer questions they are asked but sometimes guess the character of their interlocuters. During this stage, listeners make inferences and attempt to reveal the relationship between propositions, through which they construct their own meaning representation.

Anderson adds that these three processes basically operate in a serial order from perception to utilization but also work simultaneously where necessary because message

[2] According to Field (2013), the term *utilization* is originally from Clark and Clark (1977).

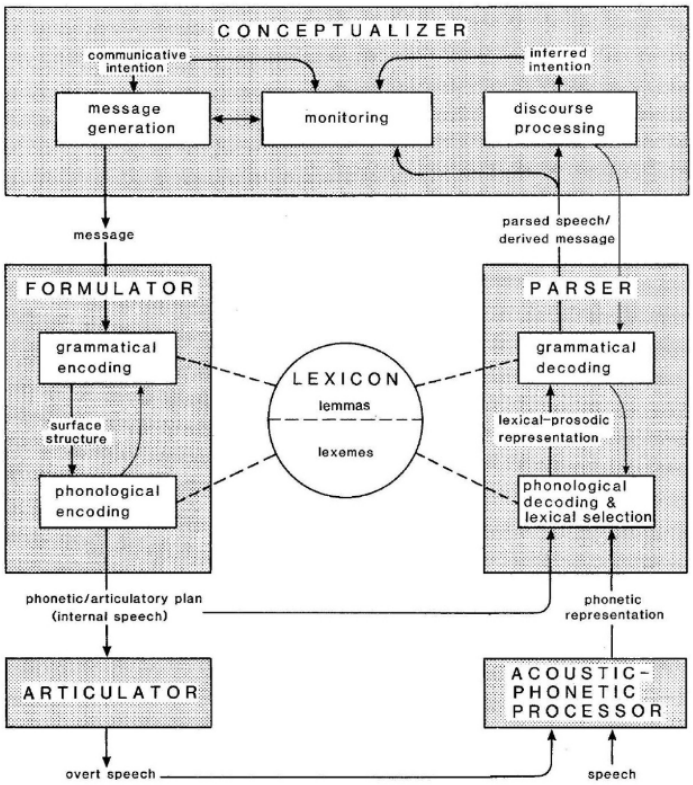
interpretation occurs immediately after the language input is received (the principle of *immediacy of interpretation*). For example, since syntactic structures are often ambiguous (e.g., garden-path sentences), semantic information passed down from the utilization stage is also used to parse the speech. Moreover, perception of words can be promoted by vocabulary knowledge or contextual information. However, Anderson emphasizes that the perception stage can cause us a greater problem when comprehending spoken language than written language. This is due to unstable features of phonemes and to listener responsibility in detecting word boundaries that are not overtly presented in connected speech. With regard to this point, L2 listeners are far more handicapped than L1 listeners due to their limited linguistic knowledge.

The Levelt Model of Speech Production and Comprehension. Levelt (1993) also proposed a model of spoken language use in L1 (Figure 2.1). This is the refined version of his earlier model (Levelt, 1989; see Appendix 2A), with which he intended to clarify the processes involved in oral conversation. While the original model did not fully specify what processes are involved in listening comprehension, the elaborated version covers not only speaking but listening processes by illustrating processing components that are similar to Anderson's three-stage comprehension model.

First, *acoustic-phonetic processor* analyzes acoustic signals of incoming speech and allows listeners to have *phonetic representations*. To this end, listeners examine such phonological properties as word onsets, spectral peaks, frequencies, and formants, figuring out which consonants and vowels they have heard. Since the model indicates no arrow coming down from the higher-processing to this processor, this process is accomplished based purely on the acoustic signals. Therefore, this process takes a very similar view to the traditional perspective on speech perception.

Figure 2.1

Schematic Representation of the Processing Components Involved in Spoken Language Use (Levelt, 1993, p. 2)



Second, *parser*, by drawing on the phonetic representations, derives literal messages from the input through three decoding processes. During the *phonological decoding* and *lexical selection* phase, word candidates that have been chosen from listeners' lexicon are narrowed down by using higher information such as context and word frequency to be recognized. The subsequent *prosodic decoding*, which is not overtly indicated in the chart but explained in his article, allows listeners to assign intonational information to each phrase (or construct *lexical-prosodic representations*) so that the listener can elicit communicative intention of their interlocutor. Finally, the role of the *grammatical decoding* is to construct meaning representations based on the textual information. To accomplish this, listeners exploit not only syntactic but also semantic cues as indicated by the downward arrow from discourse processing.

One important characteristic of Levelt's model is that it specifies lexical selection, or word recognition, to belong to the parsing stage, while it was not clear in Anderson's model.

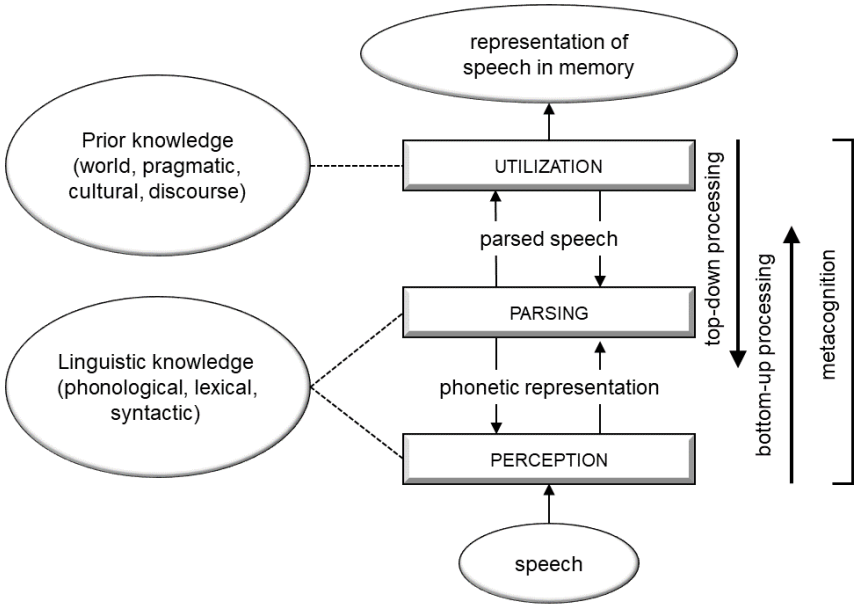
Third, *conceptualizer*, associated with both production and comprehension, plays a similar role to the utilization process. For listening, it involves two kinds of processings; *discourse processing* and *monitoring*. The former refers to understanding interlocutor's communicative intention. A representative phenomenon of this processing is inferencing indirect referents in such an utterance by a restaurant server as "*The hamburger wants the bill*," implying "*The guest who ordered a hamburger wants the bill.*" Appropriateness of these inferred messages are then examined by the monitoring processing, which in turn gives assistance to message generation process in speaking. The role of conceptualizer is identical with utilization in that listeners relate the input to the outside-text context.

Vandergrift and Goh's Comprehension Model for L2 Listening. Theoretical development achieved throughout the 20th century has led to emergence of more sophisticated models for L2 listening. The comprehension model of Vandergrift and Goh (2012), primarily based on Anderson's three-stage model, defines perception as the process that "listeners use bottom-up processing to recognize sound categories (phonemes) of the language, pauses, and acoustic emphases, and hold these in memory" (p. 21). In other words, the role of perception is to analyze phonological features of the input. They regard parsing as the process that "listeners parse the phonetic representation of what was retained in memory and begin to activate potential word candidates" (p. 22). This definition shows that their model assumes word recognition to be part of parsing rather than that of perception.

Vandergrift and Goh's model is more expedient than Anderson's in two respects. First, it clarifies the kinds of knowledge sources used in each stage. As shown in Figure 2.2, the lower-level processings (i.e., perception and parsing) are driven by *linguistic knowledge*, which

Figure 2.2

Cognitive Processes and Knowledge Sources in Listening Comprehension (Based on Vandergrift & Goh, 2012, p. 27)



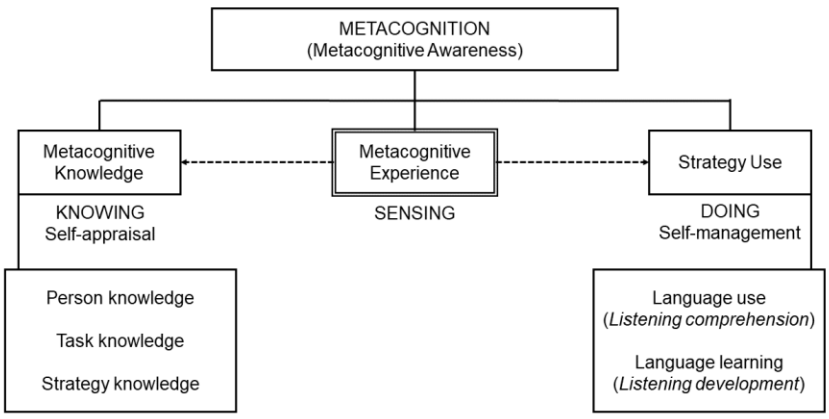
comprises phonological, lexical, and syntactic knowledge, while the upper-level processing (i.e., utilization) is activated through *prior knowledge*, which consists of world, pragmatic, cultural, and discourse (or script) knowledge. Compared to L1 listeners, L2 listeners are handicapped with the linguistic knowledge and often fail to make a sense of spoken input successfully by exploiting the linguistic information, or by bottom-up processing. In such a case, compensatory processing derived from the prior knowledge, or top-down processing, backs up the lower-level processings by providing them with a conceptual framework regarding the topic. However, since learners' prior knowledge is often bound by their cultural background, the process may hinder comprehension when they are preoccupied by their expectations about passage meaning and unable to reform their mental representation as listening proceeds. Vandergrift and Goh mention that listeners must become able to interweave bottom-up processing and top-down processing depending on their listening goal and ability.

Second, they added a new component *metacognition*, which means "listener awareness

of the cognitive processes involved in comprehension, and the capacity to oversee, regulate, and direct these processes" (Vandergrift & Goh, 2012, p. 23). Based on the framework of Flavell (1979), Vandergrift and Goh suggest the following three components (Figure 2.3). First, *metacognitive experience* refers to listeners' perception about past learning. This includes in what situations they succeeded and failed in listening as well as how they reacted to them. This experience enables learners to cope with new listening tasks and plays a central role in the metacognition system since it contributes to the development of the other two components. Second, *metacognitive knowledge*, obtained as a consequence of metacognitive experience and stored in long-term memory, is about learners themselves (person knowledge), task characteristics (task knowledge), and strategies to achieve a task goal (*strategy knowledge*). Third, *strategy use* means learners' ability to put their strategy knowledge into practice for listening comprehension (language use) as well as for development of their learning skill (listening development). Given that learners with more strategic knowledge are able to use more strategies, listening instruction should be aimed at providing learners with learning experience in order to increase their metacognitive knowledge.

Figure 2.3

A Metacognitive Framework for Listening Instruction (Based on Vandergrift & Goh, 2012, p. 85)

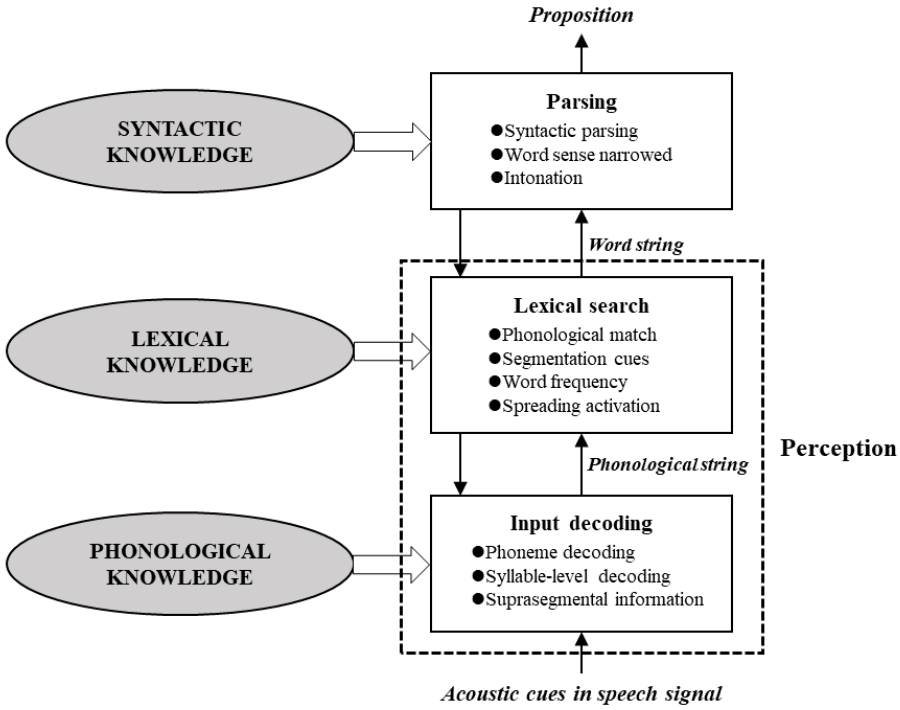


Field's Five-Stage Model. The three-stage comprehension model originally proposed by Anderson is well-known, but it does not specify what processings the perception stage comprises. To emphasize the importance of this stage, Field (2013) elaborated the model by dividing it into two processes, *input decoding* and *lexical search*. Furthermore, Field points out that the term utilization is misleading, thus he reconceptualized it by splitting it into two processes, *meaning construction* (i.e., application of world knowledge and inference) and *discourse construction* (i.e., integration of the text comprehension into the ongoing context). Therefore, his model consists of five stages. The model given in Figure 2.4 illustrates the relationship among the lower-level processes of his model.

The initial processing is *input decoding*, which refers to the process that "the listener transforms acoustic cues into groups of syllables, some marked for stress and others not" (p. 95). In other words, it means that listeners perform phonological analyses on spoken input at three levels: phoneme, syllable, and prosody. Phonological string is then constructed to be passed on to the subsequent processing called *lexical search*, during which "the listener identifies the best word-level matches for what has been heard, based on a combination of perceptual information and word boundary cues" (p. 95). In a nutshell, listeners draw upon their lexical knowledge and search for the word that corresponds to the phonological string; thus, this process is identical with word recognition. Just as well-known word recognition models suggest, words with high frequency are more likely to be activated thus recognized easily by listeners than those with low frequency (e.g., "heard" is recognized more easily than "herd"). An important processing performed at this level, but challenging for learners, is word segmentation. As noted later, this difficulty is primarily due to the features of spoken English such as phonological modification (e.g., assimilation and reduced forms). Field states that prosodic information such as word stress and syllable duration provides listeners with important cues for this process.

Figure 2.4

Model of Lower-Level Processes in L2 Listening (Based on Field, 2013, p. 97)



Note. In the original chart, input decoding is at the top while parsing is at the bottom; yet, the order is reversed upside down in this figure so that the perception stage will be at the lowest level in accordance with Anderson's (2005) three-stage model. The dotted square line is drawn by the author to indicate that the two processes correspond to the perception stage.

The Field model is valuable in that it has succeeded in describing the perception stage in detail using an orderly chart. Furthermore, the distinction between input decoding and lexical search seems to be convincing thus helps us understand what processes are involved during speech perception. However, it is preferable to view these two processings as one stage because they almost co-occur in actual listening and are extremely difficult to assess them separately in studies. For this reason, this dissertation holds the perspective that speech perception is one stage consisting of two sub-processes, input decoding and lexical search (word recognition), rather than sees it as two separate stages.

2.1.3 Definition of Speech Perception for the Current Research

Overview of the listening comprehension models suggests that each model has its own characteristic although they consist of similar processes in line with Anderson (2005). The model of Vandergrift and Goh (2012) was the most comprehensive because it covers some processings that do not appear in the other models such as the knowledge sources and metacognition. Nevertheless, the detail of each stage is not illustrated in the chart, so there is a room for refinement.

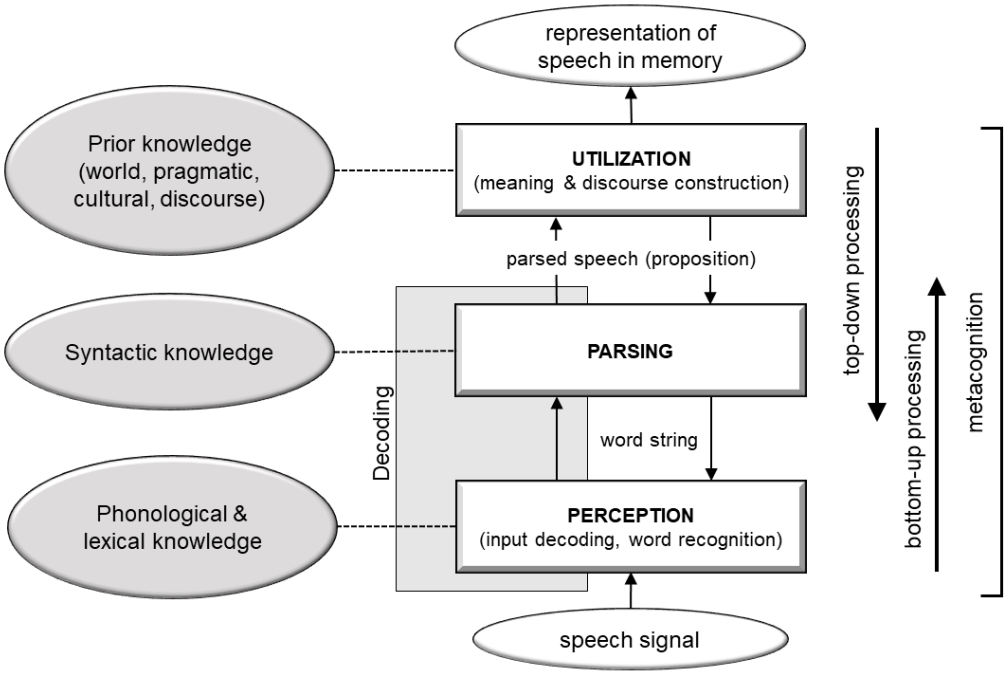
As for definition of speech perception, there are two kinds of perspectives with regard to whether it includes word recognition. While Levelt (1993) and Vandergrift and Goh (2012) incorporated it into part of the parsing stage, Field assumed it (lexical search) to constitute the perception stage with input decoding (i.e., analyzing of phonological features). Field's perspective seems more plausible because words are recognized as soon as their phonological features are identified, or even before that, thus word recognition has a stronger relationship with phonological processing than with grammatical processing.

Based on all the considerations given so far, the author has proposed an integrated model of L2 listening comprehension shown in Figure 2.5. In the model, Field's view on the perception and utilization stages is incorporated into Vandergrift and Goh's model by specifying the sub-processes constituting the two stages. Beside it, there are several changes from their original model. One of them is that lexical search has been replaced with word recognition since the latter is more commonly used in the literature. Another change is that the knowledge sources for parsing and perception are separated in accordance with Field's model. Furthermore, the term *phonetic representation* has been replaced with *word string*, which is borrowed from the Field model, because words are already identified as a result of perception. Finally, perception and parsing stages are labeled as *decoding*, which refers to the process that "listeners gradually build meaning from phonemes to words to increasingly larger units of meaning (full sentences

and larger chunks of discourse)" (Vandergrift & Goh, 2012, p. 18). This term needed to be defined because it is useful when referring to the bottom-up processing based on linguistic information. Some researchers use bottom-up processing for this meaning, but this is misleading in that it expresses only the direction of information processing (Field, 2008).

Figure 2.5

An Integrated Model of L2 Listening Comprehension Based on Vandergrift & Goh's (2012) and Field's (2013) Models



Note. Arrangement was made for directions of the arrows indicating the relationship between the three stages so that the rectangle for decoding can cover the bottom-up processing during the perception and parsing stages.

In conclusion, this dissertation regards perception as a combined process of input decoding (i.e., phonological analysis of the spoken input) and word recognition (i.e., identification of words using phonological information) driven by phonological and lexical knowledge.

2.2 Characteristics of L2 Listening

The previous section introduced some listening comprehension models to define speech perception for this research. Perception is the most fundamental process in listening, thus listeners, whether L1 or L2, must have a very good command of this ability. However, many ESL/EFL listeners find it very difficult and often fail to understand spoken messages properly. Especially, segmentation of connected speech into words is a great challenge for them. Perception failure is primarily due to the "ephemeral, one-shot nature" of listening (Lynch & Mendelsohn, 2010, p.180), but there are other reasons in the case for ESL/EFL listeners. This section will overview characteristics of L2 listening while focusing on what challenges learners are likely to face.

2.2.1 Perception for L2 Listeners

When we listen in L1, we usually pay attention to meaning but rarely to what words are uttered by the interlocuter. This is because our perception ability in L1 is automatized thus can be performed unconsciously. In contrast to them, not a few L2 listeners have great difficulty in this process for many reasons. For them to be advanced listeners, development of perception ability is crucial. This section will present why this ability is important for L2 listening and what perception difficulty learners are likely to be face with.

Bottom-Up vs Top-Down Models: 'Modified' Interactive Model of Listening. Listening comprehension is considered to be an interrelated process among the three stages. There are two kinds of ways that the information obtained through each stage interacts, *bottom-up* and *top-down*. According to Rost (2016), bottom-up processing means "a form of information processing that is guided by input in real time, and proceeds in subsequent stages" (p. 278) while top-down processing refers to "information processing guided by higher level

mental processes as we construct representations, drawing on our experiences and expectations" (p. 306). This definition applies to both reading and listening; however, especially for listening, bottom-up processing means "perceiving and parsing the speech stream at increasingly larger levels" (Nation & Newton, 2009, p. 40).

According to Flowerdew and Miller (2005), the classical view on language comprehension in 1940s and 1950s was that language meaning can be interpreted only by bottom-up processing. This is called the *bottom-up model*. After several decades, research on L1 reading yielded abundant evidence for the roles of contextual information and prior knowledge in understanding language messages, contributing to development of the *top-down model*. During 1980s and 1990s, there was a dispute over which model was more convincing (See Rubin, 1994 for a review of extensive research on this issue); however, research uncovered that good listeners exploit both bottom-up and top-down processings to make up the deficits that occur in any processing stage. This type of view on comprehension is called *the interactive model*. Flowerdew and Miller state that usefulness of this model is its flexibility; thus, it can be applied to any listeners no matter what learning styles or strengths/weaknesses they may have, or whether they listen to L1 or L2.

As for L2 listeners, it is often the case that they rely too much on top-down processing due to their poor decoding ability (Field, 2008; Flowerdew & Miller, 2005; Nation & Newton, 2009). This notion is derived from the work by Stanovich (1980), who proposed a theory called the *Interactive-Compensatory Model*. With this theory, Stanovich tried to explain how L2 readers compensate for their skill shortage in understanding texts. After reviewing extensive research, he hypothesized that learners with poor word recognition skill are likely to depend on contextual cues. As he wrote:

> The compensatory assumption states that a deficit in any knowledge source

> results in a heavier reliance on other knowledge sources, *regardless* of their level in the processing hierarchy. Thus, according to the interactive-compensatory model, the poor reader who has deficient word analysis skills might possibly show a *greater* reliance on contextual factors. In fact, several studies have shown this to be the case. (p. 63)

Stanovich concluded that a necessary condition to be a good reader is to acquire context-free word recognition. Grabe and Stoller (2011) refer to the interactive models based on this notion as *'modified' interactive models*, which "highlight the number of processes, particularly automatic processes, being carried out primarily in a bottom-up manner with little interference from other processing levels or knowledge resources" (p. 27). Given these notions, L2 listeners need to automatize their decoding ability so that they do not have to depend excessively on top-down processing based on contextual information and prior knowledge. Perception of speech input is performed at the lowest-level of decoding thus plays the most fundamental role in listening comprehension because the other processings will not be activated unless listeners have some linguistic information at hand.

Controlled vs Automatic Processing. The 'modified' interactive model of listening holds the belief that automatization of learners' perception ability is crucial for efficient listening. When learners are not very adept or handy at some skill, it can be said that their processing related to the skill is still controlled. According to Shiffrin and Scheneider (1977), *controlled processing* is "usually serial in nature with a limited comparison rate" and "strongly dependent on load" (p. 127). This is contrasted with *automatic processing*, meaning "relatively well learned in long-term memory," "parallel in nature," and "virtually unaffected by load" (p. 127). The distinction between these processings is similar to the one between *declarative* and

procedural knowledge. Ellis (2008) explains the shift of one skill in language learning from declarative knowledge to procedural knowledge as follows: "…language learning, like other kinds of skill, is characterized by a progression from an initial declarative knowledge stage involving controlled processing, to a final procedural stage where knowledge is automatic" (p. 480). He adds that 'practice' is necessary for skills to be automatized and that automatization of one skill (e.g., listening) may give a positive influence to another skill (e.g., speaking).

Given these, an important objective of listening instruction is to aid learners to enhance their perception ability until it gets beyond the controlled level so that they can focus on higher-level processing. To this end, teachers need to have their learners work on activities aimed at development of speech perception ability. The literature contends that dictation and shadowing are the best activities for this purpose. An overview of studies on the effects of these activities will be given in the later section.

Working Memory and its Role in L2 Listening. A rationale under the notion that automaticity in speech perception contributes to better comprehension is that it frees up listeners' *working memory*, allowing them to pay more attention to the passage meaning. In other words, listening is an on-going process out of listeners' control, so effortless perception (or decoding) is a prerequisite for the functioning of working memory and then for efficient and accurate listening comprehension. Several researchers refer to the association between speech perception and the working memory in listening comprehension. For example, Field (2008) states:

> If a listener is able to decode the input effortlessly, the result is to leave a great deal of working memory free for thinking about larger issues such as the overall meaning of the text. If (as with a novice L2 listener) decoding is uncertain and makes heavy demands upon attention, then it leaves no memory

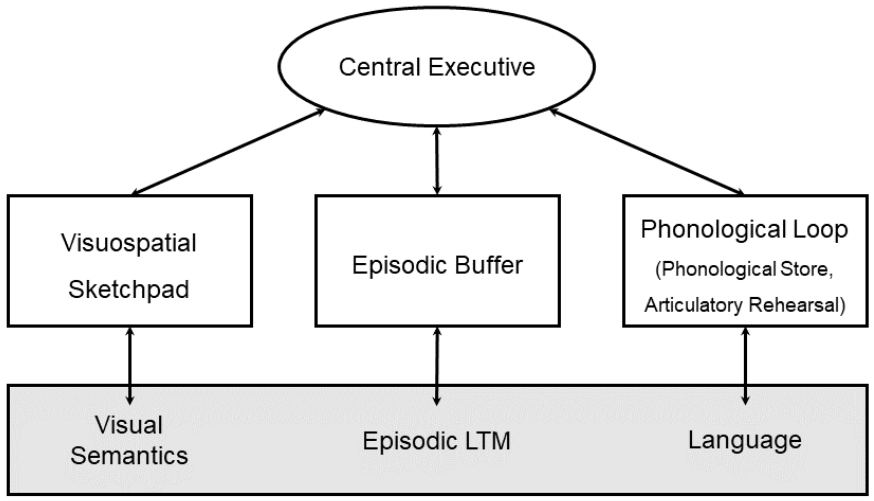
resources spare for interpreting what has been heard or carrying forward a recall of what was said earlier. (p. 136)

Likewise, Rost (2016) mentions that memory capacity is one of individual differences that have a critical impact on listening performance.

According to Baddeley (1992), working memory is "a brain system that provides temporary storage and manipulation of the information necessary for such complex cognitive tasks as language comprehension, learning, and reasoning" (p. 556). Its newer model (Baddeley, 2000) consists of four main components shown in Figure 2.6. The *central executive* controls the execution of three subordinate processings. The *visuospatial sketchpad* is a storage for visual information, while the *phonological loop* holds verbal and auditory information. The *episodic buffer* integrates visual and auditory information stored in the other two components relaying the information to the long-term memory (i.e., the *episodic LTM*).

Figure 2.6

The Multi-Component Working Memory Model (Based on Baddeley, 2000, p. 421)



The phonological loop, comprising two subsystems, plays a significant role in

language processing. One of the subsystems is the *phonological store*, where perceived speech can be held for further processing; however, this information fades away within a few seconds. To avoid this, the other subsystem called the *articulatory rehearsal process* contributes to maintenance of the phonological information by rehearsing it in mind. Since this inner process is usually done without vocalization, it is also called *subvocal rehearsal*. Written input must also undergo this transformation process into phonological information to be comprehended. Baddeley (2007, 2012) claims that the phonological loop is responsible for efficient L2 learning, introducing two studies that showed its importance in vocabulary acquisition (Papagno, Valentine, & Baddeley, 1991; Papagno & Vallar, 1992).

As Osaka (2002) states, working memory capacity varies from learner to learner, and it expands as their proficiency level increases. She also claims that development of the phonological loop is crucial for language learning especially at its initial stage. One way to do so for listening is to engage in shadowing practice. According to Kadota (2007, 2012), an important effect of shadowing is to promote the subvocal rehearsal process by externalizing it as overt speech. Kadota illustrates how this can occur, which will be explained in the later section about shadowing (section 2.3).

Perception Problems. Field (2008) also states that decoding difficulty in listening results from either *text problems* or *process problems*. The former problems occur when a learner does not have enough knowledge to understand linguistic items used in the spoken text, while the latter ones refer to the case that the learner does have the knowledge but is incapable of applying that knowledge to the decoding process. An example given by Field is the case that learners fail to understand such sentence as “I’ve lived in Italy for ten years” properly. This sentence is a present perfect form thus implies that the speaker lives in Italy at the time of utterance. There are two possible reasons for such failure. First, the learner did not know that

the grammar refers to the present status of the speaker (text problem). Second, the learner knew the grammar in written format but mistook the contracted form "I've lived" as "I lived" (process problem); thus, this error is derived from perception inability. Field states that, while text problems can usually be solved by teaching the linguistic item, solving of process problems, which seem to be neglected in listening instruction, requires intensive exposure to the linguistic item in question. To this end, he recommends incorporating dictation into the listening instruction.

There is an empirical study that employed this framework to examine what kinds of decoding errors L2 listeners tend to make. Cross (2009) asked Japanese English learners of advanced level to listen to two pieces of BBC news and write down in English what they had understood. Their mistakes were then analyzed based on the framework, but he added another category called *intrusion problems*, meaning decoding problems derived from negative transfer from the L1[3]. Examination of errors revealed that the most common error for the process problem was improper word segmentation. There were two kinds. One typical pattern was the case where a monosyllabic unstressed word appeared either before or after a familiar word (e.g., "the size" was likely to be recognized as "decide"). Another pattern was that the participants tended to separate single words into two words wrongly when the words carried a secondary stress (e.g., "household" was misunderstood as "high solid," "high sold," or "house old"). Furthermore, vowels in word-initial syllables were also difficult for the participants to perceive (e.g., /ræ/ in "rammed" was sometimes perceived as /rau/ in "round"). These typical errors are all associated with the perception process rather than the syntactic parsing, indicating that speech perception is often a critical issue for Japanese English learners.

[3] For example, inability to identify the /b/-/v/ contrast (e.g., "ban" and "van") can be an intrusion problem for Japanese English learners because these consonants are not distinguished in their L1.

Nevertheless, errors like the last example ("rammed" vs. "round") may be due to the effect of high word frequency (i.e., "round" is much more common than "rammed"). If this is the case, then it is a text problem rather than a process problem. Cross mentions that it is often difficult to figure out which category a decoding error falls into. In fact, Field (2008) refers to six possible causes for a listening problem at word level as shown in Table 2.1, but it is hard to tell whether such cause as (c) is a text problem (i.e., the learner lacks the pronunciation knowledge) or a process problem (i.e., the learner is unable to distinguish similar sounds). However, categorization based on such framework is useful when diagnosing the problems behind learners' decoding errors.

Table 2.1

Possible Causes for a Listening Problem at Word Level (Field, 2008, p. 87)

a.	the learner does not know the word
b.	the learner knows the written form of the word but has not encountered the spoken form
c.	the learner confused the word with a phonologically similar one
d.	the learner knows the spoken form of the word but does not recognise it in connected speech generally or in this utterance in particular
e.	the learner recognised the spoken form of the word but failed to match it to any meaning
f.	the learner recognised the spoken form of the word but matched it to the wrong meaning

As the Cross (2009) study suggested, word segmentation is a common problem in decoding of L2 listeners. This issue is well-established in the literature (e.g., Anderson, 2005; Field, 2013; Levelt, 1993; Rost, 2016; Vandergrift & Goh, 2012), and a common rationale held by these researchers is that the absence of visual spaces between words in spoken language makes word segmentation difficult. Instead of spaces, prosodic information provides important cues to figure out word boundaries in spoken language. However, phonological systems are often different across languages; especially, Japanese and English share very few phonological

features. In the section 2.2.3, characteristics of spoken English are presented to discuss how they hinder speech perception of Japanese English learners.

2.2.2 Skills Involved in Speech Perception

As stated in the previous section, word segmentation is an important skill in listening, yet listening involves many other skills. Some researchers categorize these skills, which are diverse in the kinds of skills included. Aitken's (1978) list, cited in Buck and Tatsuoka (1998), is a simple one consisting of seven skills such as (1) understanding, or the ability to guess, the lexis; (2) understanding syntactic patterns and morphological forms; (3) understanding clues from stress and intonation patterns; (4) identifying the speaker's purpose; (5) drawing valid inferences about the context of situation; (6) recognizing the speaker's attitude to the listener and the subject; and (7) identifying the relevant rhetorical devices used by the speaker (pp. 121-122). Out of these, the first three skills (1-3) are associated with decoding, while the other four skills (4-7) are related to utilization.

Longer lists are provided by Richards (1983). Claiming that skills required for listening vary depending on the purpose, he developed lists of different sub-skills, or what he calls "micro-skills," for conversational listening (33 skills; Appendix 2B) and academic listening (18 skills; Appendix 2C). The former list is more general in that it has all the micro-skills including ones used in lower-level processing, whereas the latter list focuses on those used in higher-level processing such as the "ability to identify purpose and scope of lecture" and "the ability to deduce meanings of words from context" (p. 229). Eight micro-skills in the former list associated with speech perception are given in Table 2.2. These skills are allied with each other rather than operate independently. For example, as stated earlier, listeners need to be able to analyze prosodic features (3, 4, 5, 6, and 7) to detect word boundaries. It should also be noted that the first ability (1) indicates the role of short-term memory.

Table 2.2

Micro-Skills in Conversational Listening Employed in Speech Perception (Richards, 1983, pp. 228-229)

1. ability to retain chunks of language of different lengths for short period
2. ability to discriminate among the distinctive sounds of the target language
3. ability to recognize the stress patterns of words
4. ability to recognize the rhythmic structure of English
5. ability to recognize the functions of stress and intonation to signal the information structure of utterances
6. ability to identify words in stressed and unstressed positions
7. ability to recognize reduced forms of words
8. ability to distinguish word boundaries

Munby's (1978) taxonomy for the language skills is probably the most comprehensive one. Based upon findings of considerable research, his list comprises 54 ability groups (or *macro-concepts*), each of which consists of several subordinate skills. This adds up to about 260 skills (or *micro-concepts*) in total. It appears that seven macro-concepts composed of 39 micro-concepts are related to perception as given in Table 2.3 (see Appendix 2D for the micro-concepts). Most macro-concepts overlap with the abilities in Richards' list; yet, they can be distinguished in two respects. First, Munby's list does not have a category for the role of memory. Second, while in Munby's list there are many descriptions about recognition of sound features, none of the ability groups refers to recognition of words. However, it is speculated to be because words start to be recognized as soon as their sound features are identified. In this sense, Munby seems to regard recognition of sounds and words as one process as indicated previously by the integrate model of listening comprehension.

Table 2.3

Macro-Concepts Related to Speech Perception (Munby, 1978, pp. 123-126)

1. Discriminating sounds in isolate word forms (5)

3. Discriminating sounds in connected speech (6)
5. Discriminating stress patterns within words (3)
7. Recognising variation in stress in connected speech (4)
9. Recognising the use of stress in connected speech (3)
11. Understanding intonation patterns: neutral position of nucleus and use of tone (10)
13. Understanding intonation patterns: interpreting attitudinal meaning through variation of tone or nuclear shift (7)

Note. Micro-concepts of each ability group are omitted here (see Appendix 2D). The number of micro-concepts included in each macro-concept is indicated in the parenthesis.

2.2.3 Characteristics of Spoken English

When L2 learners start to listen to authentic speech of the target language, they may realize how difficult it is to recognize spoken words. Even learners who are very good at reading may strive to overcome this difficulty and end up losing confidence in their ability. In reading, since words are explicitly shown on paper or screen with spaces between them, learners do not have to find word boundaries. Moreover, due to the permanent feature of written texts, learners can read back and forth to inference unknown vocabulary or to check whether their interpretation is correct. In other words, the challenge of L2 reading is not perception of words but processing above this level such as syntactic parsing and utilization; therefore, knowledge about vocabulary and grammar plays a greater role in reading comprehension. Unfortunately, learners may not be able to exploit their linguistic knowledge as well in listening as in reading because of their poor perception ability, as Vandergrift and Goh (2012) point out:

> Recognizing a word in its written form or hearing it in isolation does not necessarily mean that we will recognize that same word in the context of rapid speech. This is the real challenge of listening comprehension: L2 listeners need to be able to rapidly parse words out from a stream of sound. (p. 24)

As mentioned here, a primary cause for comprehension failure in L2 listening is perception of spoken words. This is often the case not only because spoken language lacks visual spaces between words but also because it has unique properties that do not appear in written language. Buck (2001) lists several characteristics of English spoken input that might hinder listening comprehension. Likewise, Field (2003, 2008) describes inconsistent features of spoken English. The characteristics overlap between the two researchers to a considerable degree; therefore, a summary is provided below.

Phonological Modification. Phonological modification refers to a phenomenon where pronunciation of words changes when embedded in connected speech. Some important examples are: *assimilation* (e.g., "won't you" sounds like "wonchoo"); *elision* (e.g., "next day" is often pronounced as "nexday"); *intrusion* (e.g., /r/ in "far" is not usually pronounced in British English but inserted when followed by a vowel as in "far away"); and *weak/reduced forms* (e.g., "actually" is pronounced as /æʃli/). Each language has its own modification patterns based on very complex rules. Unlike native listeners, L2 listeners are very vulnerable to these modifications; thus, their perception processing may come under its negative influence.

Accents. Variation in speakers' accents also have a crucial impact on perception. Every language has accents, and we usually pick up our own depending on the geographical and social environment we belong to. As for English, since its speakers are all over the world, there is a variety of non-standard accents. Accented input is sometimes unintelligible even for its native speakers. Needless to say, its influence is more crucial for L2 listeners.

Prosodic Features. Prosodic features of spoken language may also be an obstacle for

speech perception. Since English is a stress-timed language, unstressed words are pronounced very rapidly and undergo considerable phonological modification. This makes perception of these words difficult. Moreover, English strong-weak rhythmic pattern sometimes induces listeners to detect the beginning of words wrongly. For example, a sentence like “Liz became a star. (ˈlɪz bɪˈkeim ə ˈstɑ:)” may be recognized as “Lizbe camer star. (ˈlɪzbɪ ˈkeimə ˈstɑ:)” by learners. This is a type of *redistribution* called *cliticisation*. Another type of redistribution is *resyllabification*, in which learners misperceive the end of words (e.g., “made out” is recognized as “may doubt”). These changes may cause critical perception errors for Japanese English learners because, as Sugito (2012) states, their mother tongue is a mora-timed language.

Speech Rate. Speech rate, which is usually measured by the number of words per minute (wpm), can affect perception too. The previous research has revealed that native speakers can understand fairly fast speech up to 275 wpm but that this was not the case with L2 listeners. Simply, the faster the speech is, the more difficult it is to perceive the input. This issue is one of primary interests of this dissertation, so it will be discussed in detail later (see section 2.5).

Hesitation Phenomena. Hesitation phenomena can also hamper understanding of spoken language. There are four types: (1) unfilled pauses (i.e., silence), (2) filled pauses (i.e., the use of fillers such as ‘uh’ and ‘well’), (3) repetitions, and (4) false starts (i.e., replacement of words or phrases that have just been spoken). Previous studies showed both positive and negative evidence for the influence of these hesitation phenomena on listening comprehension or speech perception. Buck concludes that pauses can assist listening when appearing at intervals between phrases, while random pauses do not.

2.2.4 Features of Effective Listening Activities

To overcome perception difficulty derived from these features of spoken English, learners need to undergo intensive listening using various materials containing the target feature. Lynch and Mendelsohn (2010) describe five characteristics of effective listening activities for improving linguistic skills, including "discriminating between similar sounds, coping with and processing 'fast speech,' and processing stress and intonational differences" (p. 194). Rost and Willson (2013) introduce several activities that aim to develop perception ability including *shadowing*. They claim that practice contributes to expansion of working memory capacity.

Another activity recommended in the literature is *dictation*. Field (2008) states that dictation can be introduced as a *micro listening task*, meaning that teachers can conduct the task in several minutes. Table 2.4 shows some exercise examples that focus on the spoken language features using dictation. Through dictation, learners realize why the listening breakdown occurred; therefore, dictation can be used for a diagnostic purpose. In fact, dictation was initially employed as a testing method rather than a language activity (e.g., Oller & Streiff, 1975).

Table 2.4

Examples of Perception Exercise Using Dictation (Field, 2008, pp.145 and 156)

Cliticisation	Learners transcribe short extracts from authentic recordings. Choose especially clips with many instances of schwa (/ə/) and clusters of weak syllables.
Focus on chunks	After playing an authentic text for comprehension, teacher replays sections of the recording representing chunks that occur frequently in natural speech. Learners transcribe.
Reduced forms in larger chunks	Teacher identifies formulaic chunks in authentic recordings, and asks learners to transcribe them. The most useful are noted down by learners and practiced orally as items of vocabulary.

2.3 The Efficacy of Dictation and Shadowing as Learning Methods for Listening Ability

2.3.1 Dictation Implemented as a Test Task and as a Measurement Task

One activity that has been shown to be beneficial for improving speech perception is dictation, which has a long history as an instrument for measuring language skills. According to Green (2014), the 1913 edition of the Certificate of Proficiency in English (CPE) included a 30-minute dictation. Dictation was used as a test task in the CPE from 1934 to 1966, but it received heavy criticism in the 1960s (Taylor, 2013). For instance, Oller (1971), who had stated that dictation was useful as a task for measuring overall English proficiency, received criticism from Breitenstein (1972) and others later. In response, Oller and Streiff (1975) reanalyzed the data and attempted to prove its utility.

In subsequent research, dictation received some recognition as a test task. Templeton (1977) tested the reliability and validity of partial dictation as a listening test. Furthermore, Morris (1983) conducted an analysis of learners' errors in dictation and classified them into four types (comprehension errors, meaning errors, structural errors, and spelling errors), making him conclude that dictation can demonstrate a variety of learners' abilities. Similarly, Weir (1993) noted that although dictation has limitations in terms of the speed with which the text is read and the length of the English sentences, its advantages include the fact that it enables the measurement of various language skills and is easy to implement and score.

Some researchers expressed similar opinions in the 2000s as well. Hughes (2003) described the characteristics of dictation, as a test task, as 'rough and ready' (p. 168), and stated that dictation makes it possible to determine the challenges faced by learners, such as weak forms. Brown and Abeywickrama (2010) also noted that the level of dictation difficulty can be easily manipulated by adjusting the positions of questions and the length of pauses, and that it is a useful task strongly linked to other language skills. According to Green (2014), dictation was introduced into the Pearson Test of English-Academic (PTE-A) in 2010, along with essay

and reading comprehension questions. In Japan, however, dictation is not included in large-scale tests, such as the Common Test for University Admission, and is widely recognized as a learning method rather than a test task. That also seems to be the general consensus worldwide.

Dictation is also used in research as a task to assess speech perception. For example, Sebastián-Gallés, Dupoux, Costa, and Mehler (2000) asked native Spanish speakers, who were learning a foreign language, to listen to audio recordings containing time-compressed speech in their target language, and write down what they heard, to assess their speech perception. In addition, Siegel and Siegel (2015) conducted learning activities aimed at extending bottom-up processing, and had students take dictation tests before and after the intervention to assess its efficacy.

2.3.2 Effectiveness of Dictation for Listening Ability

Dictation, as a learning method, is defined as "the transcription of the exact words that a speaker utters" (Rost, 2016, p. 172). According to Morris (1983), at the same time that dictation was increasingly being criticized as a test task, it received criticism as a learning method as well. As a result of the aforementioned error analysis, Morris identified three advantages of dictation as a learning method. The first advantage is that it improves short-term memory, which is a basic listening skill that enables us to process speech more efficiently. The second is that it improves the abilities to understand the context and utilize grammatical knowledge. This refers to the ability to determine whether the words that were heard are appropriate in the context. The third advantage is that it expands the knowledge learners required for spelling words.

Similarly, Davis and Rinvolucri (1988) stated 10 reasons why dictation is beneficial, focusing primarily on its application in a classroom (Table 2.5). Furthermore, as Field (2008) mentions, dictation is a highly convenient activity that can be completed in a short amount of

time between activities because it can be done in a variety of ways. Nation and Newton (2009) illustrate 10 useful exercises that integrate dictation. One of them is *guided dictation*, a type of partial dictation in which learners transcribe words that are missing in the transcript written on the blackboard.

Table 2.5

Ten Good Reasons of Using Dictation in a Classroom (Davis & Rinvolucri, 1988, pp. 4-8)

1. The students are active during the exercise
2. The students are active after the exercise
3. Dictation leads to oral communicative activities
4. Dictation fosters unconscious thinking
5. Dictation copes with mixed-ability groups
6. Dictation deals with large groups
7. Dictation will often calm groups
8. Dictation is safe for the non-native teacher
9. For English, it is a technically useful exercise
10. Dictation gives access to interesting text

Renewed research on dictation as an activity to improve the ability to perceive speech shed light on its significance in listening (e.g., Brown, 2011; Buck, 2001; Field, 2008; Nation & Newton, 2009; Rost, 2016; Suzuki & Kadota, 2018; Ur, 1984; Vandergrift & Goh, 2012). One ardent proponent of dictation is Field (2008). Field argued that dictation should be used to enhance the ability to accurately perceive speech in order to prevent running into the process problem (see 2.2.2 for types of decoding problems). Field also noted that traditional listening instruction prioritized pre-listening and listening, while neglecting post-listening, which is aimed at reflecting on listening problems. He advocated that dictation during post-listening can help determine what caused the problems.

Similarly, in relation to the features of dictation, Nation and Newton (2009) argued

that “Dictations facilitate language learning by making learners focus on the language form of phrase and clause level constructions, and by providing feedback on the accuracy of their perception”（p. 59). The authors also asserted that dictation is effective in developing speech perception. Rost (2016) also cited dictation and shadowing examples of intensive listening activities, centered on language forms. Dictation of short phrases, according to Ur (1984), is an activity aimed at improving speech perception; however, when longer English sentences are used, learners have to pay attention to the meaning, so it can also be used for comprehension training. In fact, most of the empirical research on dictation is centered around the development of listening comprehension, which is discussed in the following section.

Empirical research on the effects of dictation on listening skills began to flourish in the 1980s, and can be divided into the following four broad categories: (1) research focused on the development of listening comprehension (Chino, 2006; Kiany & Shiramiry, 2002; Marzban & Abdollahi, 2013; Oyama, 2009; Sugawara, 1999; Takeuchi, 1997; Yonezaki, 2014), (2) research focused on the development of speech perception ability (Kakehi et al., 1981; Suenobu et al., 1982), (3) research focused on both the above-mentioned (Cohen, 2015), and (4) research focused on the development of other language skills as well as listening (Brown & Hilferty, 1986; Jafarpur & Yamini, 1993; Mohammed, 2015; Rahimi, 2008).

A summary of these studies is presented by Table 2.6. As it shows, majority of the studies found that dictation was efficacious in improving listening skills, despite differences in terms of participant characteristics (native language, proficiency, and number of participants), training (amount, duration, method, and materials), and tests that were administered. This may indicate that dictation is an efficacious learning method for many learners, regardless of how it is implemented. However, comparisons with other skills revealed that dictation has a significant effect not only on listening but also on the other language skills and knowledge. This could be because dictation is an integrated-skills activity, and because learners' attention was also

directed at vocabulary, grammar, and other linguistic aspects through written output.

To amplify the effect of dictation for listening skills, it may be effective to increase the speed with which the teaching material is presented, so that learners' attention toward sounds will be enhanced. The purpose of this research is to compare the effectiveness of dictation and shadowing for improving speech perception ability. While shadowing is an "online" task where learners are engaged in listening and speaking simultaneously, dictation is an "offline" task where learners usually perform writing after listening. Dictation using fast-paced speech, referred to as "accelerated speech dictation" in the present research, may help to fill this gap and exhibit an equal, or even better, to shadowing.

Dictation, as previously stated, has two aspects: it is a task for both measurement and learning. In this study, dictation is implemented in both ways; thus, to avoid confusion, the term *written reproduction task* is used when referring to dictation implemented as a measurement task, distinguishing it from dictation implemented as a learning method[4].

2.3.3 History of Shadowing as a Listening Activity and its Theoretical Background

Compared to dictation, the history of shadowing for L2 learning is much shorter. According to Yashima (1988), shadowing had been implemented as a basic training of simultaneous interpretation. Her attempt to introduce shadowing to English education was novel; however, it did not receive as much attention as today until the 2000s, which was the decade when there were many shifts in the policy of English education in Japan. The most significant one may be that a listening test was newly introduced into the National Center Test

[4] The dictation task is also called "listening recall" (e.g., Henning, Gary, & Gary, 1983) or just "(written) recall" (e.g., Dupoux & Green, 1997; Pallier et al., 1998; Sebastian-Gallés et al., 2000). For the same reason as in the previous note, the term "recall" was avoided in this study. Partial dictation, which was derived from a reading cloze test, was called the "cloze procedure" (e.g., Templeton, 1977).

Table 2.6

Summary of Studies Revealing Positive Effects of Dictation on Listening Skills

Author	L1	*N*	Training	Training Duration	Measurement	Chief Result
Kakehi et al. (1981)	Japanese	68-75[a]	Dictation	6 months (monthly; 4 sessions)	Written reproduction	Error rates of function words decreased more sharply than content words.
Suenobu et al. (1982)	Japanese	60 -Upper: 20 -Lower: 40	Dictation	5 months (weekly)	Written reproduction	Reproduction rate of function words especially improved.
Takeuchi (1997)	Japanese	207 -Upper: 97 -Lower: 110	Simple dictation (n = 72); Dictation with translation (n = 66); Clued dictation (n = 69)	13 weeks (45 mins × 2 classes per week)	Unknown	All training groups with dictation showed significant improvement; However, there was an interaction between training type and proficiency.
Kiany & Shiramiry (2002)	Iranian	60 -Exp: 30 -Con: 30	Exp: Dictation Con: Listening	11 classes	Comprehension	Only the experimental group showed significant improvement.
Marzban & Abdollahi (2013)	Iranian	60 -Exp: 30 -Con: 30	Exp: Partial dictation Con: Listening	20 classes (11 dictation sessions)	TOEFL Listening	Only the experimental group showed significant improvement.
Yonezaki (2014)	Japanese	16	Dictation	8 weeks (30 minutes per week)	Center Listening	Significant improvement with a large effect size. Most errors were related to function words.
Cohen (2015)	Japanese, Chinese	34	Exp: Dictation Con: Minimal pairs	12 weeks (10 minutes per week)	TOEFL Bridge, Written reproduction	The experimental group showed better score improvement in both tests.
Mohammed (2015)	Iraqi	50 -Exp: 25 -Con: 25	Exp: Dictation Con: Listening	10 months (10 minutes × 50 sessions)	Grammar, Vocabulary, Reading, Listening (all are in TOEFL-like formats)	The experimental showed significant score improvement in all the tests, while the control group showed it in the vocabulary test. The score gains of vocabulary and reading tests were greater than those of grammar and listening tests.

Note. [a]The number of participants differed depending on the test.

in January of 2006. With such changes, there was an increase of interest in how to foster students' communicative ability in English, and shadowing started to gain a wider recognition among English teachers. In fact, we can find many practical reports in English education magazines published in the early 2000s (e.g., Kougo & Kubono, 2004; Oshima, 2003; Takei, 2002), describing how to introduce shadowing into the class effectively.

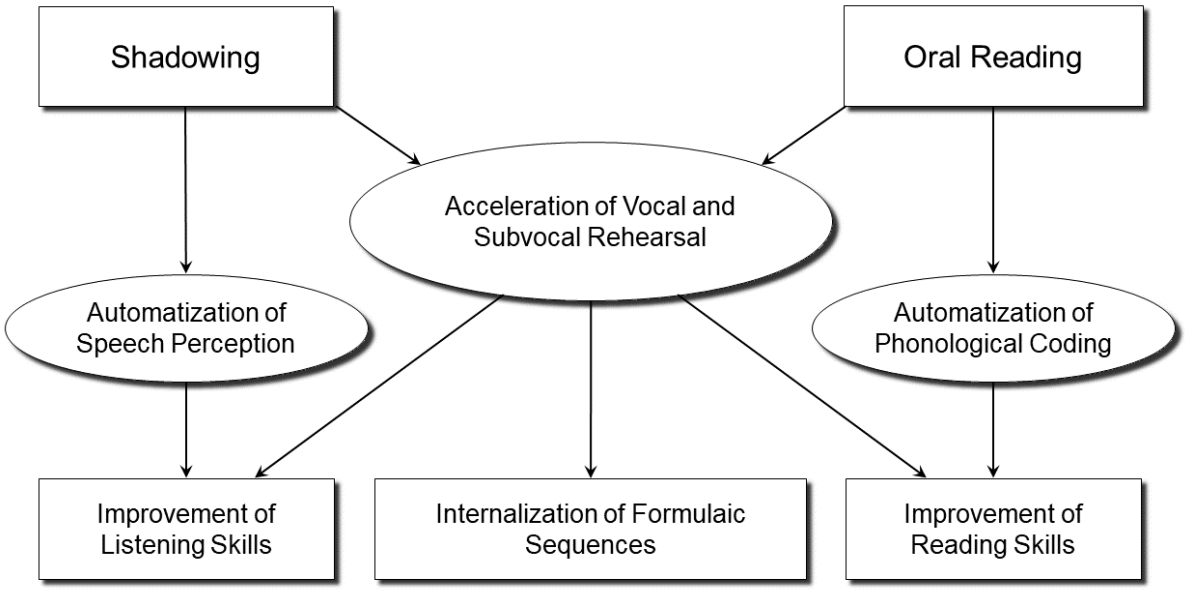
Research on shadowing also flourished in the 2000s. The landmark event was definitely the publishment of Kadota's (2007) book on shadowing. In this book, he attempted to provide an elaborate and comprehensive explanation about why shadowing is effective by referring to numerous findings from the psycholinguistic research, while comparing with oral reading. One of his major claims was that shadowing would help learners develop their speech perception ability by riveting their attention on sound features of the passage and enhancing the phonological loop. The author myself conducted a series of studies to test Kadota's claim by exploring the cognitive processes involved in shadowing (Oki, 2010a, 2010b; O'ki, 2011, 2012b, 2014). A major finding of these studies was that shadowing did promote learners' phonological processing rather than semantic processing.

Several years later, Kadota published another book (Kadota, 2012), in which he presented a model that illustrated how shadowing and oral reading contribute to language acquisition (see Figure 2.7). According to him, the effectiveness of the two activities is twofold. First, by performing the activities repeatedly, learners can develop the ability to construct phonological representation automatically from the input. This leads to automatization of either speech perception (in the case of shadowing) or phonological coding (in the case of oral reading). He adds that automatization of these processes enables learners to save their cognitive resource for meaning processing. Second, learners can enhance the efficacy of their vocal and subvocal rehearsal processes, which plays a significant role in storing learning items in their long-term memory. Kadota assumes that this process allows learners to internalize so-called

formulaic sequences, or chunks. Kadota and Tamai (2004) state that the first stage of L2 learning is to memorize lexical chunks such as "What did you say?" as a whole, naming this developmental stage *holistic chunk learning*. This is likely to take place during shadowing practice, as O'ki (2012b) revealed that lower-level learners were able to shadow familiar phrases more successfully than those unfamiliar to them, suggesting that they depended on phrasal knowledge during shadowing.

Figure 2.7

The Effects of Shadowing and Oral Reading (Based on Kadota, 2012, p. 135)



Note. This model is a revised version of the one on Kadota (2007, p. 34).

2.3.4 Effectiveness of Shadowing for Listening Ability

The author's works uncovered some characteristics of the cognitive processes performed in shadowing, but they were not aimed at revealing its training effect on the improvement of listening ability. With Yashima's (1988) work as a starter, several studies have investigated whether shadowing actually develops learners' listening ability (Hamada, 2017; Sato & Nakamura, 1998; Suzuki, 2007; Tamai, 1992, 1997, 2005; Tateuchi, 2005; Yanagihara,

1995), which are summarized in Table 2.7.

Two things should be noted. First, although the studies vary in terms of the sample size and the training length, the results generally indicate that shadowing contributed to the improvement of listening ability. This is evidenced by the result that the shadowing group demonstrated significant improvement in many of the studies. Several studies compared shadowing with other tasks such as dictation and comprehension activities, but the results are somewhat miscellaneous. For example, the shadowing group outperformed the comprehension group in such studies as Yanagihara (1995) and Tateuchi (2005), while they failed to do so in some studies (Sato & Nakamura, 1998; Suzuki, 2007).

Second, as suggested by Tamai's (1997, 2005) and Suzuki's (2007) studies, shadowing can exhibit the effect with training of several days. Tamai theorized that this was because shadowing had enhanced learners' working memory, which plays a crucial role in decoding the input. Decoding is a fundamental sub-skill of listening comprehension, so its development may appear early, in advance of the development of listening ability. Except Tamai's (2005) experiments, there is little evidence to back up this hypothesis. The fact that correlations between the shadowing test and the listening test tended to be weak could be an indirect evidence because the low correlation indicates that the shadowing skill and the listening ability may not develop in parallel[5].

Third, the effectiveness of shadowing may be affected by learner's proficiency level or the difficulty of training material. Three studies (i.e., Sato & Nakamura, 1998; Tamai, 1997; Yanagihara, 1995) yielded the finding that shadowing was more beneficial for lower-level learners. On the other hand, mixed results were observed in Suzuki's (2007) study. That is, the low- and middle-level group students learned best by shadowing the materials they had already

[5] Takayama (2007) also reports a weak correlation (r = .09).

studied, whereas the upper-level students were able to manage the new materials. Inherently, the two variables are associated with each other for the reason that whether a learner considers a material to be difficult is determined by his/her proficiency level. In order to provide learners with shadowing materials of an appropriate level, teachers must evaluate the level of their learners precisely or establish an environment where learners can choose their own materials matching their level.

Table 2.7

Summary of the Research on the Effect of Shadowing on the Listening Ability

Author(s) (year)	*N*	Grade	Length	Progress	vs. Other Groups	Proficiency
Yashima (1988)	9-16	College	2 months	Yes	-	-
Tamai (1992)	94	High S.	3.5 months	Yes	> Dictation	-
Yanagihara (1995)	90	College	2 months	Yes	> Dictation, > Comprehension	Low[a]
Tamai (1997)	25	College	5 days	Yes	-	Low
Sato & Nakamura (1998)	131	College	1 year	N/A	= Comprehension	Low
Tamai (2005)						
Experiment 1	93	College	3 months	Yes	= Dictation, > NI	-
Experiment 2	51	College	5 days	Yes	> NI	-
Tateuchi (2005)	77	College	10 weeks	Yes	> Comprehension	-
Suzuki (2007)						
Practical Study 1	27	High S.	5 days	Mixed[b]	* Comprehension	-
	112	High S.	3 months	No	< Comprehension	Mixed
Practical Study 2	114	High S.	3 months	No	= R&L, = Repetition	Mixed
Hamada (2017)						
Classroom Exp. 1	43	College	1 month	Mixed[c]	-	-
Classroom Exp. 2	43	College	1 month	Mixed[d]	-	Low

Notes. "Progress" indicates whether the shadowing group showed significant improvement; "High S." refers to high school students; N/A = Not analyzed; The symbols such as ">, <, =" "Proficiency" means whether the proficiency level of shadowing group affected their results on

the post-tests.

[a] The lower-level group showed the most remarkable improvement when the post-test consisted of linguistic items they learned in the training, but this effect did not appear when the post-test consisted of new items.

[b] Significant improvement was observed when shadowing was incorporated before studying the material.

[c] There was significant improvement in the phonemic perception test, but the scores in the comprehension test improved only for the easy passages.

[d] There was significant improvement in the phonemic perception test, but the significant score improvement was observed only for the low-level group when listening to the easy passages in the comprehension test.

2.3.5 Hamada's Shadowing Experiments on Speech Perception Ability

Among the studies listed above, Hamada (2017) is the only work that aimed to gauge the effect of shadowing on speech perception ability. He conducted two experiments. In the first one (Classroom Experiment 1: Phoneme Perception and Listening Skills Improvement), he engaged 43 national university students in shadowing exercise for a month (two sessions a week, nine sessions in total; each session is 15 to 20 minutes). Improvement of participants' abilities for speech perception was measured through a phoneme perception test (test using a partial written reproduction task) and comprehension tests consisting of easy and difficulty passages, using a pretest-posttest design. The analyses revealed that, while the average score of the phoneme perception test improved significantly, the comprehension tests showed that only the average score of the easy passages showed significant increase.

In his second experiment (Classroom Experiment 2: Lower Listening Proficiency Learners' Improvement), using the same data obtained in the previous experiment, Hamada examined whether the score improvement in the two test tasks differed depending on the proficiency level (i.e., low- and middle-levels). For the phoneme perception test, a two-way ANOVA showed no interaction between time (pre-test, post-test) and proficiency (low-level,

middle-level) but significant difference for both main effects, indicating that the two proficiency groups improved equally. In contrast, the comprehension test revealed contrastive results for easy and difficult passages; only the low-level learners significantly improved their scores with the easy passages, whereas neither group showed improvement with the difficult passages.

In both experiments, Hamada concludes that shadowing is effective for improving phonemic perception skill; yet, this is still open to question. To prevent the participants from guessing target words from the context, they were asked to listen for only function words such as articles and prepositions. However, function words have a limited variety and can be guessed easily by syntactic information or phrasal knowledge; thus, perception of these words may be more top-down than bottom-up. Furthermore, as pointed out earlier, the major problem in perception is word segmentation. Segmentation errors such as cliticization (e.g., "Liz became a star." is recognized as "Lizbe camer star.") and resyllabification (e.g., "made out" is recognized as "may doubt") are made only when two or more words are connected. Also, unstressed words near familiar words tend to be misheard by L2 listeners. For these reasons, it is more reasonable to have learners listen for content words as well, especially when in combination with function words.

2.4 Influence of Speech Rate on Listening

There are many factors that affect listening in a foreign language, such as the listener's native language, proficiency level, difficulty of the teaching materials, and presence of noise. The "speech rate" of the teaching materials is considered to be one such factor (e.g., Buck, 2001; Rost, 2016; Rubin, 1994). The speech rate in English is usually expressed in words per minute (wpm), which is calculated by dividing the total number of words by the speaking time (minutes). Although the wpm derived by this formula does not faithfully reflect the actual speech rate and is somewhat simplistic (O'ki, 2012), it is widely used as a measure of the speed

of speech because it is practical and easy to calculate. The average listening score for the second level of the EIKEN Test in Practical English Proficiency (commonly known as "EIKEN") is around 140 wpm. For the National Center Test administered by Daigaku Nyuushi Center, the average throughout all sections was around 160 wpm (Komori, 2010). The range of the TOEFL iBT seems to be quite large, from less than 120 wpm to more than 200 wpm (Sawaki & Nissan, 2009). According to Tauroza and Allison's (1990) standard, a speed of 200 wpm is the fastest (i.e., faster than normal) for a monologue format (i.e., one person speaking one way, not a dialogue), as in radios or lectures. It is the speed at which many Japanese learners of English have difficulty in terms of listening comprehension.

Studies regarding the role of speech rate when listening to native speakers often use artificially time-compressed speech (known as time-compressed speech or accelerated speech), such as that produced by devices or computers. Such studies have shown that native speakers can hear even fairly fast speech. For example, in a study by Wingfield and Nolan (1980), 28 university students were asked to listen to English speech at a rate of 185 wpm compressed to 80%, 70%, and 60% of its length (equivalent to 231, 264, and 308 wpm, respectively). Then, the students were asked to reproduce what they heard aurally when the tape was stopped. The results showed that the students were able to reproduce 85% of the words at 80% (231 wpm) and 75% at 60% (308 wpm). Beatty, Behnke, and Goodyear (1979) showed that there was no difference in listening comprehension between 140 and 245 wpm when 300 native speakers of English were asked to listen to speech at different speeds (140, 175, 210, 245, and 280 wpm) and answer true or false questions about the content.

2.4.1 Empirical Studies on the Effect of Speech Rate on Listening

The results of these studies show that native speakers can understand even fairly fast speech of around 250 wpm. However, non-native speakers and learners are unable do the same.

Some studies suggest that the speed at which it is easy for learners to listen is much lower (e.g., Zhao, 1997). The following paragraphs review seven studies in which the effect of speech rate on English learners' listening ability was examined. Types of English sentences that are difficult for learners in terms of speech rate are also explored below.

Kelch (1985). Kelch focused on the use of "foreigner talk" by native speakers when speaking to non-native speakers and examined whether its characteristics, speech rate, and linguistic adjustments promoted speech perception. The participants were 26 ESL learners who were native speakers of Japanese, Chinese, Ilocano, and Spanish (with TOEFL scores ranging from 490 to 580). Four types of sentences were used, and two types of features were multiplied: adjusted speech rate (±speed; +speed: 191/200 wpm, -speed: 124/140 wpm) and linguistic adjustment, such as paraphrasing to easier expressions (±modification). The results of a two-way analysis of variance (speed × modification) showed that only the main effect of speed was significant, indicating that sentences read at a slower speech rate were easier to understand. The results suggest that learners may find it easier to listen if the speech rate is slowed down. This approach was more effective than paraphrasing or simplifying the syntactic structure.

Conrad (1989). Conrad compared the effects of speech rate for English learners and native English speakers. The participants were 28 native Polish EFL learners (high-level skill = 17; medium-level skill = 11) and 29 native English speakers. Sixteen English sentences consisting of 9–11 words to be read at approximately 180 wpm were prepared (e.g., "At school, the dormitories are quiet during exam time"). All were compressed to 91%, 83%, 71%, 56%, and 40% of their length (equivalent to 196, 216, 253, 320, and 450 wpm, respectively). Subsequently, the participants were asked to listen to each sentence once—in order, beginning with that at the fastest speech rate. Then, they were asked to reproduce the sentences in writing.

The results of the analysis (see Table 2.8) showed that the native speakers reached a 96% reproduction rate at 320 wpm (Trial 2) and were able to reproduce almost 100% of the words thereafter, but the reproduction rate of the learners was much lower. Even in trial 5, only 61% of the words were reproduced. There was also a marked difference by proficiency level; learners with medium-level skills could reproduce only 44% of the words even at the slowest speed, suggesting that the speed of 196 wpm was a major barrier for learners with medium to low-level skills.

Table 2.8

Mean Percentage of Sentence Items Recalled by Group for Each of Five Speech Rates

		Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
		40%	56%	71%	83%	91%
Group	*n*	(450 wpm)	(320 wpm)	(253 wpm)	(216 wpm)	(196 wpm)
Native Speakers	29	66	96	99	100	100
High-Level Skill	17	5	27	49	64	72
Medium-Level Skill	11	1	11	23	35	44

Note. This table is based on Conrad (1989, p. 7).

Blau (1990). Blau conducted two experiments on the effect of speech rate. In the first experiment, 72 Polish university students and 100 Puerto Rican university students studying ESL were asked to listen to English sentences at two different speeds (approximately 145 and 170 wpm) and answer multiple-choice comprehension questions. To examine the effect of syntactic structure as well as speech rate, three types of English sentences with different syntactic features were prepared (version 1 = sentences with easy syntax structure; version 2 = sentences with syntactic complexity and cues to the structure; version 3 = sentences with syntactic complexity and no cues). An analysis of covariance between speed (speed) and English sentence type (version), with proficiency test scores as the covariate, showed no significant main effect of speed or version in either participant group (i.e., Polish students and

Puerto Rican students). In other words, the results showed that there was no difference in listening comprehension between 145 wpm and 170 wpm.

In the second experiment, three types of English sentences were prepared to examine the effects of speech rate and pauses. The first was spoken at a natural speed (200 wpm), whereas the second was slowed down to 185 wpm. The speech rate in the third sentence was 185 wpm, with a three-second pause inserted (roughly) after every 23 words, lowering the wpm to 150. The sample included 106 people who had participated in the first survey. The results of the analysis of covariance showed that the scores for the third English sentence (150 wpm) were significantly higher than those for the other two English sentences. In other words, the insertion of pauses made listening comprehension easier. Interestingly, scores did not improve when the speed was reduced from 200 wpm to 185 wpm.

Griffiths (1990, 1992). Griffiths conducted two experiments regarding speech rate with elementary school teachers undertaking English training at a university in Oman. In the first experiment (Griffiths, 1990), 15 participants listened to three passages of 350–400 words each at different speeds (*slow* = 100 wpm; *average* = 150 wpm; *fast* = 200 wpm) and were asked to answer 15 true or false questions per passage. The results of the experiment showed that performance at the *fast* rate was significantly lower than that at the other two speech rates, but there was no significant difference between *average* and *slow* rates (i.e., 200 wpm < 150 wpm = 100 wpm). In the second experiment (Griffiths, 1992), a similar study was conducted with 24 participants using different speech rates. The three speech rates used in this experiment were *slow* = 127 wpm, *average* = 188 wpm, and *fast* = 250 wpm. The results of the experiment showed that performance at the *slow* rate was significantly higher than performance at the other two speech rates, but there was no difference in performance for *average* and *fast* rates (i.e., 250 wpm = 188 wpm < 127 wpm).

What is unique about Griffiths' experiment is that various speech rates were tested under similar conditions. Although the participants and teaching materials were not exactly the same, the combined results of the two experiments suggested that there was no difference in performance for the three speeds above 180 wpm (i.e., 188, 200, and 250 wpm) and the three speeds below 150 wpm (i.e., 150, 127, and 100 wpm). It is interesting to note that the results at 200 wpm and 250 wpm were equivalent to those at 188 wpm. This finding suggests that there may be a difficulty barrier at 180–190 wpm. Based on the results of the above experiment, Griffiths argued that it is better to speak to intermediate or lower learners at a speed below the average (120–130 wpm).

Zhao (1997). Zhao, on the other hand, believed that the right speed depends on the listener and investigated whether there was a positive effect on listening comprehension when participants were allowed to choose a speed that they found easy to listen to. The participants were 15 ESL learners (from China, Colombia, Korea, Taiwan, Turkey, and Venezuela). In addition to the choice of speed, Zhao added the presence or absence of repetition as a factor and compared the four conditions (see Table 2.9). The results of the experiment showed that the scores were significantly higher when a speed adjustment (Conditions 2 and 3) was allowed than when it was not (Conditions 1 and 4), indicating that listening at a speed that is easy to listen to makes it simpler to understand. Incidentally, the average speed chosen by the participants for easy listening was around 130–140 wpm. On the other hand, there was no effect from repeated listening, indicating that even if there was an opportunity to listen repeatedly, a full understanding would not be achieved if the speed was too fast. The original speed of the English text used in this experiment was 194 wpm, so English text read at this level of speed may be too difficult for many learners.

Table 2.9

Types of Listener Control in Four Conditions

Type of control	Condition 1	Condition 2	Condition 3	Condition 4
Speed	No	Yes	Yes	No
Dynamic	No	No	Yes	No
Repetition	No	No	Yes	Yes

Notes. This table is based on Zhao (1997, p. 53). Speed = listener's control over speech rate before listening; Dynamic = listener's control over speech rate during listening.

Koya (2017). Similar to Zhao in the aforementioned study, Koya tested whether it was easier to listen when learners were allowed to choose their preferred speed. The participants were 41 first-year students in the science and technology department of a Japanese university. Koya stated they were beginner-level English learners because their average TOEIC score upon entering the university was 310. Koya prepared two types of EIKEN Level 2 listening questions of equal difficulty (both with an average speed of around 130 wpm and 15 questions each) and had the participants answer them consecutively. In the first found, students were asked to listen at the original speed, and in the second round, they were asked to listen at their preferred speed to determine whether their listening comprehension improved in the second round. The results showed that 32 out of 41 participants chose the slower speed in the second round (about 110 wpm on average at 0.85x speed), and their scores increased significantly. When a questionnaire was conducted after the session, many of the participants who slowed the speed mentioned that they could hear the words more easily, which suggested that adjusting the speech rate made speech recognition easier.

2.4.2 Summary of the Speech Rate Studies

The following table (Table 2.10) summarizes the results of the study described above. Although there are differences between studies in terms of the participants' native languages, the English learning environment, size of the study, and type of task, most of the studies

reviewed herein showed a significant drop in performance when the speech rate was 180 wpm or faster[6]. O'ki (2012a), while noting the ambiguity of the unit wpm, stated that this speed seems to be equivalent to the speed at which the holistic sound-processing system proposed by Kohno (1993, 2001, 2007) becomes active. This system is an auditory perceptual system where listeners recognize linguistic stimuli as a coherent whole when syllables are inputted every 330 milliseconds or faster[7]. Learners who hear an English sentence that exceeds this speed can make use of it and perceive the sentence efficiently. To do so, they must have a wealth of linguistic knowledge, such as that of a native speaker; however, learners with limited linguistic knowledge may not be able to keep up with the processing. The participants in Kelch's study had high TOEFL scores (490–580), but even learners at this level had difficulty listening to English sentences at 191–200 wpm, suggesting that 180 wpm is a significant barrier for most learners. Given these, dictation may exhibit a better effect on speech perception ability when using faster speech than when using normal-speed speech.

2.5 Links to the Present Research

This chapter has discussed how speech perception ability is crucial in listening comprehension. In this dissertation, perception is viewed as a process consisting of two processings; input decoding (phonological analysis of the input) and word recognition (identification of words based on the sound information). It is well-known in the literature that segmenting words from connected speech is a major problem for L2 listeners in perception, caused by spoken English features such as phonological modification (e.g., assimilation and

[6] An exception is the study by Koya (2017), but it is difficult to put it in the same category as other studies because of the low proficiency level of the participants and slow speech rate (130 wpm) for English the text used in the test.

[7] This occurrence rate of syllables is comparable to roughly 180 syllables per minute (6000 milliseconds). O'ki (2012a) estimated that this rate would be close to 180 wpm because wpm includes pause time (see his article for further explanation).

Table 2.10

Summary of the Studies on the Speech Rate Role in L2 Listening

Author	L1	*N*	Task	Result
Kelch (1985)	Various	26	Written Reproduction	Sig. between 140 and 191 wpm
Conrad (1989)	Polish	28	Written Recall	High: 72% recalled at 196 wpm; Middle: 44% recalled at 196 wpm
Blau (1990)	Study 1: Polish, Spanish	172	Multiple-Choice	No Sig. between 145 and 170 wpm
	Study 2: Polish, Spanish	106	Wh-Questions	Sig. between 150 and 185 wpm
Griffiths (1990)	Unknown	15	T/F Questions	Sig. between 150 and 200 wpm
Griffiths (1992)	Unknown	24	T/F Questions	Sig. between 127 and 188 wpm
Zhao (1997)	Various	15	Multiple-Choice	Listener's control of speech rate outweighed the effect of task repetition

Notes. Multiple-Choice = multiple-choice comprehension questions; T/F = true or false; Sig. = significant difference.

elision). Redistribution, meaning that a listener misperceive the word boundaries, is a typical phenomenon of word segmentation failure. Researchers claim that automatization of speech perception ability contributes to the release of working memory, enabling L2 listeners to focus on meaning processing. In the same vein, expansion of working memory capacity through developing the phonological loop will make listening easier.

As useful exercises for improving speech perception ability, many researchers and practitioners have been interested in dictation and shadowing. There is extensive research that has attempted to investigate the effectiveness of these exercises, most of which has suggested that they are efficacious for the improvement of listening comprehension ability. However, there is little research aimed at revealing its effect on speech perception ability. One exception is Hamada (2017), who studied whether shadowing would improve learners' phonemic perception skill. His study was pioneering but limited in that it focused on perception of only function

words, although segmentation failure often involves perception of content words. Therefore, it is necessary to examine whether dictation and shadowing will promote perception of various words.

Dictation and shadowing are identical in that both are integrated-skills tasks involving listening; however, they can be distinguished in terms of whether the two language skills (i.e., listening and writing/speaking) are performed simultaneously. Dictation, where writing is usually performed offline, may be cognitively less demanding than shadowing. Studies on the role of speech rate in listening suggest that dictation exercises using accelerated speech beyond 200 wpm will bring out a better effect on speech perception ability.

Given these backgrounds, the present research attempts to compare the effects of accelerated speech dictation and shadowing on speech perception ability. In Study 1, in an attempt to confirm that accelerated speech hampers learners' performance in perception, university students were asked to transcribe English passages of three kinds of speeds. Participants' errors were analyzed to figure out what problems learners tend to face in perception. Study 2 was then aimed at examining the effectiveness of a short-term training (only one class) with either accelerated speech dictation or shadowing on speech perception ability, which was measured through a written reproduction task. Since this task is an integrated-skills task involving writing thus affected by test takers' spelling knowledge, Studies 3 and 4 aimed to develop a new measurement task that can be implemented without productive skills. To this end, validity and reliability of *word count task* and its revised version were compared with those of written reproduction tasks. In the final study, Study 5, using measurement tasks recommended by the previous studies, investigated the effectiveness of long-term training (two months) with accelerated speech dictation and shadowing. Text analyses on participants' weekly journals were also conducted to obtain qualitative evidence for the training effect.

Chapter 3

Study 1: The Effect of Speech Rate on Speech Perception

3.1 Study Goal

As discussed in the previous chapter, numerous studies on the role of speech rate in listening comprehension find that English learners are likely to experience greater difficulty as the speech rate increases. Based on Kohno's (2001) theory, the author noted the possibility that speech rates faster than 200 wpm could be those at which the holistic sound-processing system works more dominantly than the analytic sound-processing system. Thus, its perception becomes much more difficult compared to when English learners listen to speech slower than that rate. As linguistic knowledge is considered to play an important role in perceiving fast speech, the impact of speech rate may be more crucial for learners with lower proficiency. Moreover, it is unclear what kinds of words or sounds are especially difficult for learners to read at very fast rates.

Based on these backgrounds, the research questions (RQs) of this study are as follows:

RQ1-1: Does the reproduction rate of English speech decrease drastically when the speech rate exceeds 200 wpm? Is the influence of speech rate on perception more critical to lower-level learners?

RQ1-2: What kinds of words are difficult to perceive when the speech rate exceeds 200 wpm?

3.2 Method

3.2.1 Participants

The participants of this study were 40 university freshmen in English education (18 men and 22 women) who were taking a basic grammar course. The mean score of the listening comprehension section of the TOEFL ITP (Institutional Testing Program) held one month prior to data collection was 45.33 (*SD* = 4.69) out of 68, with a range from 33 to 56. The participants were divided into two proficiency groups based on their TOEFL listening scores. The average score of the upper group (*n* = 21) was 48.57 (*SD* = 2.44), while that of the lower group (*n* = 19) was 41.74 (*SD* = 3.90). Since the score distribution of both groups was not normal, a Mann-Whitney U-test was performed to compare listening ability. The test found a significant difference between the mean ranks of the upper group (30.00) and lower group (10.00), $U = 0.00$, $z = -5.44$, $p = .000$, with a large effect size ($r = .86$).

3.2.2 Materials

In this study, participants engaged in a written reproduction task with three different speech rates (approximately 135, 175, and 215 wpm). To negate the practice effect of listening to the same passage repeatedly, three different passages selected from the Grade 2 of the STEP (the Society for Testing English Proficiency) Eiken listening test administered in the fall of 2011 were used (The passages are cited in 3.3.2).

As summarized in Table 3.1, the speech rates of these passages were controlled at less than 140 wpm (Passage A = 136.6 wpm; Passage B = 135.0 wpm; Passage C = 130.3 wpm). However, the passages varied slightly with respect to the topic, word level, and grade level (FKGL). To minimize the possible effect of passage difficulty by counterbalancing the order of presentation of the three passages among participants, two faster versions were produced for each passage. For this purpose, a free audio-editing software called Audacity

(https://www.audacityteam.org/) was used to accelerate each passage by 30% and 60%. As a result, each of the three passages had three speech rates, that is, a total of nine materials were obtained. The average speech rate of the passages speeded up by 30% was 174 wpm, whereas that by 60% was 214 wpm. The nine materials were divided into three sets (Set 1-3), each of which comprised three passages with different speech rates (see Table 3.2).

Table 3.1

Difficulty of the Three Passages (A, B, and C) Used in the Written Reproduction Task

	Topic	Total Words	Duration (seconds)	JACET 1000 tokens	Words / sentence	FKGL	wpm [×1.3, ×1.6]
A	Employee Meeting	66	29	79.1%	11.0	5.3	136.6 [178, 219]
B	Business Program	63	28	92.1%	12.6	5.9	135.0 [176, 216]
C	Snow Noise	63	29	79.7%	12.6	7.1	130.3 [169, 208]

Note. "JACET 1000 tokens" refers to the number of tokens that are listed in 1000-word level based on the JACET 8000 Word List (checked at: http://someya-net.com/wlc/index_J.html). FKGL = Flesch-Kincaid Grade Level. wpm = words per minute. Numbers in [ ] are speech rates of the faster versions.

Table 3.2

Results of the Written Reproduction Task by the Three American Students

Material	Original	30% UP	60% UP
Set 1 (Student A)	[Passage A] 65/66 words (98.5%)	[Passage B] 63/63words (100.0%)	[Passage C] 62/63 words (98.4%)
Set 2 (Student B)	[Passage B] 61/63 words (100.0%)	[Passage C] 63/63 words (100.0%)	[Passage A] 66/66 words (100.0%)
Set 3 (Student C)	[Passage C] 61/63 words (96.8%)	[Passage A] 65/66 words (98.5%)	[Passage B] 63/63 words (100.0%)

To confirm that these accelerated aural passages maintained sufficient intelligibility, three exchange students from the United States (one female and two male) were recruited.

Each student listened to one of the material sets (Set 1-3) through headphones on a computer, and then transcribed the words while being permitted to pause the audio where necessary. All of them completed the task in approximately five minutes and, as Table 3.2 indicates, were able to transcribe almost all the words accurately. While they failed to reproduce a few words, most of which were unstressed function words (e.g., *the* and *but*), it was evident that these errors were not caused by the acceleration of the speech rate, as the exchange students missed the same words even at the original rate. In other words, the errors were probably careless mistakes due to lack of attention, and it can be concluded that the accelerated passages were as intelligible as the original ones.

3.2.3 Procedure

The study was conducted in a computer room equipped with sufficient computers for each participant. First, 40 participants were randomly divided into three groups (Set 1 = 12, Set 2 = 14, and Set 3 = 14) by receiving a test booklet on which one of the three sets was labeled (see Appendix 3A). To compare the listening abilities of these three groups, a one-way ANOVA was conducted after the study on their TOEFL ITP listening scores (Set 1 Group, $M = 45.75$; Set 2 Group, $M = 45.14$; Set 3 Group, $M = 45.14$), and the results revealed that all these means were statistically equal, $F(2, 39) = 0.07$, $p = .935$, $\eta_p^2 = .004$.

The author explained the study's aim and asked the students to do their best, although their performance in the study would not affect their course grades. When all the computers were turned on, the participants downloaded the designated material set from the university's public folder and adjusted the volume while listening to sample audios through stereo earphones. Subsequently, the participants engaged in the written reproduction task using the three passages. For each passage, they were given 10 minutes, during which they tried to transcribe as many words as possible. They were permitted to pause and repeat the

audio at any time but not to return to the previous passages.

3.2.4 Scoring

Participants' performance on the written reproduction task was evaluated by calculating the percentage of words that were reproduced correctly. Three raters, including the author, performed the evaluation, and each rater evaluated one of the three sets. To simplify scoring and maintain inter-rater reliability among the three raters, all misspelled words were judged as incorrect. Hence, the omission of inflectional morphemes, such as plural *-s* and past tense *-ed*, was perceived as an error. However, a point was given when participants noted a homophone (e.g., *their*) instead of the correct word (e.g., *there*) because they were able to perceive the sound correctly (i.e., the input decoding process was successful). For the same reason, the absence of periods and commas was also ignored. As a result of the scoring, the internal consistency of each set measured by Cronbach's α was maintained at a sufficiently high level, as shown in Table 3.3, indicating that the reliability of the test and scoring was assured.

Table 3.3
Internal Consistency of Each Material Set Measured by Cronbach's α

Material	Original	30% UP	60% UP
Set 1	[Passage A] = .80	[Passage B] = .87	[Passage C] = .84
Set 2	[Passage B] = .79	[Passage C] = .75	[Passage A] = .72
Set 3	[Passage C] = .88	[Passage A] = .91	[Passage B] = .86

3.2.5 Analyses

To examine RQ 1-1, a mixed design of a two-way ANOVA in a 2 (Lower, Upper) × 3 (Normal, 30% UP, 60% UP) was conducted. To answer RQ 1-2, the reproduction rate for each word in the faster speeds (i.e., 30% UP and 60% UP) was also calculated, and further

analysis was conducted to find typical errors.

3.3 Results and Discussion

3.3.1 Reproduction Rates in the Three Speech Rates

Table 3.4 and Figure 3.1 show the reproduction rates of the two proficiency groups for the three speech rates. The reproduction rates of both the upper and lower groups declined gradually as the speech rate increased. The influence of speech rate seems to be slightly critical for the lower group, because its reproduction rate dropped by more than 14%, whereas that of the upper group dropped by only 10%.

Table 3.4

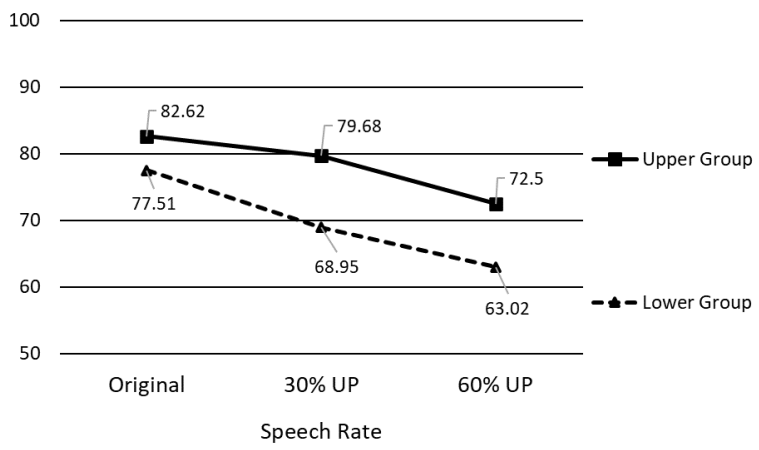
Mean Reproduction Rates of the Upper and Lower Groups in the Dictation Tasks (N = 27)

	n	Original	30% UP	60% UP
Upper Group	21	82.62% [9.02]	79.68% [13.67]	72.50% [11.99]
Lower Group	19	77.51% [14.09]	68.95% [13.93]	63.02% [12.22]
Total	40	80.20% [11.83]	74.58% [14.66]	68.00% [12.87]

Note. Numbers in the brackets are standard deviations.

Figure 3

Mean Reproduction Rates of the Upper and Lower Groups in the Written Reproduction Tasks (N = 27)



Although the influence of speech rate seemed to be stronger for the lower proficiency group, the two-way ANOVA revealed no significant interaction between the two variables (Table 3.5), indicating that the speech rate affected both proficiency groups equally. Conversely, the main effects of proficiency level and speech rate were significant (p = .003, .000 respectively). Multiple comparisons with Bonferroni correction for speech rates (Table 3.6) revealed a significant difference between the original rate and the fastest rate (p = .000). Moreover, while the mean difference between 30% UP and 60% UP was close, it did not reach the significant level (p = .076). Thus, the assumption that the perception of English speech will become challenging for learners when the rate goes beyond 200 wpm is debated.

Table 3.5

Summary Table for Two-Way Analysis of Variance of the Effects of Proficiency Levels and Speech Rate Conditions on Mean Reproduction Rate

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2
(A) Proficiency	1	710.92	710.92	10.07	.003	.209
(B) Speech Rate	2	3024.75	1515.37	11.56	.000	.233
(A) × (B)	2	174.01	87.01	0.67	.517	.017
Total	76	9942.42				

Table 3.6

Results of Multiple Comparisons Between the Three Speech Rates

(I) Speech Rate	(J) Speech Rate	(I-J)	*SE*	*p*	95% CI	
					Lower	Upper
Original	30% UP	5.75	2.53	.086	-0.58	12.08
	60% UP	12.31	2.32	.000	8.51	18.10
30% UP	60% UP	6.55	2.82	.076	-0.50	13.60

Nonetheless, the performance of the lower group at the original rate (77.51%) was better than that of the upper group at the fastest rate (72.50%). An independent samples t test

revealed that the gap was statistically significant, t (38) = 5.81, p = .000, with a large effect size, d = 1.84. This suggests that English speech at speeds greater than 200 wpm is difficult, even for upper-level learners, and therefore, the assumption is still worth examining.

3.3.2 Error Analyses on Difficult Words in the Faster Conditions

To reveal difficult words for learners to perceive at faster speech rates, the total number of participants who reproduced these words successfully in the two fast conditions (i.e., 30% UP and 60% UP) and their percentages were calculated for each word, and then the characteristics of the errors were analyzed. In the text below, words with a percentage of less than 30% are underlined (see Appendix 3B for all results).

Passage A:

Good morning, everyone. Thanks for coming to this employee meeting. Unfortunately, there was an accident last night in the restaurant's kitchen. One of the cooks burned his hand badly when a pot of hot soup was knocked over. I just want to remind you all to follow our safety rules at all times. We want you to work quickly, but safety is the most important thing.

Passage B:

Ted studies business at college. For him, the most interesting thing about the program is that he sometimes gets to work in real companies. He is learning a lot about how companies work. He thinks this will be a valuable experience for his future. He has to work very hard, though, and he does not have much time to relax at home anymore.

Passage C:

Many people find falling snow very beautiful. But animals in the ocean may find it annoying. Researchers have discovered that snowflakes hitting the ocean's surface create a noise. For animals under the surface, this sound can be very loud. The researchers do not yet know for sure whether the noise harms or disturbs the animals, but they know the animals can hear it.

(Grade 2 of the STEP Eiken listening test in the fall of 2011)

Compared to the other two passages, Passage B seems to have been more learner-friendly because the participants were able to reproduce most words, except *companies* and *valuable*. There are two possible reasons why this passage is rather easy. First, its vocabulary level was lower than that of the other two passages. As summarized in Table 3.1, 92% of the vocabulary in Passage B was categorized into the JACET 1000-word level, whereas Passages A and C contained less than 80% of the vocabulary at this level. Another reason may be that participants found the topic familiar. Passage B was a story about a college student interning at a company. Several university students volunteer to work in schools, so an internship is of interest to them. In contrast, Passage C, which was the most difficult passage, explains a phenomenon in nature, that is, the way snow makes noise for ocean animals. The participants were all English majors; therefore, they might have found this topic difficult. Passage A, which looks like an announcement made by the restaurant manager in an employee meeting, was probably less familiar to the participants because it is assumed that not many students had experience working in a restaurant.

The error analyses on Passages A and C suggested that most errors can be categorized into four types. First, words containing certain vowels were perceived poorly. The perception of English vowels can be problematic for Japanese learners of English. This is because Japanese has only five vowels while English has fourteen stressed vowels (Celce-Murcia, Brinton, & Goodwin, 1996) or more than twenty stressed vowels (Katayama, Nagase, & Joto, 1996; Takebayashi & Saito, 2008). In this study, the word *pot* (/pɑt/) in Passage A, which only 29% of the participants were able to perceive correctly, was recognized as various words such as *part* (/pɑːrt/), *put* (/pʊt/), and *poor* (/pʊər/). The contrast between the low central vowel (/ɑ/) and its long version (/ɑː/) may become more perplexing as the speech rate increases, because compression of an aural passage shortens its total duration, and thus the vowel lengths, causing learners to fail to distinguish those vowels.

Conversely, distinguishing it from the high back vowel (/ʊ/) should be easier since, according to Takebayashi and Saito (2008), the place and manner of articulation of these two vowels are quite different from each other (i.e., /ɑ/ = low central unrounded vowel; /ʊ/ = high back rounded vowel). Inherently, the high back vowel, which does not exist in the Japanese phonemic system, may be indistinguishable for Japanese learners. The word *sure* (/ʃuər/) in Passage C, which only 27% of the participants were able to perceive correctly, was mistaken for several other vowels such as *share* (/ʃeər/), *show* (/ʃou/), and *shark* (/ʃɑ:rk/). Based on these observations and considerations, it can be speculated that learners must undergo extensive training to successfully perceive English vowels at extremely fast speech rates.

Second, the perception of some consonants was challenging for participants. The most prominent example is the distinction between the two English liquids, /l/ and /r/. These consonants are not distinguished in the Japanese language, and thus, this contrast is often difficult for Japanese native learners to perceive (e.g., Goto, 1971; Mochizuki, 1981; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997). For example, in this study, several participants spelled the words *snowflakes* as *snowfrakes* and *create* as *clear* or *cleared* in Passage C. A common characteristic of these errors is that consonants are located in consonant clusters (i.e., /fl/ and /kr/) at the beginning of words or syllables. Kato (2009), who reviewed research on the perception of English liquids by Japanese learners of English, found that this perception is the most challenging when they occur in this phonetic environment.

Another difficult consonant is the voiceless alveolar stop /t/. Voiceless stops such as /p/, /t/, and /k/ are usually pronounced with strong exhalation or aspiration when accompanied by a stressed vowel (Celce-Murcia, Brinton, & Goodwin, 1996). According to Raphael (2005), aspiration gives listeners a significant clue to perceiving voiceless stops, which, in turn, means that L2 listeners may have greater difficulty perceiving these consonants when not aspirated. In fact, in this study, several participants spelled *Unfortunately* (/ʌnfɔ:rtʃənətli/)

as *Unfortunally* or *Unfortunery* (/restərənts/) and *restaurant's* as *restlan's* in Passage A. In both cases, the consonant /t/ in the middle of the words dropped out, possibly due to the lack of aspiration, and thus it was imperceptible. For the same reason, *knocked* (/nɑkt/) in Passage A, which contains the voiceless velar stop /k/, was misheard by several participants as *not* (/nɑt/) where the stop consonant is gone.

The unaspirated /t/ in words such as *hitting* in Passage C can also be troublesome for learners. In North American English, /t/ preceded by a vowel, followed by an unstressed syllable (e.g., *da̲t̲a* and *cit̲y*), often changes to a sound called a flap or tap that is represented by /ɾ/ (Celce-Murcia, Brinton, & Goodwin, 1996). Flaps are usually voiced and sound like /d/ (Shizuka, 2019); thus, participants tended to spell *hitting* as *heading* or *hedding*. These errors also suggest that Japanese learners may confuse the short vowel /ɪ/ with /ɛ/ because, as Katayama, Nagase, and Joto (1996) mention, the pronunciation of the former vowel is close to that of Japanese [e].

Third, reproduction rates of vocabulary with inflectional morphemes (e.g., possessional and plural) were likely to be low. Table 3.7 shows the common errors observed in this study. These errors can be classified into two types. First, the participants were able to perceive the sounds correctly, but wrote down ungrammatical words. For example, *restaurant's* and *restaurants* in Passage A are identical in terms of pronunciation but differ in terms of grammatical functions. The same contrast was observed in the pairs of *companies* and *company's* in Passage B and *ocean's* and *oceans or oceans'* in Passage C. Errors of this type are associated with a lack of attention or grammatical knowledge. This may indicate that faster speech leads to a heavier cognitive load on the learners' working memory. The second error is that the participants actually failed to perceive the sounds of the morphemes and, as a result, spelled the words incorrectly. Examples of this type are *restaurant* (Passage A), *company* (Passage B), and *ocean* (Passage C). Learners who make such errors need to

improve their perception skills to become more proficient listeners. Dictation is the best activity for this because it helps learners realize the kinds of words they have trouble listening to.

Table 3.7

Common Errors for Words With Inflectional Morphemes

	Words	%	Common Errors
Passage A	restaurant's	25	restaurant, restaurants
Passage B	companies	27	company, company's
Passage C	ocean's	8	ocean, oceans, oceans'

Note. Numbers in the % column are reproduction rates.

Finally, the perception of unstressed function words seemed to be challenging for the participants. Specifically, their performance on words such as *our* (Passage A) and *find it annoying*, *create a noise*, *or, they* (Passage C) was fairly poor as their reproduction rates fell between 0% (*they*) and 27% (*it, or*). Unlike content words, function words seldom receive prominence in sentences because they carry little meaning. Particularly, when syllables ending with a consonant are followed by unstressed vowels (e.g., *find it* and *create a*), there is likely to be a linking (Takebayashi & Saito, 2008), which can cause segmentation failure. To avoid this, learners must engage in activities where they are required to reproduce function words, such as dictation and shadowing.

3.4 Summary of the Findings: Chapter 3 (Study 1)

3.4.1 Answer to RQ1-1

RQ1-1: Does the reproduction rate of English speech decrease drastically when the speech rate exceeds 200 wpm? Is the influence of speech rate on perception more critical to lower-level learners?

The performance of both proficiency groups declined steadily and almost parallelly with the increase in speech rates. No significant interaction between the proficiency group and speech rate indicated that the influence of speech rate was equal for both proficiency groups. However, multiple comparisons between the three speech rates revealed a significant decline between the original and fastest rates, but the gaps between the others were not significantly different. From these results, the author's assumption that learners' performance in speech perception deteriorates sharply when its rate exceeds 200 wpm is still open to question. Nonetheless, the performance of the upper group at the fastest rate was significantly poorer than that of the lower group at the original rate. Thus, it can be plausibly concluded that English speech faster than 200 wpm is demanding even for upper-level learners.

3.4.2 Answer to RQ1-2

RQ1-2: What kinds of words are difficult to perceive when the speech rate exceeds 200 wpm?

Analyses of the participants' scripts revealed four types of common errors. The participants had difficulty perceiving (1) vowels such as the low central (/ɑ/) and high back (/ʊ/) vowels; (2) consonants such as the contrast between the two liquids (/l/-/r/) in consonant cluster positions and unaspirated voiceless stops (/t/ and /k/); (3) inflectional morphemes such as possessive and plural *-s*; and (4) unstressed function words. These results correspond to beliefs about the kinds of difficulties Japanese learners tend to experience in listening or pronunciation. When English speech is presented at a very fast rate (e.g., faster than 200 wpm), these difficulties are reinforced, and thus, learners will be unable to understand the passage accurately. Activities focused on speech perception ability, such as dictation and shadowing, can solve this problem, but which of these activities is more effective for this purpose is still not known.

3.4.3 Study Limitations

This study has two limitations. First, the proficiency gap between the upper and lower groups was not sufficiently large, even though the statistical analysis demonstrated a significant difference. This may have led to the finding that the interaction between proficiency level and speech rate was not significant. Hence, participants of various proficiency levels must be included. Second, the materials taken from the Grade 2 of a STEP Eiken test were somewhat easy for the participants because the mean reproduction rate of all participants remained at 68%, even with the fastest rate. In future research, more challenging material should be used for learners of this level.

Chapter 4

Study2: Short-Term Training Effect of Accelerated Speech Dictation and Shadowing on Speech Perception Ability

4.1 Study Goal

Study 1 revealed that perception of EFL speech tended to become more difficult as speech rate increased. In particular, it was found that when the rate exceeded 200 wpm, the impact of speech rate on perception became critical. As O'ki (2012) theorized, at this speed, EFL listeners are under a greater psychological load, and thus they often depend on their holistic-sound processing system rather than analytic-sound processing system. According to Kadota (2007), this happens to learners in shadowing. That is, shadowing imposes a greater psychological load than ordinary listening activities do, causing the holistic sound-processing system to be more dominant. In light of these considerations, it can be assumed that training by dictation using fast speech (accelerated speech dictation) may also have an identical effect on learners' speech perception ability with shadowing. Therefore, this study piloted the effects of accelerated speech dictation and shadowing on speech perception ability when they were performed for a short period.[1]

In this context, this study addressed the following RQs:

RQ2-1: Will the training using accelerated speech dictation and shadowing improve learners' speech perception ability?

RQ2-2: Will learners appreciate the effectiveness of accelerated speech dictation and shadowing?

[1] This study is based on the experiment reported by O'ki and Izumi (2015).

4.2 Method

4.2.1 Participants

The participants in this study comprised 26 sophomores and one junior (8 men and 19 women) majoring in English education at a private university. They were attending an elective English course taught by the author to prepare for the TOEFL ITP and practiced all sections of the test (i.e., listening comprehension, structure and expression, and reading comprehension) using a commercial textbook. Instructions for the listening comprehension section in each class comprised a comprehension exercise and an explanation by the author of the text using the script. The students had no opportunity to dictate or shadow as a class activity prior to the data collection. All participants were divided into one of the two groups after the pre-test: the accelerated speech dictation group or the shadowing group. None of the participants reported any hearing impairments.

4.2.2 Materials

Pre- and Post-Test. To compare the effects of accelerated speech dictation and shadowing on the improvement of speech perception ability, the participants performed a written reproduction task in the pre- and post-tests. Considering the level of materials used in the course, a passage for both tests was extracted from Grade Pre-1 of the STEP Eiken listening test administered in the spring of 2013 (see below). Only the first word of each sentence was printed on the test sheet (Appendix 4A) so that the participants could listen to the passage and fill out the rest. The original speed of this passage was approximately 150 wpm. The difficulty of this passage is summarized in Table 4.1.

It is predicted that two-thirds of the world's population will be living in cities by the year 2030. Many people worry this will lead to increased greenhouse-gas

emissions and greater environment damage. A recent study carried out in the U.K., however, suggests the reverse may be true. According to the study, well-planned cities can actually have lower CO_2 emissions per person than suburban or rural areas.

In cities, two of the biggest sources of CO_2 are emission from vehicles and domestic waste. However, environmentalists are now realizing that city planning can play an important role in reducing CO_2 emissions. In the U.S. city of Denver, for example, CO_2 emissions per person are almost twice those in New York City. This is because Denver is spread out and its residents rely on cars for transportation, while New York City is densely populated and has an efficient public transportation network.

(Grade Pre-1 of the STEP Eiken listening test in the spring of 2013)

Table 4.1

Difficulty of the Audio Passage Used in the Pre- and Post-Tests

Total Words	Duration (seconds)	JACET 1000 tokens	Words/sentence	FKGL	wpm
149	59 seconds	72.6%	18.5	11.5	150.5

Note. Compound words, such as "two-thirds" and "greenhouse-gas," were counted as two words. "JACET 1000" refers to the number of words listed in 1000-word level based on the JACET 8000 Word List (checked at: http://someya-net.com/wlc/index_J.html). FKGL = Flesch-Kincaid Grade Level. wpm = words per minute.

Training Materials. The same passage was used in the training session between the pre- and post-tests. However, for the accelerated speech dictation group, a faster version of the audio was prepared by compressing it using Audacity software (https://www.audacityteam.org/). The rate was accelerated to exceed 200 wpm, the speed at which most English learners have difficulty perceiving the speech, as found in Study 1. The shadowing group was instructed to practice at the original speed (150 wpm); however, if they found shadowing too difficult or the passage too fast, a slower version of the passage (110 wpm) was also prepared by stretching it using the software.

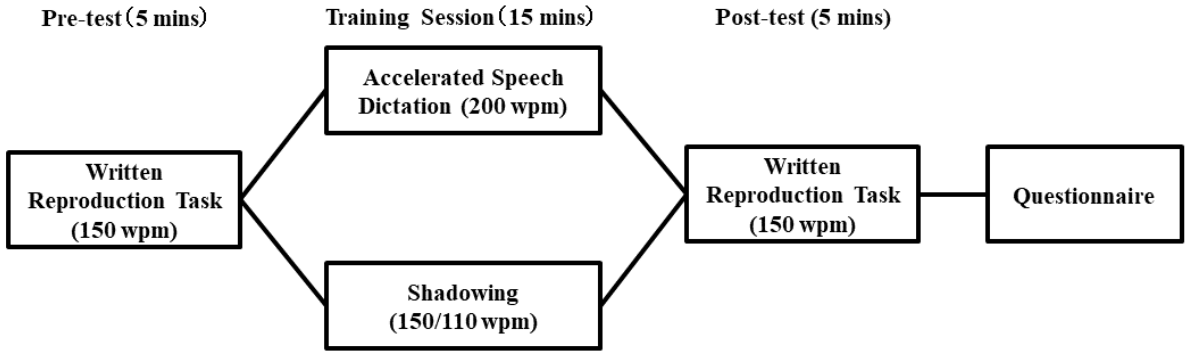
4.2.3 Procedure

The study was conducted as part of an in-class listening activity in a computer lab where regular classes were held. Each student was assigned a computer with a headset and had access to the shared folder that allowed them to download class materials prepared by the author. At the beginning of the class, a booklet labeled either A or B was randomly distributed to the participants to divide them into two groups: the accelerated speech dictation group (ASD group; $n = 14$) and the shadowing group (SH group; $n = 13$).

The study was conducted in four steps, as outlined in Figure 4.1. First, in the pre-test, the participants were allowed five minutes to listen to the passage described in the previous section and write as many words as possible on the test sheet. Both groups were allowed to stop and replay the audio at any time. Second, during the training session, the participants engaged in the activity assigned previously for 15 minutes. The ASD group was told to listen to only the faster speech (200 wpm), while the SH group was given two speech rate options (150 or 110 wpm) and allowed to use these alternately, if necessary. As the same text was used in the subsequent post-test, its script was not provided until the entire study was completed. In contrast to the pre-test, the ASD group was allowed to pause and replay the audio when it was being played, while the SH group was instructed not to do so. Third, after the completion of the training session, the post-test was conducted in the same manner as the pre-test using the same passage. Finally, the participants responded to a questionnaire asking them to rate the effectiveness of the activity on a five-point scale from 1 (not effective) to 5 (very effective) and to note their impression of the activity in a freely written format. After collecting the test sheets, the author provided participants with the script so that they could check the accuracy of their perception.

Figure 4.1

Experimental Procedure of Study 2



4.2.4 Scoring

The scoring method used in Study 1 was adopted in this study as well; the performance of participants was measured through their reproduction rate, that is, the percentage of words that were spelled correctly. There were 149 words in total; however, eight words at the beginning of the sentences were excluded from the calculation. Therefore, 141 words were used for scoring.

4.2.5 Analyses

Three types of analyses were performed in this study. First, to compare the speech perception ability of the two groups before training, an independent samples *t* test was conducted to evaluate the groups' performance in the pre-test. Second, to examine whether the two types of training improved participants' speech perception ability, a mixed design of a two-way ANOVA in a 2 (pre-test, post-test) × 2 (ASD group, SH group) format was carried out. Third, an independent samples *t* test was performed on the means obtained from the responses of the two groups; this was done to evaluate how difficult the training was.

4.3 Results and Discussion

4.3.1 Results of the Pre- and Post-Tests

Table 4.2 and Figure 4.2 show the reproduction rates of the two groups in the pre- and post-tests. In the pre-test, the mean reproduction rate of the ASD group was slightly higher than that of the SH group. The independent samples *t* test showed no significant difference between the mean reproduction rates of the two groups, $t(25) = 0.56$, $p = .579$, $d = 0.22$. Hence, it was considered that the speech perception abilities of the two groups were equal before training. However, in the post-test, the gap between the two groups became marginally larger (i.e., the difference between the two groups in the pre-test was 2.37%, and 5.76% in the post-test). The difference in reproduction rates between the pre- and post-tests of the ASD group was 10.00%, whereas that of the SH group was only 6.61%, suggesting a better training effect from accelerated speech dictation.

Table 4.2

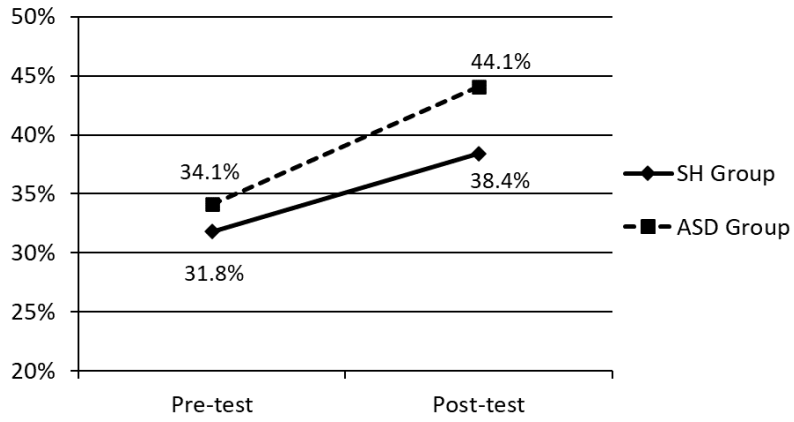
Mean Reproduction Rates of the Two Groups in the Pre- and Post-Tests (N = 27)

	n	Pre-test		Post-test	
		M	*SD*	*M*	*SD*
ASD Group	14	34.14% (48.1 words)	12.43	44.14% (62.2 words)	16.15
SH Group	13	31.77% (44.8 words)	9.12	38.38% (54.1 words)	10.67
Total	27	33.00% (46.5 words)	10.82	41.37% (58.3 words)	13.84

Note. ASD = accelerated speech dictation; SH = shadowing. The total number of words used was 141.

Figure 4.2

Mean Reproduction Rates of the Two Groups in the Pre- and Post-Tests (N = 27)



Note. ASD = accelerated speech dictation; SH = shadowing.

Shapiro-Wilk tests for normality were performed; they revealed that score distributions at all levels were normal. The two-way ANOVA was therefore analyzed as previously planned (see Table 4.3). Although the ASD group showed better improvement, the analysis found no significant interaction between the two variables, F (1, 53) = 1.26, p = .272, η_p^2 = .048, or the main effect of the training group, F (1, 53) = 0.80, p = .381, η_p^2 = .031. However, the main effect of pre- and post-tests was significant, F (1, 53) = 30.43, p = .000, η_p^2 = .549, indicating that the reproduction rate improved significantly in the post-test, regardless of the type of training.

Table 4.3

Summary Table for Two-Way Analysis of Variance of the Effects of Training Type and Pre- and Post-Test Conditions on Mean Reproduction Rates

Source	*df*	*SS*	*MS*	*F*	*p*	η_p^2
(A) Training Type	1	222.87	222.87	0.80	.381	.031
(B) Pre-, Post-Tests	1	930.46	930.46	30.43	.000	.549
(A) × (B)	1	38.61	38.61	1.26	.272	.048
Total	53	8972.15				

The two types of training exerted a similar effect for two possible reasons. First, the training duration (15 minutes) was too short to reveal a difference. In particular, the shadowing group may have needed more time to get used to the unusual task, wherein they had to listen and speak simultaneously. Tamai's (1997) study suggested that shadowing was effective even after a short time, but the results were obtained after a training period of five days. If the participants had more time to practice, they might have performed better. Therefore, in future studies, training that allows participants more time to practice tasks will be necessary.

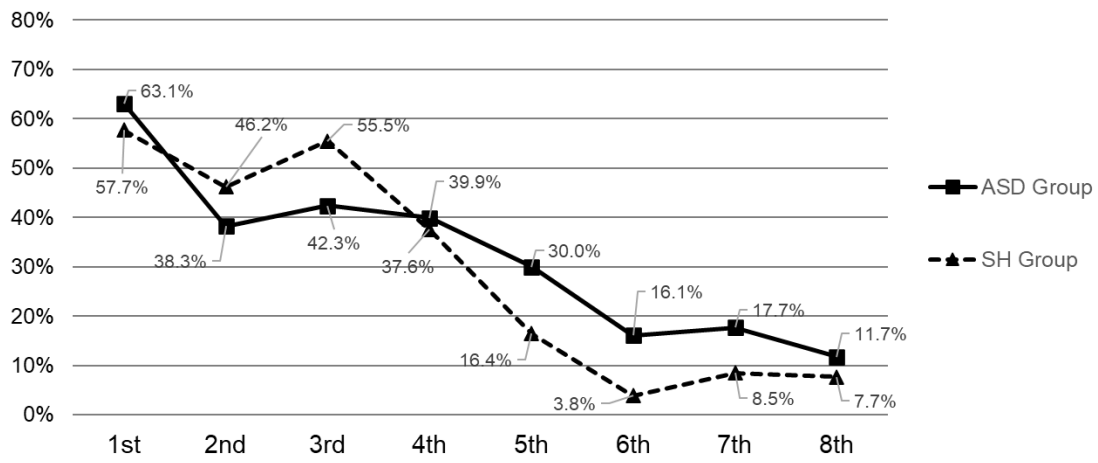
Second, the testing time in the pre- and post-tests (i.e., 5 minutes) was not sufficient to complete the task. Figure 4.3 shows the mean reproduction rate in the post-test for each sentence. As can be seen from the graph, the score decreased in later sentences. Inspection of the test sheets submitted by participants revealed that many participants wrote fewer words in later sentences, which implies that the test time had run out before the participants could work on those sentences. Interestingly, the decline was more prominent in the SH group. That is, their reproduction rate was as high as 55.5% for the fifth sentence but dropped to below 10% for the sixth to eighth sentences. This result appeared unique when considering that during the training session, the participants could not pause the audio in the midst of each shadowing to listen to the entire passage several times. The sharp decline probably occurred because shadowing was such a cognitively demanding activity that the learners failed to remain attentive in the later sentences. Furthermore, it is likely that the text, obtained from Grade Pre-1 of the STEP Eiken test, was too difficult for the participants.

The study's limitation—the absence of a control group for comparison—implied that the improvement could have been a practice effect caused by using the same passage repeatedly in the pre-test, training session, and post-test. In particular, the ASD group engaged in the reproduction task at all the stages; therefore, they may have become

accustomed to the task itself. This could be why the ASD group demonstrated a slightly better improvement than the SH group, although the interaction was not significant. To be sure that both training sessions contributed to the improvement, it would be necessary to compare them with, for example, a listening comprehension group.

Figure 4.3

Mean Reproduction Rate in Each Sentence in the Post-Test



Note. ASD = accelerated speech dictation; SH = shadowing.

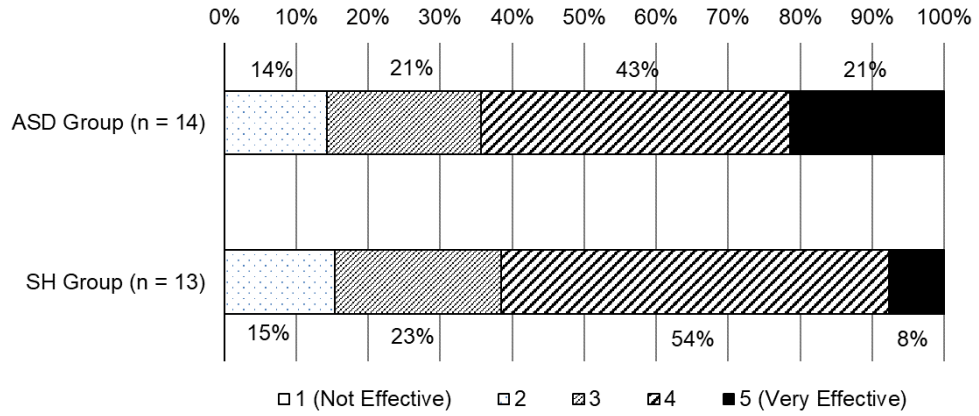
4.3.2 Responses to the Questionnaire

Ratings of the Effectiveness of Training. Figure 4.4 shows the participants' ratings of the effectiveness of their training. In both training types, more than 60% of the participants answered either 4 (moderately effective) or 5 (very effective), while none answered 1 (not effective). The mean score of the ASD group was moderately high (M = 3.71, SD = 0.99) and close to that of the SH group (M = 3.54, SD = 0.88). Because the ratings of the SH group were not normally distributed, a Mann-Whitney U non-parametric test was performed instead of a t test. The test revealed no significant difference between the mean ranks (ASD group = 14.71, SH group = 13.23), U = 81, z = -0.52, p = .603, with a very small effect (r =.10), indicating that both groups equally appreciated the effectiveness of their training.

However, in each group, approximately 15% of the participants questioned the effectiveness of their training. This was possibly because, as mentioned in the previous section, the training was so difficult that a longer duration was needed for it to take effect and for some participants to feel its effectiveness. The following section reports that in the freely written question, numerous participants responded that the training was very challenging.

Figure 4.4

Ratings by the Participants on the Effectiveness of Speech Perception Training



Note. ASD = accelerated speech dictation; SH = shadowing.

Responses to the Freely Written Question. All responses to the last question, which asked participants for voluntary feedback on the training in a freely written format, are listed in Table 4.4 (the ASD group) and Table 4.5 (the SH group). In both tables, descriptions referring to the difficulty of training are underlined by the author.

For the ASD group (see Table 4.4), although many participants admitted that the training was challenging primarily because of the passage speed (e.g., participants D, E, J, M, and N), they could overcome this difficulty because they were allowed to hear the passage at the original rate during the pre-training phase. Furthermore, participants L and N mentioned that repetition of the training enabled them to process the fast speech. The instructional tips in

these participants' feedback were to encourage learners to continue working on the same material and, if the material was too fast or beyond their listening ability, to present it at the original or slower rate. Yet, this might reduce the effect of accelerated speech dictation to direct learners' attention to sound information.

Interestingly, participants A and J acknowledged that the fast speech rate promoted the processing of linguistic information, such as syntactic structures and important vocabulary. This may support the assumption of this study mentioned in section 4.1. That is, like shadowing, accelerated speech dictation induces learners to exploit the holistic sound-processing system in which they have to perceive aural input efficiently by focusing on larger syntactic structures and important words. However, the responses of participants C and E suggested that they tried to separate the input into smaller units and perceive it in detail. Because accelerated speech dictation requires transcription of every single word heard, learners must keep activating their analytic sound-processing system. This means that in accelerated speech dictation, learners must operate both types of sound-processing systems (i.e., holistic and analytic) simultaneously; consequently, it is a cognitively demanding activity that requires adequate instruction and careful selection of material.

Table 4.4

Feedback of the ASD Group on the Training (n = 14)

A)	*Although the speed was very fast, it helped me figure out the sentence structure and understand vocabulary that I could not hear in the pre-test.* (4)
B)	*I was able to write more words in the training than in the pre-test because I remembered some of the content.* (3)
C)	*I had difficulty hearing details of the passage like "the."* (4)
D)	*Because I listened to the sentences at a slow speed in the pre-test, I did not find it difficult to listen to the faster speech. However, the words that I could not hear at the original speed felt much harder at the fast speed.* (5)

E) *All the words sounded connected, so it was hard to understand them.* (5)

F) *I felt it was easier to listen to the faster speech (probably because I heard it once).* (2)

G) *We were allowed to stop the audio, so I felt little difference between the speeds in the pre-test and in the training.* (3)

H) *I got tired because dictation was an exhausting activity. However, I think I would learn from it if I continued it.* (5)

I) *I heard the sentences once, so I was able to understand the meaning even when the speed increased. It was good that I got to listen to the whole text.* (4)

J) *It was hard to hear the fast speech, but it enabled me to focus on the important words.* (4)

K) *Dictation made me realize how much I could not understand. I would like to make use of this opportunity for my future study.* (4)

L) *The more I listened to it, the more I could hear it.* (4)

M) *I was able to grasp some sentences because I had heard them once, but it felt very difficult because it was quite fast.* (2)

N) *It was difficult. I gradually got used to the speed of speech as I listened to it several times.* (3)

Note. Descriptions referring to training difficulty are underlined and the ratings of each participant shown in brackets. Original responses written in Japanese are shown in Appendix 4B.

As the underlines in Table 4.5 indicate, several participants in the SH group also experienced difficulty completing the training. For example, participants S and Y failed to follow the aural input up to the end of their training. This difficulty can be attributed to the passage speed (participants P, Q, and U), unfamiliar words (participant Q), and the length of the sentences (participant Y). Some educational implications can be drawn from these responses to alleviate the difficulty of shadowing. First, teachers should allow their learners to adjust the passage speed to a comfortable (but not too easy) level or to choose materials at an appropriate level. Second, repetition of the same material can promote perception, but its effect is probably limited; Shiki, Mori, Kadota, and Yoshida (2010) and O'ki (2014) reported

that the reproduction rate in shadowing leveled off within a few or several repetitions. In that case, the presentation of written texts may help improve the reproduction rate, but teachers should not rush doing so because learners will depend on it before they make their best efforts to perceive the speech.

Table 4.5

Feedback of the SH Group on the Training (n = 13)

O) *Shadowing requires you to listen to the English sentences and repeat them. Therefore, I think it will improve your listening skills.* (5)

P) *The original speed of the passage was so fast that sometimes I could not keep up with it.* (4)

Q) *I could not keep up with the original speed, so I practiced only with the slow speed. Even the slower passage sounded a little fast to me, but I was able to understand it when I really focused on it. However, I got stuck when there were unfamiliar words.* (4)

R) *I was able to understand more words as I worked on the dictation test.* (4)

S) *I found it easier to shadow the passage when I practiced with the original speed first and then with the slow speed.* (4)

T) *I practiced with both speeds alternately, but I could not reach the level at which I could shadow it successfully.* (4)

U) *Even the slower speech was very difficult for me to shadow, but I managed to do it.* (2)

V) *I was able to remember the sentences as I listened to them repeatedly. Furthermore, I could predict what sentence would come next.* (4)

W) *There were some words that I could not really hear.* (3)

X) *Due to the difference in speed, I was able to understand the original speech better after listening to the slower one.* (3)

Y) *The sentences were too long for me to catch up with them.* (2)

Z) No response. (4)

AA) *Dictation of passages with unfamiliar vocabulary is more difficult than those with only familiar vocabulary.* (3)

Note. Descriptions referring to training difficulty are underlined and the ratings of each participant are shown in brackets. The original responses, written in Japanese, are presented in Appendix 4C.

4.4 Summary of the Findings: Chapter 4 (Study 2)

4.4.1 Answer to RQ2-1

RQ2-1: Will the training using accelerated speech dictation and shadowing improve learners' speech perception ability?

The pre- and post-tests revealed that both the ASD and SH groups equally improved their reproduction rates in the post-test. There are two possible explanations for this finding. First, the training truly contributed to the improvement of participants' speech perception ability. Second, the improvement was merely a practice effect derived from repetition of the same material. In this sense, a control group that received no training or engaged in listening activities focused on comprehension skills was necessary for comparison. Furthermore, more time may be needed for each training session to demonstrate a larger effect. In particular, shadowing takes longer to get accustomed to; thus, participants needed to be trained for at least five times, as suggested by Tamai's (1997) study.

4.4.2 Answer to RQ2-2

RQ2-2: Will learners appreciate the effectiveness of accelerated speech dictation and shadowing?

Participants' ratings of the training's effectiveness demonstrated that regardless of the training type, a majority of them considered the training effective. However, a few participants in each group were reluctant to admit its effect, probably because the training session was too short despite the task difficulty. Their feedback on the freely written question provided some educational tips for both types of training: learners engaged in either training may benefit from (1) working on the same material repeatedly, (2) listening to the passage at a slower rate, and (3) looking at the written script to check unfamiliar vocabulary. However, (1) may not be as effective in shadowing as in dictation; previous studies on shadowing have

revealed that reproduction rates nearly stopped improving after learners had some practice. Furthermore, it is strongly advised to employ (2) and (3) only after learners have made their best efforts to understand the aural input at the original speed.

4.4.3 Study Limitations

This study has four limitations. First, the number of participants in each training type was small; future research needs to recruit more participants. To clarify the effects of speech perception training, a control group that receives no training or engages in listening comprehension activities is required for comparison.

Second, the training duration of 15 minutes was too short. According to previous research, shadowing requires training to take place approximately five times for it to take effect; therefore, it is imperative to conduct a semi-longitudinal study. During this period, it will be useful for learners to keep a journal because it provides further information about their development.

Third, some participants found the aural passage selected from Grade Pre-1 of the STEP Eiken test difficult. To accommodate a variety of learners' levels, teachers should allow them to choose their material or to adjust the speech rate to a comfortable but reasonably demanding level.

Fourth, learners' speech perception abilities should be measured using methods other than a written reproduction task; an integrated-skills task requires both listening and writing. Therefore, the task performance does not reflect only participants' speech perception ability, but also their knowledge of spelling. This issue is further discussed in the following chapter on developing a non-integrative task.

Chapter 5

Study 3: Development of a New Task for Speech Perception Ability (Part 1)

5.1 Study Goal

The previous two studies aimed to measure the participants' speech perception ability through a written reproduction task. However, because performance in this task is affected by learner's spelling knowledge, an alternative approach may be necessary to assess learner's speech perception ability more accurately. In the following two sections, types of speech perception tasks used in the literature are reviewed while referring to their advantages and disadvantages (5.1.1); and then a new type of speech perception task is suggested for further research (5.1.2).

5.1.1 Types of Speech Perception Tasks

There is a bulk of first language (L1) and L2 research that has focused on speech perception ability in English. Some of these studies are listed in Table 5.1. As shown, the types of speech perception tasks adopted in these studies can be categorized as: (1) discrimination/identification tasks, (2) oral reproduction tasks (repetition tasks), and (3) written reproduction tasks (dictation tasks). Each task has both strengths and weaknesses, which are mostly associated with characteristics of audio stimuli, means of output, and others (e.g., task difficulty and preparation costs). These tasks are discussed next.

Table 5.1
Types of Speech Perception Tasks Used in L1 and L2 Research

Task Type	Examples of Studies	L1
Discrimination/	Tsukada et al. (2005)	Korean
Identification task	Baker et al. (2008)	Korean
	Cebrian & Carlet (2014)	Spanish/Catalan

Oral reproduction task (Repetition task)	Cummins et al. (1984)	Japanese
	Peelle & Wingfield (2005)	English
	Shi & Farooq (2012)	Various
Written reproduction task (Dictation task)	Dupoux & Green (1997)	English
	Pallier et al. (1998)	Various
	Sebastian-Gallés et al. (2000)	Spanish
	Habibi, Nemati, & Habibi (2012)	Iranian
	Siegel & Siegel (2015)	Japanese

Note. L1 refers to the first language(s) of the participants.

Discrimination/Identification Tasks. In these types of tasks, participants are presented with two or more similar sounds (usually in word forms), or minimal pairs, and are asked whether they are identical or to match them with words that have the same phonemes. For example, Baker et al. (2008) conducted what the researchers call a cross-language perceptual identification task, where Korean children and adults listened to monosyllabic English words and reported Korean vowels that sounded closest to those in the stimulus words. This type of identification task is also called an AX task, where participants listen to "A" as a stimulus and tell what the unknown sound "X" is[1].

The strong points of this task are as follows: first, audio stimuli are usually short (i.e., phonemes or words); thus, it is assumed that the load on participants' memory is small; second, the output does not necessarily require L2 production, meaning the task can focus solely on the receptive ability of the participants; third, phonological knowledge needs to be defined when designing the task, so that the objective of the research is clear. This makes the testing of hypotheses easy.

However, this task also has certain weak points. First, it is not suitable for measuring perceptual knowledge or ability at a level higher than phonemes. In a nutshell, being able to distinguish different sounds at word level does not always mean one will be able to do the same

[1] If there are two stimuli, it is called an ABX task.

at sentence level. Therefore, this task can be used for limited purposes. Second, someone with expertise must design these tasks so that the hypotheses can be tested. For example, Tsukada et al. (2005) selected several pairs of English vowels that seemed to be either difficult or easy for Korean learners of English in order to examine whether the length of residence in America affects the discrimination of these English vowels. To design this kind of experiment, the researcher must be familiar with the phonology of both languages and have the ability to make meaningful vowel pairs for the research.

Oral Reproduction Tasks (Repetition Tasks). In this task, participants are presented with audio material (usually sentences) and are asked to repeat it as accurately as possible[2]. For example, Shi and Farooq (2012) examined the effects of speech rate and noise on speech perception of native and non-native English speakers.

The strengths of the oral reproduction task are threefold. First, contrary to discrimination/identification tasks, audio stimuli in this task are usually long (sentences or passages); hence, perception ability at a level higher than phonemes can be measured. Second, because the output is in the form of spoken language, and we can speak faster than we can write, oral reproduction tasks can be implemented in less time when compared to written reproduction tasks explained in the next section. Third, thanks to modern technology, it is becoming easier to obtain audio materials even in an EFL environment; thus, they can be prepared with less effort than before.

Despite these strengths, the oral reproduction task also has several limitations. First, since participants have to rehearse the input in mind, the task may impose a heavy load on

[2] Repetition tasks often go by names with the word "recall," but this term can be misleading because recall tasks (whether oral or written) usually refer to an experiment method to measure listening comprehension where participants reproduce the "content" of a listening material as much as possible.

memory. Second, because it is an integrative task (i.e., listening and speaking), participants must have a good command over speaking the target language. Pronunciation, in particular, can have a huge impact on the evaluation of oral performance; hence, rater reliability must be calculated. Third, it is impossible to record the oral performance of many participants simultaneously unless there are enough recording devices. Moreover, a quiet environment is essential; otherwise, peripheral noise will distract participants from listening and degrade the quality of recordings.

Written reproduction tasks (Dictation tasks). At present, the written reproduction task is utilized in both L1 and L2 research aimed at speech perception ability. In an L1 setting, Dupoux and Green (1997) conducted three experiments that focused on the perception of time-compressed speech by American university students. In an L2 setting, Pallier et al. (1998) carried out multiple experiments with participants from different L1 backgrounds (Spanish, Catalan, French, and English) and tried to reveal the influence of L1 and habituation (i.e., pre-task training session) on the perception of time-compressed speech.

Written reproduction tasks are advantageous in four respects. First, like oral reproduction tasks, written reproduction tasks usually use sentences (or passages) as input; hence, they can measure speech perception ability at sentence level. Second, output in written language format benefits both EFL learners and raters. EFL learners, who are often handicapped by opportunities to speak English, may feel more comfortable writing what they have heard than speaking. For raters, scoring is easier when compared to evaluating oral tasks because evaluation of accented speech can be unstable and tedious, raising questions on rater reliability. Third, unlike oral reproduction tasks, written reproduction tasks do not require recording devices, so they are low cost. Fourth, the difficulty level of written reproduction tasks can easily be adjusted by changing the number of words to be filled. For example, in the experiments

conducted by Cai (2012) and Afsharrad and Benis (2014), participants were asked to fill in each blank with one word.

Written reproduction tasks have similar limitations as oral reproduction tasks. First, participants have to endure a heavy load on memory while rehearsing the input until they reproduce it. The load could be heavier than oral reproduction tasks because writing something takes longer than speaking. What happens is that participants forget what they hear, which impacts their performance. Second, the evaluation of written reproduction tasks is inevitably affected by how to deal with spelling errors. Buck (2001) recommends deducting points depending on the weight of each error. However, this method has a critical problem; as Brown and Abeywickrama (2010) list, there are several types of errors, and decisions on how many points should be deducted could have a large impact on the outcome. Moreover, the decision has to be made again when using different passages, which is time consuming. To maintain rater reliability and practicality, scoring criteria need to be as clear and consistent as possible. For this reason, the present study gives points to only those syllables that are spelled correctly in the written reproduction tasks.

5.1.2 A New Task for Measuring Speech Perception Ability: Word Count Task

As we have seen, each task has its own strengths and weaknesses, and there is no perfect task for measuring speech perception ability. Owing to its popularity, the written reproduction task may work best for Japanese learners. However, integrative tasks may not be able to measure speech perception ability precisely; therefore, a new discrete-point task intended for measuring speech perception ability is required.

Based on the research background, the author developed an original task called the *word count task*. In this task, instead of writing down the words, participants count and report the number of words in blanks. One may wonder if this simple task reflects the construct of L2

listening ability. To answer this question, it is useful to refer back to the list of micro skills used in conversational listening proposed by Richards (1983) shown in Table 2.2. The idea of counting words is derived from No. 8 (ability to distinguish word boundaries), but this ability is linked with other perception skills. In other words, to determine the boundaries between words, listeners need to be able to retain chunks (No. 1); recognize unique sounds and prosodic features of L2 (No. 2, 3, 4, 5, and 6); and know when a word changes its sound form (No. 7).

The word count task has several merits. First, since it does not require L2 production, it can probably measure auditory skills more precisely than integrated-skills tasks. Second, because participants are exempted from L2 production, it may be less challenging and stressful for them than integrative tasks. In other words, it is beginner friendly. Third, as scoring is easy and objective, it is expected that both practicality and reliability can be preserved. Fourth, task difficulty can easily be adjusted by increasing or decreasing the number of words to be counted. Despite these possible merits, there is also a possibility that, like other multiple-choice tests, participants will luckily answer the correct number and obtain higher scores than their actual ability. Overestimation of participants' ability attributed to wild guessing lowers the score reliability.

5.1.3 Research Question

This study attempted to examine the validity and reliability of a prototype of word count task and reveal how the task should be revised. As explained, there is no established task to measure speech perception ability; therefore, the validity of a word count task was tested based on its relationship with a written reproduction task. As already stated, written reproduction tasks have shortcomings as a measure of speech perception ability, but due to their popularity in educational settings in Japan, using one is the best counterpart to the word count task.

Keeping in mind these considerations, this research attempted to investigate the validity and reliability of the word count task. To accomplish this goal, the present research tested the following RQ:

RQ3: Is a word count task valid and reliable as a measure of speech perception ability?

5.2 Method

5.2.1 Participants

The participants were 123 freshmen (59 men and 64 women) from a private university in Japan, who were majoring in education. To measure the English proficiency levels of new students, the university administers a TOEFL ITP shortly after their entrance every year. The participants took the test in the same month when this study was conducted. Their average score in the listening comprehension section was 43.34 (*SD* = 4.56) out of 68, with their individual scores ranging from 32 to 61. None of them reported hearing impairment.

5.2.2 Materials

Two sets of materials were used in this study: one for the word count task, or the CNT, and the other for the written reproduction task, or the RPD (see Appendix 5)[3]. The set used for the CNT consisted of 10 sentences extracted from an official Test of English for International Communication (TOEIC) training book published by Educational Testing Service (2012). These sentences were all taken from Part 1 (the picture-sentence matching task), with each having a blank to be filled (e.g., "A man is _____." for "A man is painting a gate.").

[3] CNT and RPD were chosen to use as abbreviations for the word count task and the written reproduction task respectively (they are contractions of "count" and "reproduction") because WCT (word count task) and WRT (written reproduction task) look alike and may be confusing for readers.

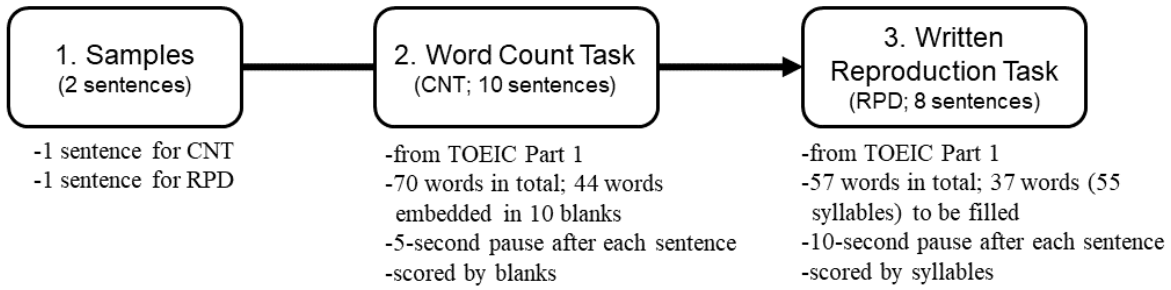
The participants were to count the number of words in the blanks, which ranged from three to five; the average number of syllables was 6.8. To prevent guessing of answers from visual information, pictures were not presented to the participants. The set used for the RPD included eight sentences selected from Part 1 of another test in the same TOEIC training book. As in the CNT, each sentence had blanks that participants filled by counting the number of words they had heard. The number of words ranged from four to five, and the average number of syllables was 6.9. Most sentences in the two sets were spoken with the standard American accent, while a few were spoken with a standard British accent.

5.2.3 Procedure

As illustrated in Figure 5.1, the study was conducted in three steps. First, two sample questions were provided to familiarize participants with the procedures of the two listening tasks. Second, the RPD followed, in which the participants listened to eight sentences twice. In order to allow participants time to jot down the words, a 10-second pause was given each time they listened to a sentence. Finally, the participants took the CNT, in which they listened to 10 sentences and counted the number of words replaced with blanks. They heard each sentence twice with a 5-second pause. Throughout the experiment, participants were permitted to take notes to reduce memory load.

Figure 5

Experimental Procedure of Study 3



5.2.4 Scoring

In the RPD, syllables that were spelled correctly were regarded as correct. The number of correct syllables was divided by the total number (55 syllables) to obtain the reproduction rate. In the CNT, one point was given for each correct word count (total points = 10).

5.2.5 Analyses

Two kinds of analyses were performed. First, to measure the *external aspect* of the CNT with the RPD[4], the Pearson's correlation coefficient between the average scores of the two tasks was calculated. Additionally, correlations of these tasks with the TOEFL listening test were analyzed to examine whether the CNT and the RPD really measure speech perception ability rather than general listening ability. Second, to compare the reliability of the CNT and the RPD, internal consistency using Cronbach's α was tested for both tasks.

5.3 Results and Discussion

5.3.1 Descriptive Statistics and Validity

Descriptive statistics of participants' scores in the three tests are shown in Table 5.2. The mean score of the CNT was 4.80 (out of 10); the standard deviation was 1.95. As shown in Table 5.3, there was a significant positive correlation between the scores of the CNT and the RPD (r = .46, p = .000), implying that the participants performed similarly in these perception tasks. However, given the fact that the same participants took these two tests, the coefficient was not high enough to conclude that both of these tasks have the same measuring ability.

[4] External aspect, an aspect of validity in Messick's framework, corresponds to the criterion-referenced validity in the classical test theory (Hirai, 2017).

Table 5.2

Descriptive Statistics of the Three Listening Tests (N = 123)

Test	*M*	*SD*	*SE*	*Median*	*Min*	*Max*
TOEFL	43.34	4.56	0.41	44	32	61
RPD (%)	50.04	12.83	1.16	49.09	12.73	89.09
CNT	4.80	1.95	0.18	5	0	10

Note. TOEFL = Listening section of TOEFL ITP; RPD = written reproduction task; CNT = word count task.

Table 5.3

Correlation Coefficients Among the Three Tests

	TOEFL	RPD	CNT
TOEFL	-		
RPD	.44*	-	
CNT	.22*	.46**	-

Notes. TOEFL = Listening section of TOEFL ITP; RPD = written reproduction task; CNT = word count task.

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

It is possible that the two tasks evaluate somewhat different aspects of listening ability because their correlations with the TOEFL listening score differ slightly. To be exact, the RPD score correlated more strongly with the TOEFL listening score (r = .44, p = .000) than the CNT score (r = .22, p = .015). These results may imply that the RPD, in which two skills are involved (i.e., listening and writing), does not assess speech perception ability alone.

This result seems to echo the findings of the KATE Research Promotion Committee (2018), which revealed that dictation scores of high school seniors had significant positive correlations with their performances in various tests such as listening comprehension, Japanese-to-English translation, and picture description. The committee also found that dictation scores were associated with both reading and listening scores obtained in the National Center Test. Based on these findings, it was speculated that dictation could be a convenient way to measure learners' general English ability. If so, it is not appropriate to evaluate the validity of the CNT

solely by its relation to learners' performance in a written reproduction task. For this reason, in the next study (Study 4), a listening strategy survey was also employed to reveal what cognitive processes are involved in the tasks.

5.3.2 Reliability

To measure the reliability of the two listening tasks, internal consistency was calculated for each task. The reliability coefficient by Cronbach's α was determined to be .45 for the CNT and .70 for the RPD. The coefficient of the CNT, as an objective test, is rather low, meaning the CNT is not always reliable. The difference in reliability between the CNT and the RPD can be attributed to the following two reasons.

First, in the CNT, the participants did not have to write what they heard but gave answers only in numbers; therefore, the chance of random guess scores might have been high. As sentences used in Part 1 of the TOEIC are usually short (i.e., less than 10 words), stimulus sentences used in this study were somewhat homogeneous in terms of the number of words to count (mostly four or five words). Furthermore, sentences in the TOEIC Part 1 are also limited in their variation of grammatical structure (most are in present progressive form), which may have allowed students to guess answers easily. To prevent this, it was necessary to increase the variety of items in the CNT regarding (1) the number of words to count and (2) their grammatical structures. This is reflected in the subsequent study.

Second, the limited number of items could also have lowered the reliability. The CNT consisted of only 10 sentences, each of which was scored binarily. Therefore, the score of each participant fell between 0 and 10, resulting in small variance in the total score of the participants. This is in contrast to the RPD, where there were 55 syllables to be scored, and thus the reproduction rate varied from participant to participant. An efficient way to increase the number of items in the CNT is to embed several blanks in a passage, rather than using only one blank

in a single short sentence. The reliability of the RPD can also be increased using passages, but this will cost participants too much time and energy for writing. To reach a compromise, partial written reproduction tasks (or partial dictation tasks), where learners reproduce some fragments of the passage rather than the whole sentence, can be an alternative to the RPD. Hence, Study 4 will explore the validity and reliability of a revised version of the CNT by comparing it to a revised version of the RPD.

5.4 Summary of the Findings: Chapter 5 (Study 3)

RQ3: Is a word count task valid and reliable as a measure of speech perception ability?

Even though the correlation between the RPD and the CNT was positively significant, the two tasks seemed to measure different abilities for the reason that the coefficient was not high. The result that the RPD demonstrated a higher correlation with the TOEFL listening implies that the RPD measures general listening ability rather than speech perception ability. For these reasons, validity of the CNT is still unclear. Investigation based on only the correlation analysis is not sufficient; thus, another approach aimed at learner's cognitive processes such as a strategy survey is also needed to evaluate the validity of CNT.

In terms of the reliability, the CNT exhibited insufficient coefficient due to limited variation of the items and the small number of them. Using passages with more blanks might resolve these problems and increase the reliability of the CNT. The partial written reproduction task using passages is also an alternative to the RPD for the higher reliability.

Taking these into consideration, the next chapter examines the validity and reliability of improved versions of the CNT and the RPD.

Chapter 6

Study 4: Development of a New Task for Speech Perception Ability (Part 2)

6.1 Study Goal

This study attempted to evaluate the validity and reliability of a revised version of the word count task (CNT-R) by comparing it to a revised version of the written reproduction task (RPD-R). These new tasks are different from their original versions in the following two respects: it has (1) more variations in the number of words to count or write and in the grammatical structures and (2) more blanks to be filled using passages rather than single short sentences. It was expected that these revisions would increase the score variance and reliability coefficient of the two tasks. As for validity, the results from Study 3 suggested that language skills used in the CNT may be different from those used in the RPD. To confirm this, not only the correlation analyses but also a strategy survey was conducted, and participants' responses were compared between the CNT-R and the RPD-R. The final goal is to decide which task is more suitable for measuring speech perception ability.

Based on these backgrounds, this study examined the following RQ:

RQ4: Is the revised version of the word count task (CNT-R) valid and reliable as a measure of speech perception ability?

6.2 Method

6.2.1 Participants

The participants were 76 university students (40 men and 36 women), enrolled in the same university as in Study 3. Among them, 34 were freshmen majoring in business, and 20

were sophomores and 22 juniors were majoring in education. Their average age was 19.23 (*SD* = 1.01), and none of them reported any hearing impairment. There were no TOEFL records to illustrate their listening level, but the same written reproduction task as in Study 3 (i.e., the RPD) was administered to compare their speech perception ability with that of Study 3. As a result, it was found that students in this study were able to reproduce 47.92% (*SD* = 12.80) of the target syllables, falling short of those in Study 3 by 2.12%. However, an independent *t* test proved that this gap was not statistically significant, $t(197) = 1.13$, $p = .259$, $d = 0.63$, indicating that the two participant groups possessed equal ability.

6.2.2 Materials

The material was prepared for three kinds of tasks. First, as stated earlier, the same written reproduction task as in Study 3 (the RPD) was administered to compare the results between the two studies. Second, an RPD-R was newly designed (see Appendix 6A). To accommodate the variety of participants' proficiency levels, a total of three passages were selected from Grade Pre-2, 2, and Pre-1 of the STEP Eiken tests held in the fall of 2014. The characteristics of these three passages are summarized in Table 6.1. As indicated by (2) in the table, the numbers of words to be filled in the tests were 11, 15, and 24, respectively (total= 50 words, 64 syllables). The blanks were diverse in terms of their grammatical structure (e.g., subject + verb, verb + object, noun phrase, to-infinitive verb phrase, and adverbial phrase), and most of them contained one or more function words. All passages were spoken with a standard American accent.

Table 6.1

Characteristics of the Three Passages Used in the RPD-R

STEP Eiken Grade	Pre-2	2	Pre-1
	(a)	(b)	(c)
(1) Total words	58	69	99
(2) Words/syllables in blanks	11/15	15/16	24/33
(3) Words per sentence	14.5	13.8	13.8
(4) Alphabets per word	4.5	4.1	4.9
(5) Flesch Reading Ease	73.9	82.4	58.4
(6) Flesch-Kincaid Grade Level	6.5	5.1	8.5
(7) Duration (sec.)	25	31	45
(8) Words per minute	139	133	132
(9) Gender of speaker	Male	Female	Male

Third, as explained above, a CNT-R was developed (Appendix 6B). A total of five passages were chosen from the same STEP Eiken tests, but the passages were different from those used in the RPD-R. As indicated by (2) in Table 6.2, the number of blanks for each passage was 8, 6, 6, 7, and 10, respectively (total = 37 blanks). As in the RPD-R, the blanks had a variety of grammatical structures, most of which contained a function word. All the passages were spoken with an American accent. Table 6.3 displays characteristics of the original and revised versions of the word count task and the written reproduction task.

Table 6.2

Characteristics of the Five Passages Used in the CNT-R

STEP Eiken Grade	Pre-2		2		Pre-1
	(d)	(e)	(f)	(g)	(h)
(1) Total words	58	61	68	67	88
(2) Words in blanks	20	15	18	17	24
(Number of blanks)	(8)	(6)	(6)	(7)	(10)
(3) Words per sentence	14.5	15.2	13.6	13.4	17.6
(4) Alphabets per word	4.0	3.7	4.4	4.6	4.6
(5) Flesch Reading Ease	71.0	79.0	69.8	74.5	64.9
(6) Flesch-Kincaid Grade Level	6.9	6.0	6.8	6.1	8.5

(7) Duration (sec.)	28	24	29	31	38
(8) Words per minute	124	152	140	129	139
(9) Gender of speaker	Female	Male	Male	Female	Male

Table 6.3

Characteristics of the CNT, CNT-R, RPD, and RPD-R

	Word Count Task		Written Reproduction Task	
	CNT	CNT-R	RPD	RPD-R
(1) Number of words in each blank	3 to 5[a]	1 to 5	4 or 5	1 to 3
(2) Grammatical structures	Mostly present progressive	Various	Mostly present progressive	Various
(3) Text type	Short sentence	Passage	Short sentence	Passage
(4) Full mark	10 blanks	37 blanks	55 syllables	64 syllables

Note. [a]One blank consists of three words, while the other blanks consist of four or five words.

In addition to these listening tests, a strategy survey (Appendix 6C) was also conducted to assess the *substantive aspect* of both measures[1]. It consisted of 16 multiple-choice items, aimed at revealing the cognitive processes that participants underwent during the RPD-R and the CNT-R (for the items, see Table 6.6 in the Results and Discussion section). The questionnaire, originally developed by Vandergrift (2003), was arranged by Sakai (2009) to compare learners' listening strategies used in a dictation test with those used in a free written recall test. Sakai's version has 26 items written in Japanese, 15 of which were extracted for this study. One item (12. While listening, I paid attention to detailed sounds like the plural *-s* and past tense *-ed*.) was originally created for the present study to examine how closely the participants paid attention to the input. The participants were asked to choose a response from 1 (*disagree*) to 5 (*agree*).

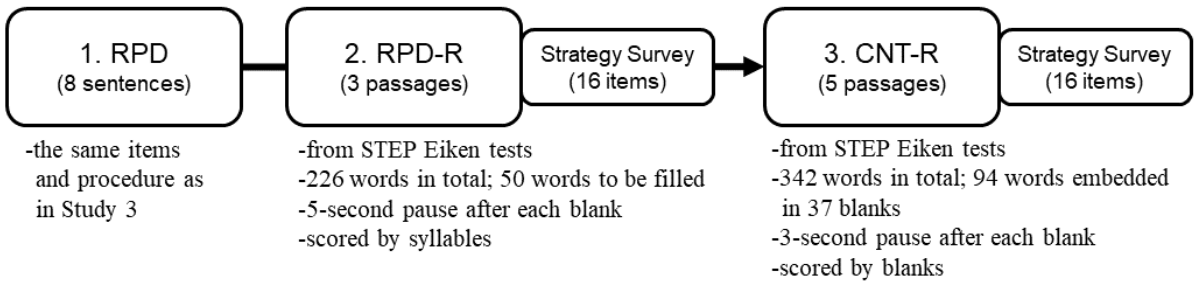
[1] According to Hirai (2017), substantive aspect of a test is a type of validity in Messick's framework. It refers to how much cognitive processes anticipated by theories are actually observed in learners' language activity, and can be assessed by post-experiment interviews and questionnaires.

6.2.3 Procedure

As illustrated in Figure 6.1, the study was conducted in three steps. First, the participants worked on the RPD. The same procedure as in Study 3 was used; they listened to each sentence twice, followed by a 10-second pause. The participants then performed the RPD-R. After being presented with a sample passage extracted from the Grade 3 of a STEP Eiken test held in the fall of 2018, they listened to three passages twice while being given a 5-second pause after each blank. After completing the task, they responded to 16 items on the strategy survey. Finally, the participants carried out the CNT-R and the strategy survey. In the same vein as the RPD-R, the task was initiated with a sample passage taken from the STEP Eiken Grade 3 test. Afterward, the participants approached the main task with five passages while being given a 3-second pause after each blank (Each passage was presented twice). After the completion of the CNT-R, the participants responded to the strategy survey and reported the difficulty of each task using a 5-point Likert scale ranging from 1 (*easy*) to 5 (*difficult*). This was aimed at examining the extent to which each task was learner friendly. Throughout the listening tasks, the participants were allowed to take notes if necessary.

Figure 6

Experimental Procedure of Study 4



Note. RPD = written reproduction task used in Study 3; RPD-R = revised version of the RPD; CNT-R = revised version of the word count task used in Study 3.

6.2.4 Scoring

For the RPD and the RPD-R, the same scoring system as in Study 3 was adopted. In other words, the reproduction rate of syllables spelled correctly was calculated for each participant. In the CNT-R, one point was given when participants were able to answer the correct number (total points = 37).

6.2.5 Analyses

To evaluate the validity of the RPD-R and the CNT-R, correlation coefficients among the scores in the three tasks were calculated. Moreover, to reveal whether cognitive processes involved in the two tasks were similar, participants' responses to the two strategy surveys were compared by performing a dependent samples *t* test for each item. Regarding reliability, internal consistency was measured for each test using Cronbach's α. Finally, to compare the difficulty level of the three tasks, a one-way repeated measures ANOVA was performed.

6.3 Results and Discussion

6.3.1 Descriptive Statistics and Reliability

Descriptive statistics of participants' scores in the three tasks and their reliability coefficients are shown in Table 6.4. The raw average score of the CNT-R was 24.07 (out of 37); thus, the correct answer rate was 65.05%. Its standard deviation was 6.29, higher than that of Study 3 (1.95). This suggests that increasing the number of items in the CNT-R contributed to larger score variances. In fact, the internal consistency of the CNT-R (α = .85) improved remarkably from that of the CNT (α = .45) and attained a somewhat higher reliability coefficient than the two dictation tasks (α = .77 and .77).

Table 6.4

Descriptive Statistics of the Three Tasks and Their Reliability Coefficients (N = 76)

Tests	*M*	*SD*	*SE*	*Median*	*Min*	*Max*	*α*
RPD (%)	47.92	12.80	1.47	49.09	10.91	74.55	.77
RPD-R (%)	51.62	15.41	1.77	54.69	6.25	84.38	.77
CNT-R	24.07	6.29	0.72	25	7	35	**.85**

Note. RPD = written reproduction task used in Study 3; RPD-R = revised version of the RPD; CNT-R = revised version of the word count task used in Study 3.

6.3.2 Validity

External Aspect (Correlation Analyses). As Table 6.5 shows, scores of the three tasks were positively correlated with each other. The CNT-R had positive correlations with the two written reproduction tasks. As a result of revising the CNT, the correlation coefficient with the RPD slightly improved from that of Study 3 (from .46 to .63). Furthermore, correlation between the CNT-R and the RPD-R was as high as .79. The fact that the CNT-R demonstrated a certain level of correlation coefficients with the written reproduction tasks indicates that the CNT-R can be an effective alternative to written reproduction tasks. However, the written reproduction task being an integrative task, further evidence is needed to conclude that the CNT-R is a valid measure of speech perception ability.

Table 6.5

Correlation Coefficients Among the Three Tests

	RPD	RPD-R	CNT-R
RPD	-		
RPD-R	.81**	-	
CNT-R	**.63****	**.79****	-

Note. RPD = written reproduction task used in Study 3; RPD-R = revised version of the RPD; CNT-R = revised version of the word count task used in Study 3.

**$p < .01$.

Substantive Aspect (Strategy Survey). As pointed out earlier, correlation analyses with written reproduction tasks are not sufficient to determine whether the CNT-R is valid for measuring the speech perception ability. To obtain further evidence, a strategy survey was conducted. The averages of participants' responses to the survey are summarized in Table 6.6.

Overall, in both tasks, the items related to phonological processing (Items 7, 12, 13, and 14) obtained higher averages than those related to meaning processing (Items 1–6, 8–10) and syntactic processing (Item 15). Item 13 (While listening, I paid attention to individual words.), in particular, showed high averages in both tasks (RPD-R = 3.79, CNT-R = 4.08). Additionally, Item 7 (I shadowed the English text.) exceeded an average of 3.50 in both tasks. This proves that both tasks can promote the subvocal rehearsal process, which is crucial for effective speech perception (Kadota, 2007). In contrast, the average of Item 10 (I tried to understand the details of each passage.) reached about 2.00 in both tasks, indicating that meaning processing tends to be restricted in both the tasks. From these considerations, it can be concluded that both the CNT-R and the RPD-R are valid methods to measure speech perception ability.

Table 6.6

Listening Strategies Used by the Participants During the RPD-R and the CNT-R and the Results of Dependent T Tests Between the Two Tasks (N = 76)

Items	(1) RPD-R	(2) CNT-R	(1)-(2)	*p*
Cognitive Strategies				
While listening, I…				
1. guessed where the story was going based on what I had understood.	2.84 [1.19]	2.82 [1.37]	0.02	.865
2. imagined the story based on words and phrases I perceived.	**3.57 [1.17]**	3.25 [1.41]	0.32	**.040**
3. used my background knowledge related to the topic.	2.71 [1.28]	2.66 [1.40]	0.05	.703
4. pictured the story in my head.	2.14 [1.20]	2.16 [1.20]	-0.01	.922

5. summarized the gist of the story.	2.42 [1.21]	2.51 [1.31]	-0.09	.403
6. translated English into Japanese.	2.75 [1.32]	2.47 [1.24]	0.28	.056
7. shadowed the English text (repeated the text in my head).	**3.59 [1.30]**	**3.88 [1.21]**	-0.29	.080
8. looked for keywords.	3.43 [1.35]	2.95 [1.31]	0.49	**.004**
Metacognitive Strategies				
While listening, I…				
9. tried to understand each sentence precisely.	2.78 [1.37]	2.66 [1.33]	0.12	.435
10. tried to understand the details of each passage.	2.13 [1.08]	2.22 [1.13]	-0.09	.330
11. tried to understand the global meaning of each passage.	**3.88 [1.14]**	**3.55 [1.35]**	0.33	**.013**
12. paid attention to detailed sounds like the plural *-s* and past tense *-ed*.	**4.13 [0.90]**	3.09 [1.30]	1.04	**.000**
13. paid attention to individual words.	**3.79 [1.09]**	**4.08 [1.14]**	-0.29	.059
14. tried to figure the phrase chunks.	3.24 [1.22]	**3.80 [1.22]**	-0.57	**.000**
15. paid attention to the grammatical structures.	3.08 [1.22]	2.82 [1.29]	0.26	**.045**
16. asked myself whether or not my understanding was correct.	3.17 [1.28]	3.05 [1.34]	0.12	.401

Note. Averages over 3.50 and *p*-values under .05 are boldfaced and underlined. RPD-R = revised version of the written reproduction task used in Study 3; CNT-R = revised version of the word count task used in Study 3.

In terms of which task is more suitable for measuring speech perception ability, inconclusive results were found. As for the RPD-R, the highest average (M = 4.13) was found in Item 12 (I paid attention to detailed sounds like the plural *-s* and past tense *-ed*.). Among all the items, the largest gap between the two tasks (1.04) was observed for this item, and a dependent t test revealed that this difference was statistically significant, t (75) = 6.91, p = .000, with a large effect size, d = 0.93. These results seem to indicate that the RPD-R, where target words must be spelled out in a short course, enhances phonological processing to a degree greater than the CNT-R. In other words, the RPD-R promotes the input decoding process in the Field (2013) model. However, the results of other items may contradict this conclusion. In Item

13 (I paid attention to individual words.), the average of the CNT-R (M = 4.08) was slightly higher than that of the RPD-R (M = 3.79), with the gap almost reaching a significant level (p = .059). Moreover, a significantly higher average of the CNT-R was observed for Item 14, which is concerned with learners' awareness of their phonological processing at the chunk level.

Furthermore, there is also evidence that learners are more likely to utilize semantic information during the RPD-R than the CNT-R. Specifically, in the RPD-R, the averages of Items 2 and 8 were approximately 3.50, and that of Item 11 was close to 4.00. These averages were significantly larger than those in the CNT-R (p = .040, .004, and .013, respectively). Cognitive activities described in these items cannot be performed without meaning processing; thus, learners' involvement in meaning processing seems to be greater in the RPD-R than in the CNT-R. In order to measure learners' speech perception ability accurately, meaning processing should be suppressed to the highest degree possible, although it cannot be completely excluded.

In summary, the strategy survey revealed that, in both the CNT-R and the RPD-R, phonological processing was more dominant than meaning and syntactic processing. Although the RPD-R demands that a learner employs exquisite processing toward sounds, it also involves higher-level processing compared to the CNT-R. Therefore, further research is necessary to conclude which task is more appropriate for measuring speech perception ability.

6.3.3 Difficulty

At the end of the study, participants were asked to report the difficulty of each task based on a 5-point Likert scale ranging from 1 (*easy*) to 5 (*difficult*). Averages of the RPD, RPD-R, and CNT-R were 4.36 (SD = 0.74), 3.38 (SD = 0.86), and 3.21 (SD = 1.06), respectively. A one-way ANOVA revealed that there was a significant difference among the three averages, F (1.67, 122.86) = 50.91, p = .000, η_p^2 = .404. Pairwise comparisons revealed significant differences between the RPD and RPD-R (p = .000) and the RPD and CNT-R (p = .000). Clearly,

the participants considered the RPD to be the most difficult task. In all likelihood, this is attributed to the difficulty of materials that were chosen from the TOEIC preparation book. In contrast, the difference between the RPD-R and CNT-R was not significant (p = .509); thus, the difficulty of the CNT-R was moderate for those who attempted this task.

6.4 Summary of the Findings: Chapter 6 (Study 4)

6.4.1 Answer to RQ4

RQ4: Is the revised version of the word count task (CNT-R) valid and reliable as a measure of speech perception ability?

In this study, a CNT-R was compared with two kinds of written reproduction tasks. One was the RPD from Study 3, and the other was an RPD-R. From the results of these three listening tasks (i.e., the RPD, RPD-R, and CNT-R) and a survey on listening strategies, four findings were obtained. First, as a result of revising the material, the internal consistency of the CNT-R showed a remarkable improvement from Study 3 and demonstrated the highest reliability coefficient among the three tasks. Second, the CNT-R scores had moderate to strong positive correlations with the two written reproduction tasks, meaning the CNT-R has a certain level of external aspect validity. Third, the strategy survey revealed that phonological processing was more dominant than meaning processing in the CNT-R. Fourth, the difficulty of the CNT-R was found to be moderate.

The findings listed above suggest that the CNT-R can be a valid and reliable method to measure speech perception ability. Since it is friendly to raters and learners, it has several potential applications for future research. First, since learners' spelling knowledge does not interfere with their task performance, it will be easy to interpret the outcome of experiments. Second, because participants write only numbers, the task is not time consuming. This, in turn, means that a researcher can provide participants with more items than a RPD with the same

amount of time, which may contribute to higher reliability. Third, unlike written reproduction tasks, a rater does not have to be perplexed with poor handwriting; hence, inter-rater reliability need not be calculated. The CNT-R will thus help save time and energy. Fourth, the CNT-R is simple to perform, even for learners who are unfamiliar with this task; thus, the time spent for in-advance training can be minimized.

6.4.2 Study Limitations

This study has two limitations. First, this research could not specify which task is more appropriate for measuring speech perception ability: the CNT-R or the RPD-R. The RPD-R seemed to engage learners in more elaborate phonological processing than the CNT-R, while the CNT-R involved less meaning processing. Second, the validity of the CNT-R may need to be reexamined by means of methods other than those used in this research (i.e., correlations with the RPD-R and the strategy survey). As already pointed out, written reproduction tasks are integrative tasks; thus, task performance is inevitably affected by learners' spelling knowledge. Strategy surveys can only elicit subjunctive responses from learners; therefore, it is likely that the results may not reflect what was actually happening during the task.

However, there is currently no other valid and reliable research paradigm; hence, it may be safer to use both the CNT-R and the RPD-R in the further study. For this reason, the next study, in which the long-term effects of accelerated speech dictation and shadowing are compared, will use both tasks in the pre- and post-tests.

Chapter 7

Study 5-1: Long-Term Training Effects of Accelerated Speech Dictation and Shadowing on Speech Perception Ability (Quantitative Analyses)

7.1 Study Goal

Study 2 investigated the short-term effect of two training tasks—accelerated speech dictation and shadowing—on learners' speech perception ability. Although the result revealed significant improvement in both training groups, the training effect was unclear since there was no control group, thus, their improved performance in the post-test may have been a practice effect. Moreover, the lack of significant interaction between the task factor and the test factor (pre- and post-tests) was probably due to the length of the training (15 mins.). Therefore, this study aims to verify the long-term training effect of the two exercises in comparison to the absence of training. To accomplish this goal, Study 5 implements the assessment methods developed in Studies 3 and 4 (i.e., revised versions of the word count task and the written reproduction task) for the improvement of speech perception ability, and a strategy survey and written journals for a comprehensive assessment. Chapter 7 discusses the quantitative analyses of the results of the two assessment tasks and the strategy survey as Study 5-1. The qualitative analyses of the written journals will be discussed in Chapter 8 as Study 5-2.

Based on these backgrounds, this study examines the following RQs:

RQ5-1: Will long-term training with accelerated speech dictation and shadowing improve learners' speech perception ability?

RQ5-2: Will the training develop learners' strategy use in listening?

7.2 Method

7.2.1 Participants

The study participants included 84 university students (40 men and 44 women) who belonged to an education department and came from varied majors. The majority were 1st-year students (see Appendix 7A for the breakdown of their majors and grades). Fifty-six students, who participated in the experimental groups (i.e., accelerated speech dictation group or shadowing group), were attending an elective English course taught by the author. This one-semester course, held every week for 15 classes (each class was 90 mins.), started with 74 students. However, only the 56 students who completed all the mentioned tests and took the speech perception training for at least 5 weeks[1] were included in the analysis. Although the course was intended to foster the four English skills—listening, reading, speaking, and writing—using recent Grade 2 STEP Eiken tests, most of the class time was spent on reading and listening exercises. As part of the listening practice, participants chose either accelerated speech dictation (n = 27) or shadowing (n = 29), and engaged in the activity for eight weeks. The remaining 28 participants, who were not taking the course, were recruited from the same department for the control group. None of the 84 participants reported any hearing impairments.

7.2.2 Materials

Pre- and Post-Tests. The two tests developed in Study 4 were conducted in the pre- and post-tests to assess the improvement in participants' speech perception ability. Although the same materials were used for both the pre- and post-tests, the practice effect was assumed to be limited for two reasons. First, the post-test was administered two months after the pre-test and the participants were not given the correct answers, thus, their memory had little impact on

[1] The precondition for five weeks is derived from Tamai (1997), who revealed the effectiveness of shadowing exercise implemented for five classes.

their performance. Second, since the participants were not aware that they would take the same test later, it was unlikely that they had attempted to memorize the test items. In addition, there was a control group, which made it possible to assess whether the improvement was due to the training or merely a practice effect.

A new test (see Appendix 7B) was prepared for the pre-test to examine the general listening ability of the participants. This test, composed of 20 listening comprehension questions with 10 dialogues and 10 monologues, was extracted from a preparation textbook for the Grade 2 STEP Eiken test (Obunsha, 2019). These test scores were also used as a covariate in cases where there were gaps in the speech perception ability of the three groups in the pre-test.

Training Materials. Passages from three Grade 2 STEP Eiken tests administered in the 2020 academic year were used in each class. The participants listened to the passages once and answered listening comprehension questions. They were provided regular opportunities to enhance their speech perception ability. A training session was held in each class after they had answered the listening comprehension questions for about seven to eight Eiken passages. Subsequently, the passages were used as training material with which the participants engaged in either dictation or shadowing. For the dictation group, an accelerated version was prepared for each passage by increasing its speech rate by 30%. As a result of this alteration, the average passage speed reached 200 wpm (approximately).

Weekly Journal. To obtain qualitative evidence of the improvement of speech perception ability, the participants were asked to keep a journal after the training every week (see Appendix 7C). This was documented online via Google Forms. The participants were instructed to write in Japanese about how they worked on the activity, what sounds were

especially difficult to reproduce, and how their listening ability improved. To encourage the completion of this assignment, the author kept reminding the participants that their responses would be graded irrespective of the content. As stated at the beginning of this chapter, analyses of the texts obtained from the journals will be discussed in the next chapter as Study 5-2.

Questionnaire. The experimental and control groups responded to a questionnaire after the post-tests (see Appendices 7D and 7E). The questionnaire had two sections: a listening strategy survey with 16 items used in Study 4 and a free writing section. The objective of each section was different for the experimental groups and the control group. Regarding the listening strategy survey, the objective for the experimental groups was to explore the impact of the training on their use of listening strategies. However, unlike Study 4, the participants selected one of the following three options, *agree*, *mildly agree*, or *disagree,* to indicate the development of the listening strategies. In contrast, the control group was asked to rate the importance of each strategy (or ability) by choosing one from the following three options, *important*, *neutral*, and *not important*. Regarding the free writing item, the experimental groups wrote their feedback about the training, while the control group reported the type of English study they had engaged in since the pre-test to confirm that no one had engaged in dictation or shadowing regularly between the pre- and post-tests. Written responses of both groups will be discussed in the next chapter.

7.2.3 Procedure

As illustrated in Figure 7.1, this study was conducted in three phases. First, three kinds of pre-tests were administered under the author's instruction. The 56 participants of the experimental groups took the tests together, during the class, in the first week of the English course as an investigation of their listening ability. In contrast, the 28 participants of the control

group, who were recruited for this research from outside of the course, took the tests separately from the experimental groups on dates that were convenient for them owing to the availability of their schedules. However, the time gap between the pre- and post-tests was controlled to be the length of the training session, i.e., eight weeks. All the groups listened to the audio materials through speakers in the classroom, writing their answers on printed test sheets. After the pre-tests, the participants were provided an opportunity to experience both types of training so that they can choose their preferred training. Following trial training, each participant reported their preferred training type. The experimental groups (i.e., accelerated speech dictation group = ASD group, shadowing group = SH group) were determined by these results.

Figure 7.1

Experimental Procedure of Study 5 (5-1 and 5-2)

Second, in the training session, only the experimental groups engaged in speech perception training—accelerated speech dictation or shadowing—for eight weeks as part of the course. During this session, starting in the second week, the participants worked on their selected training, using the materials described in the previous section. They were permitted to work on the same materials repeatedly. Written scripts were initially provided, however, they

were strongly advised to avoid looking at them until necessary. As Study 2 revealed, shadowing can be difficult for some learners; the participants who had selected shadowing were allowed to adjust the speech rate of the audio materials. To accommodate individual learning, the participants used their smartphones or laptops and heard the audio materials through earphones.

In contrast to the experimental groups, the control group was not given any instruction by the author. To confirm that they had not undergone intensive training in dictation or shadowing during the eight weeks, their response to the last question of the questionnaire was inspected. It proved that all 28 participants were qualified for the study.

Third, in the post-test phase, the participants took the two tests to measure the improvement in their speech perception ability with the same procedure as the pre-tests. Subsequently, they responded to the questionnaire explained previously.

7.2.4 Scoring

For the revised version of the written reproduction task (RPD-R), the scoring system of Study 4 was adopted. The reproduction rate of syllables (total syllables = 64) spelled correctly was calculated for each participant. In the revised version of the word count task (CNT-R), one point was given when participants were able to give the correct number (total points = 37), while they obtained no point when the answer was wrong. The proficiency test, composed of 20 comprehension questions, was scored by counting the number of correct answers.

7.2.5 Analyses

To examine the two RQs, two analyses were conducted. To study the influence of the training on score improvement of the two tasks (i.e., RPD-R and CNT-R), a mixed design of a two-way MANCOVA in a 2 (pre- and post-test; within-subject variable) × 3 (ASD group, SH

group, and control group; between-subject variable) format, with the proficiency test score as a covariate, was planned because the improvement of speech perception ability may vary according to the general listening ability. An assumption of the analyses, using a covariate, is the independence between the covariate and the independent variable (Field, 2009). To examine this, a one-way ANOVA was performed to test whether the three treatment groups differed in the proficiency test scores (see Table 7.1). The analysis revealed a significant difference, F (2, 83) = 3.76, p = .027[2], indicating that the two variables are not independent of each other and, thus, the previous assumption was not fulfilled. Therefore, a two-way MANOVA was conducted instead of the planned MANCOVA.

Table 7.1

Descriptive Statistics of the Listening Proficiency Test (N = 84)

	n	Part 1 (Dialogue)	Part 2 (Monologue)	Total
ASD Group	27	4.96 (2.26)	4.52 (2.64)	9.48 (4.36)
SH Group	29	5.17 (2.22)	3.83 (2.16)	9.00 (3.92)
Control Group	28	6.29 (2.03)	5.39 (2.22)	11.68 (3.40)
Total	84	5.48 (2.23)	4.57 (2.40)	10.05 (4.04)

Notes. ASD = Accelerated Speech Dictation, SH = Shadowing; Standard deviations are presented in parentheses. The maximum total score is 20.

To reveal whether the training contributed to a shift in listening strategy use of the two experimental groups, a chi-square test for independence between the training type and the participants' response pattern was performed for each strategy item. As for the control group, their responses were analyzed by calculating the percentage of participants for each answer option to investigate which strategy they recognized to be important in listening.

[2] Multiple comparisons with the Tukey method revealed that the control group significantly outperformed the SH group (p = .030), while the other pairs were insignificant.

7.3 Results and Discussion

7.3.1 Improvement in Speech Perception Ability (Performances in RPD-R and CNT-R)

Descriptive Statistics. Participants' performances in the two tasks in the pre- and post-tests are summarized in Table 7.2. As the graphs illustrate (see Figures 7.2 and 7.3), in the pre-tests of both the tasks, the control group demonstrated the highest performance while the SH group had the lowest performance. The gap between the control group and the two experimental groups in the pre-test was larger in the RPD-R than in the CNT-R. In the RPD-R, where the reproduction rates of all the groups increased in the post-test, the control and SH groups improved at a parallel rate (4.40% and 5.76%, respectively); however, the ASD group had a more dramatic increment (9.72%). This indicates that accelerated speech dictation was more effective than shadowing. However, in the CNT-R, even though the increase of all the groups was minimal (i.e., the development of the total participants was 1.51 points), the SH group showed a relatively larger development (3.38 points), performing at par with the other groups in the post-test.

Table 7.2
Descriptive Statistics of the Speech Perception Tasks in Pre- and Post-Tests (N = 84)

	n	RPD-R		CNT-R	
		Pre-test	Post-test	Pre-test	Post-test
ASD Group	27	52.20% (18.66)	61.92% (18.43)	26.19 (5.99)	27.11 (6.57)
SH Group	29	50.81% (11.65)	56.57% (13.33)	22.79 (5.48)	26.17 (5.31)
Control Group	28	64.29% (13.86)	68.69% (13.58)	29.07 (4.12)	29.21 (3.50)
Total	84	55.75% (15.96)	62.33% (15.86)	25.98 (5.80)	27.49 (5.35)

Notes. ASD = Accelerated Speech Dictation, SH = Shadowing, RPD-R = the revised written reproduction task, CNT-R = the revised word count task; Standard deviations are presented in parentheses. The maximum total score of the revised word count task is 37.

Figure 7.2

Reproduction Rate Improvement of the Three Groups in the Revised Written Reproduction Task (RPD-R)

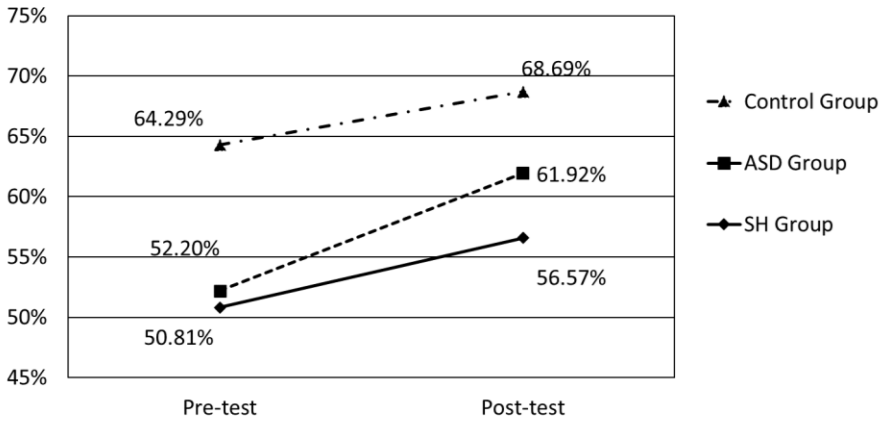
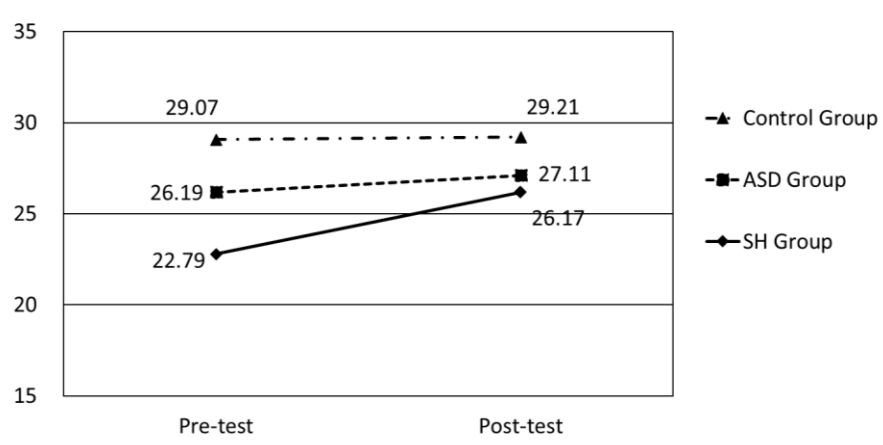


Figure 7.3

Score Improvement of the Three Groups in the Revised Word Count Task (CNT-R)



Multivariate Effect. A two-way MANOVA examined participants' performances in the two tasks as dependent variables (DVs), and the types of speech perception training and the scores in the pre- and post-tests as independent variables (IVs). The multivariate analysis revealed significant interaction between the two IVs, Pillai's Trace = .236, F (4, 162) = 5.42, p = .000, η_p^2 = .118. A two-way univariate ANOVA was performed for each DV, revealing a significant interaction for both the RPD-R, F (2, 81) = 3.90, p = .024, η_p^2 = .088, and the CNT-R, F (2, 81) = 6.90, p = .002, η_p^2 = .146 (see Table 7.3). These results indicate that, in both tasks,

the improvement depended on the training type, therefore, follow-up analyses were conducted to test the simple main effect in each task.

Table 7.3

Multivariate and Univariate Analyses of Variance for the Speech Perception Ability Measures

	Multivariate			Univariate					
				RPD-R			CNT-R		
Source	F^a	p	η_p^2	F^b	p	η_p^2	F^b	p	η_p^2
Training Type (T)	3.56	.008	.081	5.82	.004	.126	6.28	.003	.134
Pre-, Post-Tests (P)	38.77	.000	.492	68.84	.000	.459	15.63	.000	.162
T × P	5.42	.000	.118	3.90	.024	.088	6.90	.002	.146

Note. Multivariate F ratios were generated from Pillai's statistics.

[a]Multivariate $df = 4, 162$. [b]Univariate $df = 2, 81$.

Tests for the Simple Main Effect (RPD-R). As stated earlier, all the groups showed higher reproduction rates in the post-test of the RPD-R. To confirm whether the rate increase was statistically significant, independent sample *t* tests with Bonferroni correction adjusting the alpha level to be .025 were performed, indicating that the improvement of each group was significant (ASD group, $p = .000$; SH group, $p = .000$; control group, $p = .001$). Significant improvement in both the ASD and SH groups corresponds to the finding of Study 2, which revealed the short-term effects of accelerated speech dictation and shadowing on better dictation performances.

However, the improved performance of the control group in the post-test, who received no training, seems to suggest that the improvement emerged as a practice effect. Nevertheless, there were differences in the rate of increase among the three groups, whereby, the ASD group showed the largest increase. To examine the differences among the three groups in each test, one-way ANOVAs were conducted, revealing the significant group difference for both the pre-

test, $F(2, 67.64) = 6.82$, $p = .002$[3], and the post-test, $F(2, 83) = 4.53$, $p = .014$. Post-hoc multiple comparisons with Scheffe's method were performed for both the pre- and the post-tests. As summarized in Table 7.4, although the analyses revealed significant differences between the two experimental groups and the control group in the pre-test (p-values were .014 with the ASD group and .004 with the SH group), the contrast between the ASD group and the control group in the post-test became insignificant ($p = .263$). This result suggests that long-term training through accelerated speech dictation exhibited a greater effect on the development of speech perception ability than shadowing.

Table 7.4

Results of Multiple Comparisons Among the Three Groups in the RPD-R

Tasks	(I) Group	(J) Group	(I-J)	p	95%CI	
					Lower	Upper
Pre-test	ASD	SH	1.39	.941	-8.56	11.34
		Control	-12.09	.014*	-22.12	-2.05
	SH	Control	-13.48	.004**	-23.34	-3.62
Post-test	ASD	SH	5.35	.426	-4.81	15.51
		Control	-6.77	.263	-17.02	3.47
	SH	Control	-12.12	.014*	-22.18	-2.06

*$p < .025$., **$p < .005$.

Tests for the Simple Main Effect (CNT-R). In contrast to the predominance of the ASD group in the RPD-R, only the SH group took advantage of the training in the CNT-R. Independent samples t tests for the pre- and post-test differences revealed that only the SH group showed significant improvement in the post-test (ASD group, $p = .241$; SH group, $p = .000$; control group, $p = .792$). In addition, a one-way ANOVA for the group differences in the pre-test revealed that a significant score gap was observed between the SH group and the

[3] Since Levene's test indicated inequality of variances among the three groups in the pre-test, Brown-Forsythe test was conducted instead of the one-way ANOVA.

control group (see Table 7.5). However, a Brown-Forsythe test revealed that the group differences in the post-test were insignificant, F (2, 65.65) = 2.46, p = .093. These results indicate that only the SH group increased their scores and achieved a performance level comparable with the control group in the post-test.

Table 7.5

Results of Multiple Comparisons Among the Three Groups on the Pre-test of the CNT-R

Tasks	(I) Group	(J) Group	(I-J)	*p*	95%CI	
					Lower	Upper
Pre-test	ASD	SH	3.39	.060	-0.11	6.89
		Control	-2.89	.132	-6.42	0.64
	SH	Control	-6.28	.000**	-9.75	-2.81

Note. The group differences in the post-test were insignificant (p = .093).
*p < .025., **p < .005.

7.3.2 Results of the Listening Strategy Survey (Experimental Groups)

The listening strategy survey, consisting of 16 items and conducted at the end of this study, aimed to reveal the contribution of the two-month training to the development of listening strategies, especially those related to speech perception, and compare the effects between the two experimental groups. The items can be grouped into two categories: (1) those related to cognitive strategies (Items 1–8) and (2) those related to metacognitive strategies (Items 9–16). Results of each category are presented next.

Cognitive Strategies (Items 1-8). Table 7.6 and Figure 7.4 display the response of the two experimental groups to the eight items concerned with cognitive strategies. More than 70% of the participants in both the groups answered either "*agree*" or "*mildly agree*" for each item, although Item 1 in the SH group fell short of this percentage by 1%. Moreover, none of the association coefficients between the training groups and the response patterns was significant

(except for Item 7), indicating that majority of both the groups recognized the training effect on the development of each strategy. The items in this category included strategies necessary for top-down listening such as the inference of passage meaning based on context (Item 1) and the use of background knowledge (Item 3). The development of their strategy use for top-down listening probably resulted from the speech perception training, given the theoretical assumption that the better the speech perception ability, the more cognitive resource can be assigned to comprehension while listening (e.g., Kadota, 2007).

Table 7.6

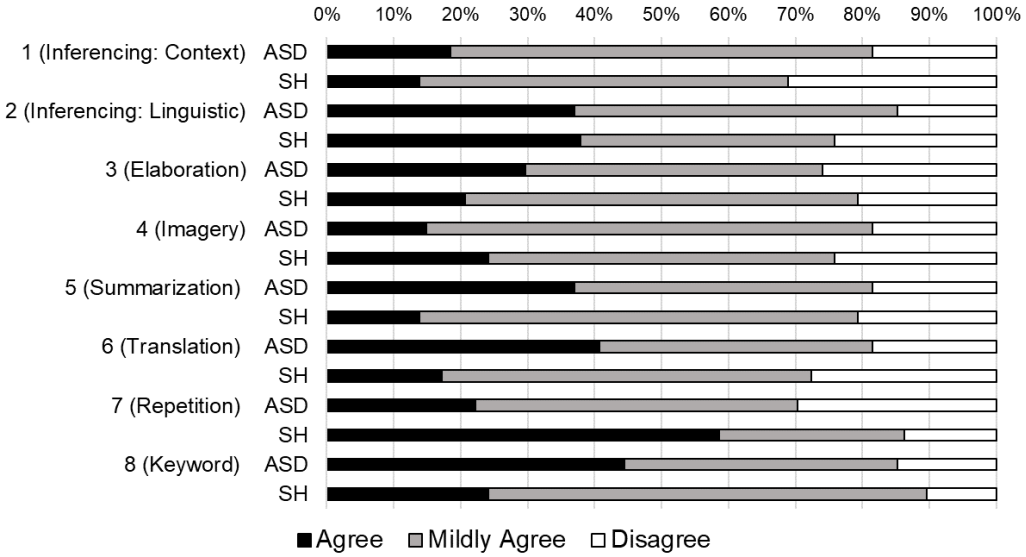
Responses of the Experimental Groups to the Listening Strategy Questionnaire (Cognitive Strategies: 1-8) and the Results of the Tests of Independence Between the Training Groups and the Response Pattern

	ASD Group (n = 27)			SH Group (n = 29)			V
Strategy	*Agree*	*Mildly*	*Disagree*	*Agree*	*Mildly*	*Disagree*	
1 - Inferencing: Context	5 (18.5%)	**17 (63.0%)**	5 (18.5%)	4 (13.8%)	**16 (55.2%)**	9 (31.0%)	.15
2 - Inferencing: Linguistic	10 (37.0%)	**13 (48.1%)**	4 (14.8%)	**11 (37.9%)**	**11 (37.9%)**	7 (24.1%)	.13
3 - Elaboration	8 (29.6%)	**12 (44.4%)**	7 (25.9%)	6 (20.7%)	**17 (58.6%)**	6 (20.7%)	.14
4 - Imagery	4 (14.8%)	**18 (66.7%)**	5 (18.5%)	7 (24.1%)	**15 (51.7%)**	7 (24.1%)	.16
5 - Summarization	10 (37.0%)	**12 (44.4%)**	5 (18.5%)	4 (13.8%)	**19 (65.5%)**	6 (20.7%)	.27
6 - Translation	**11 (40.7%)**	**11 (40.7%)**	5 (18.5%)	5 (17.2%)	**16 (55.2%)**	8 (27.6%)	.26
7 - Repetition	6 (22.2%)	**13 (48.1%)**	8 (29.6%)	**17 (58.6%)**	8 (27.6%)	4 (13.8%)	.37*
8 - Keyword	**12 (44.4%)**	11 (40.7%)	4 (14.8%)	7 (24.1%)	**19 (65.5%)**	3 (10.3%)	.25

Notes: The values in parentheses indicate percentages in the group; "*Mildly*" means "*Mildly Agree*"; *V* is Cramer's *V*; the highest frequency and percentage in each item are bold-faced and underlined; Item 1 (the ability to guess where the story is going based on what I have

understood); Item 2 (the ability to imagine the story based on words and phrases I have perceived); Item 3 (the ability to use my background knowledge related to the topic); Item 4 (the ability to picture the story in my head); Item 5 (the ability to summarize the gist of the story); Item 6 (the ability to translate English into Japanese); Item 7 (the ability to shadow the English text); Item 8 (the ability to look for keywords).
$^{*}p < .05$.

Figure 7.4
Responses of the Two Experimental Groups to the Strategy Survey (Cognitive Strategies)



However, there are a few differences between the two groups. The ASD group responded positively to Items 6 (the ability to translate English into Japanese) and 8 (the ability to look for keywords) as more than 40% of the participants selected "*agree*" for these strategies. Although translation into L1 is not necessarily effective for fluent listening, it cannot be accomplished without attention to meaning. Similarly, to look for keywords, learners must pay attention to the macro-structure of the passage's meaning and extract important information. Both results indicate that the participants in the ASD group developed strategies to process the passage meaning through dictation training. This corroborates the claim made in the previous section that language activities and measurement tasks using dictation promote meaning processing, which was why the ASD group demonstrated the greatest improvement in the RPD-

R.

The SH group showed a distinctive response to Item 7 (the ability to shadow the English text) with 58.6% of the participants (17 out of 29 participants) selecting "*agree*" for the improvement of this strategy; the highest percentage among all the 16 items. As the result of the chi-square tests for independence, only this item showed significant association between the training type and the response pattern, X^2 (df = 2, N = 56) = 7.72, p = .021, with a large coefficient, Cramer's V = .37. Furthermore, analyses of the adjusted standardized residuals revealed the gap between the observed value and the expected value for the cell to be |2.8|, indicating that the percentage was statistically large. The result stipulates that the majority of the group acknowledged that, thanks to the shadowing training, they had become more skillful at repeating the aural input in their mind. As discussed earlier, the significant improvement of the SH group in the CNT-R can be attributed to the discipline of their working memory through long-term shadowing, which is evidenced in the result of this survey.

Metacognitive Strategies (Items 9-16). Table 7.7 and Figure 7.5 display the response of the two experimental groups to the eight items concerned with metacognitive strategies. Items 12 (the ability to pay attention to detailed sounds like the plural -*s* and past tense -*ed*), 13 (the ability to pay attention to individual words), and 14 (the ability to figure the phrase chunks) were aimed at investigating whether the training had promoted strategy use for accurate speech perception. Percentages of the participants who rated either "*agree*" or "*mildly agree*" for these items ranged from 72.4% (SH group on Item 12) to 86.2% (SH group on Item 13) with an average of approximately 80%. The high percentages of both the experimental groups imply that the long-term training made the participants more aware of the perception strategies and, thus, enhanced the training effect.

Table 7.7

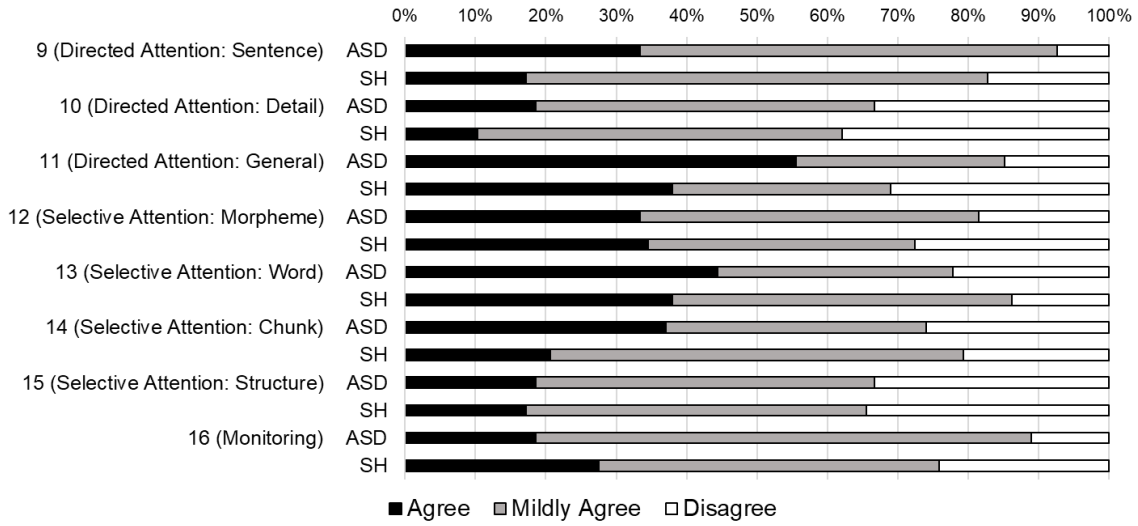
Responses of the Experimental Groups to the Listening Strategy Survey (Metacognitive Strategies: 9-16) and the Results of Tests of Independence Between the Training Group and the Response Pattern

Strategy	ASD Group (n = 27)			SH Group (n = 29)			*V*
	Agree	*Mildly*	*Disagree*	*Agree*	*Mildly*	*Disagree*	
9 - Directed Attention: Sentence	9 (33.3%)	**16 (59.3%)**	2 (7.4%)	5 (17.2%)	**19 (65.5%)**	5 (17.2%)	.22
10 - Directed Attention: Detail	5 (18.5%)	**13 (48.1%)**	9 (33.3%)	3 (10.3%)	**15 (51.7%)**	11 (37.9%)	.12
11 - Directed Attention: General	**15 (55.6%)**	8 (29.6%)	4 (14.8%)	**11 (37.9%)**	9 (31.0%)	9 (31.0%)	.21
12 - Selective Attention: Morpheme	9 (33.3%)	**13 (48.1%)**	5 (18.5%)	10 (34.5%)	**11 (37.9%)**	8 (27.6%)	.12
13 - Selective Attention: Word	**12 (44.4%)**	9 (33.3%)	6 (22.2%)	11 (37.9%)	**14 (48.3%)**	4 (13.8%)	.16
14 - Selective Attention: Chunk	**10 (37.0%)**	**10 (37.0%)**	7 (25.9%)	6 (20.7%)	**17 (58.6%)**	6 (20.7%)	.23
15 - Selective Attention: Structure	5 (18.5%)	**13 (48.1%)**	9 (33.3%)	5 (17.2%)	**14 (48.3%)**	10 (34.5%)	.02
16 - Monitoring	5 (18.5%)	**19 (70.4%)**	3 (11.1%)	8 (27.6%)	**14 (48.3%)**	7 (24.1%)	.13

Notes: The values in parentheses show percentages in the group; "*Mildly*" means "*Mildly Agree*"; *V* is Cramer's *V*; None of the association coefficients was significant; Item 9 (the ability to understand each sentence precisely); Item 10 (the ability to understand the details of each passage); Item 11 (the ability to understand the global meaning of each passage); Item 12 (the ability to pay attention to detailed sounds like the plural *-s* and past tense *-ed*); Item 13 (the ability to pay attention to individual words); Item 14 (the ability to figure the phrase chunks); Item 15 (the ability to pay attention to the grammatical structures); Item 16 (the ability to ask whether or not the understanding is correct).

Figure 7.5

Responses of the Two Experimental Groups to the Strategy Survey (Metacognitive Strategies)



Fewer participants acknowledged the training effect on strategies 10 (the ability to understand the details of each passage) and 15 (the ability to pay attention to the grammatical structures) with less than 20% of the participants of both the groups choosing "*agree*" for these items. The former strategy refers to deeper meaning processing, while the latter is concerned with grammatical processing. Both the strategies operate at higher levels than speech perception. Therefore, it can be concluded that speech perception training, using accelerated speech dictation and shadowing, can enhance listening strategies related to speech perception rather than those related to semantic and grammatical processing.

Although the associations between the training type and the response pattern in all the eight items (Items 9-16) were found to be insignificant, the two experimental groups reacted differently to the two items. For items 9 (the ability to understand each sentence precisely) and 11 (the ability to understand the global meaning of each passage), a larger percentage of participants answered "*agree*" in the ASD group. Moreover, a total of 85.2% (23 out of 27 participants) of the group chose either "*agree*" or "*mildly agree*" for Item 11, while 69.0% (20 out of 29 participants) of the SH group chose those responses. These two strategies are

distinctive from each other in terms of the depth of meaning processing but akin to each other as both refer to strategies used for comprehension of the passage's meaning rather than for perception of individual words. As discussed earlier, accelerated speech dictation is more likely to promote semantic processing than shadowing. The above results provide auxiliary evidence for this claim.

7.3.3 Results of the Listening Strategy Survey (Control Group)

While the survey for the experimental groups was aimed at exploring the development of their strategy use, the survey for the control group was intended to investigate the importance of each strategy (or ability) as per the participants. As indicated in Table 7.8 and Figure 7.6, the participants' ratings varied markedly across strategies. For example, more than 80% of the group recognized the importance of strategies 1 (the ability to guess where the story is going based on what I have understood), 2 (the ability to imagine the story based on words and phrases I perceived), 11 (the ability to understand the global meaning of each passage), and 14 (the ability to figure the phrase chunks). Except 14, the other three strategies are associated with inference or comprehension of the whole passage rather than the perception of individual words or sentences. In contrast, the importance ratings of comprehension strategies utilized to understand the details of the passage (Items 9 and 10) were fairly low with less than 20% of the group evaluating these strategies as "*important*."

These results suggest that top-down listening skills are prioritized by the control group. In fact, abilities required in bottom-up listening (Items 6, 13, and 15) were underestimated with less than 40% of the group rating them as "*important*." In addition, Item 7 (the ability to shadow the English text), which the SH group recognized as the most improved after the training, did not achieve a high rating, indicating that the control group was not very aware of its importance. Therefore, it can be concluded that long-term speech perception training can improve learners'

performance in the speech perception tasks (i.e., RPD-R and CNT-R) and enhance the strategies required for speech perception.

Table 7.8

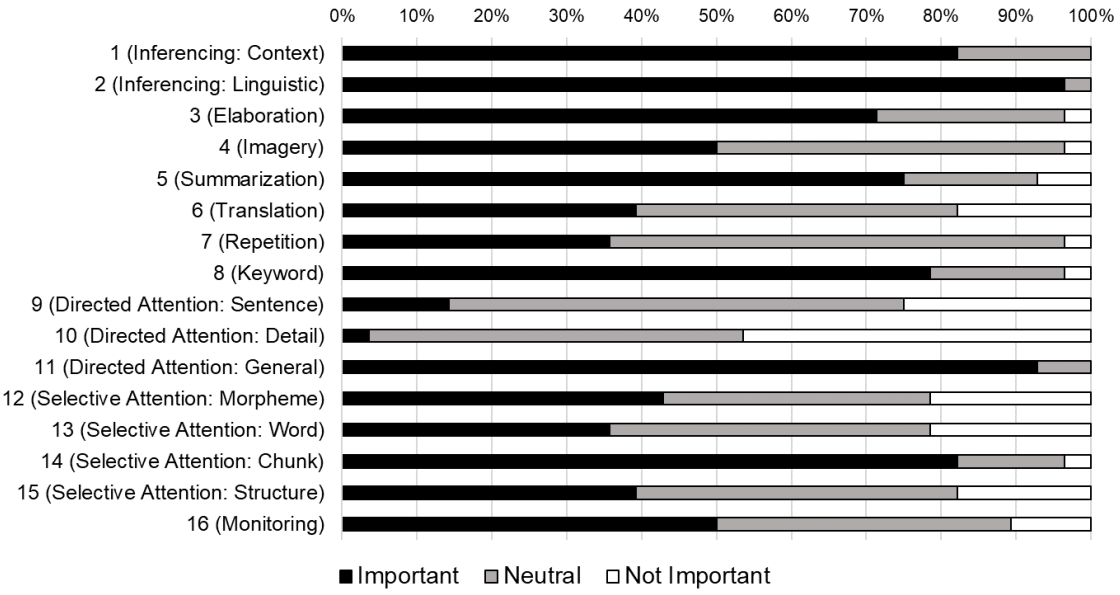
The Importance of Each Listening Skill as Recognized by the Control Group (n = 28)

Ability	Important	Neutral	Not Important
Cognitive Strategies			
1 - Inferencing: Context	**23 (82.1%)**	5 (17.9%)	0 (0.0%)
2 - Inferencing: Linguistic	**27 (96.4%)**	1 (3.6%)	0 (0.0%)
3 - Elaboration	**20 (71.4%)**	7 (25.0%)	1 (3.6%)
4 - Imagery	**14 (50.0%)**	13 (46.4%)	1 (3.6%)
5 - Summarization	**21 (75.0%)**	5 (17.9%)	2 (7.1%)
6 - Translation	11 (39.3%)	**12 (42.9%)**	5 (17.9%)
7 - Repetition	10 (35.7%)	**17 (60.7%)**	1 (3.6%)
8 - Keyword	**22 (78.6%)**	5 (17.9%)	1 (3.6%)
Metacognitive Strategies			
9 - Directed Attention: Sentence	4 (14.3%)	**17 (60.7%)**	7 (25.0%)
10 - Directed Attention: Detail	1 (3.6%)	**14 (50.0%)**	13 (46.4%)
11 - Directed Attention: General	**26 (92.9%)**	2 (6.3%)	0 (0.0%)
12 - Selective Attention: Morpheme	**12 (42.9%)**	10 (35.7%)	6 (21.4%)
13 - Selective Attention: Word	10 (35.7%)	**12 (42.9%)**	6 (21.4%)
14 - Selective Attention: Chunk	**23 (82.1%)**	4 (14.3%)	1 (3.6%)
15 - Selective Attention: Structure	11 (39.3%)	**12 (42.9%)**	5 (17.9%)
16 - Monitoring	**14 (50.0%)**	11 (39.3%)	3 (10.7%)

Note. The highest frequency and percentage in each item are bold-faced and underlined. The items refer to the same strategies as those presented to the experimental groups.

Figure 7.6

Responses of the Control Group to the Strategy Survey (Cognitive & Metacognitive Strategies)



7.3.4 Discussion of the Effects of Accelerated Speech Dictation and Shadowing

Regarding the effects of speech perception training, the two measurement tasks yielded mixed results. Specifically, the analyses of the RPD-R indicated significant improvement in all three groups, although the ASD group demonstrated the largest rate increase; the CNT-R revealed significant improvement only for the SH group. To interpret the discrepancy between the results of the two measurement tasks, it was necessary to compare the cognitive processes involved in the two training exercises and analyze how well the respective processes can be measured through the two measurement tasks. To this end, the strategy survey results obtained in Studies 4 and 5 are summarized in Table 7.9. As indicated by the symbols, the RPD-R covers more cognitive processes than the CNT-R, suggesting that it can gauge a wider variety of sub-skills, including those relating to higher-level processing (Items 2, 8, 11, and, 15). In fact, the RPD-R exhibited more overlap with the two training exercises than the CNT-R did. In other words, the RPD-R is a general-purpose measure, so it could detect significant improvement of

both experimental groups. In contrast, the CNT-R involves fewer processes but is focused on speech perception ability. Below, the effects that were expected through each training are discussed in detail.

Table 7.9

Summary of the Strategy Survey Results Obtained in Studies 4 and 5

	Study 4		Study 5	
Items	RPD-R[a]	CNT-R[a]	ASD[b]	SH[b]
1. to guess where the story was going based on what I had understood				
2. to imagine the story based on words and phrases I perceived	●		☆	☆
3. to use my background knowledge related to the topic				
4. to picture the story in my head				
5. to summarize the gist of the story			☆	
6. to translate English into Japanese			★	
7. to shadow the English text (repeated the text in my head)	○	○		★
8. to look for keywords	△		★	
9. to try to understand each sentence precisely			☆	
10. to try to understand the details of each passage				
11. to try to understand the global meaning of each passage	○	○	★	
12. to pay attention to detailed sounds like the plural *-s* and past tense *-ed*	●		☆	☆
13. to pay attention to individual words	○	○	★	☆
14. to try to figure the phrase chunks		●	☆	
15. to pay attention to the grammatical structures	△			
16. to ask myself whether or not my understanding was correct				

Notes. RPD-R = revised written reproduction task; ASD = accelerated speech dictation; CNT-R = revised word count task; SH = shadowing.

[a] The black circle (●) indicates the items whose mean ratings were greater than 3.50 (out of 5) and were significantly larger than the other group's; the white circle (○) indicates the items with a mean rating greater than 3.50 but without significant difference from the other group's; △ indicates the items whose mean rating was less than 3.50 (but greater than 3.00) and significantly larger than the other group's.

[b] The stars (★ and ☆) indicate the items for which the percentage of participants who gave positive responses (i.e., "*agree*" + "*mildly agree*") reached 70%. Items that more than 30% of the participants rated "*agree*" are indicated by ☆, while the items that more than 40% of them rated "*agree*" are indicated by ★.

Effects Expected From Accelerated Speech Dictation. Accelerated speech dictation, which exhibited the greatest improvement in the RPD-R, comprises many more cognitive processes than shadowing. In all likelihood, this is because dictation is a bi-modal task where learners need to process both spoken and written languages. In other words, dictation requires them to perceive spoken words and transform them into an orthographical form. Meanwhile, learners are allowed to analyze the input syntactically and construct meaning representations. In so doing, learners engaged in dictation are expected to acquire various language skills and linguistic knowledge. However, accelerated speech dictation, which uses speedy materials, may be more geared toward enhancing the speech perception ability.

The RPD-R, which is also a dictation task, involves many cognitive processes too. As indicated by Table 7.9, accelerated speech dictation and the RPD-R share several processes. Cognitive processes common to these tasks were Items 2 (to imagine the story based on words and phrases I perceived), 8 (to look for keywords), 11 (to try to understand the global meaning of each passage), 12 (to pay attention to detailed sounds like the plural *-s* and past tense *-ed*), and 13 (to pay attention to individual words). As most of these items are highlighted with the colored stars (★) for accelerated speech dictation, these effects are highly expected from this

exercise.

The greatest improvement of the ASD group suggests that long-term training enhanced their ability to perform the aforementioned processing. In other words, accelerated speech dictation possibly improved the participants' skills of (1) input decoding (Item 12); (2) recognition of spoken words (Item 13); (3) parsing of the speech input to understand propositions (Item 2); and (4) forming meaning and discourse construction (Items 8 and 11).[4] With regard to effect (1), phonemic decoding ability was especially promoted because Item 12 refers to perception of inflectional morphemes. These effects cover all three processing stages in listening comprehension (perception, parsing, and utilization); therefore, the RPD-R suggests the effectiveness of accelerated speech dictation for the improvement of perception ability as well as general listening ability.

In the CNT-R, the ASD group showed almost no improvement (the score increase was only 0.92%). The CNT-R, which does not activate many cognitive processes, shared only three process with accelerated speech dictation: Items 11 (to try to understand the global meaning of each passage), 13 (to pay attention to individual words), and 14 (to try to figure the phrase chunks). The negligible improvement of the ASD group can be attributed to the small number of shared processes between accelerated speech dictation and the CNT-R. This is reasonable because, unlike the RPD-R, the CNT-R does not require overt language production nor the analysis of transcribed words.

Among the aforementioned three items, the processing measured exclusively by the CNT-R is the chunk perception indicated by Item 14 (to try to figure the phrase chunks). The ASD group demonstrated development of this processing, indicating the effect of accelerated

[4] As explained in Chapter 2, *meaning construction* is the application of world knowledge and inference, while *discourse construction* refers to integration of the text comprehension into the ongoing context (see 2.1.2).

speech dictation on holistic sound-processing ability. Interestingly, this processing was not dominant in the RPD-R, which is also a dictation task. This difference was probably derived from the speed of the materials; accelerated speech dictation engaged the participants in reproduction of an entire passage at much faster rates, thereby pushing them to process the speech more efficiently than in the RPD-R. This finding supports the author's assumption made in Chapter 2 that acceleration of speech imposes an extra cognitive load on learners, promoting holistic sound-processing.

The strategy survey results also suggested that the ASD group tended to respond positively to such items as 5 (to summarize the gist of the story), 6 (to translate English into Japanese), and 9 (to try to understand each sentence precisely). These indicate that the participants attempted to (1) understand the proposition of each sentence (Items 6 and 9) and (2) form discourse construction. These are the processing above the perception level; therefore, it can be concluded that accelerated speech dictation promotes higher-level processing.

Effects Expected From Shadowing. Among the three groups, only the SH group made significant improvement in both measurement tasks. Compared to accelerated speech dictation, shadowing shares fewer processes with the RPD-R: Items 2 (to imagine the story based on words and phrases I perceived), 7 (to repeat the text in my head), 12 (to pay attention to detailed sounds like the plural *-s* and past tense *-ed*), and 13 (to pay attention to individual words). It is assumed that because these processes were enhanced by shadowing, the SH group demonstrated significant improvement in the RPD-R. In other words, the significant reproduction rate increase of the SH group in the RPD-R resulted from the improvement of their skills in (1) input decoding (Items 12), (2) recognition of spoken words (Item 13), (3) enhancement of the subvocal rehearsals process (Item 7), and (4) parsing of the speech input to understand propositions (Item 2). Most of these effects are associated with the perception stage

rather than the parsing and utilization stages, indicating that shadowing is effective especially for improving speech perception ability.

As mentioned in Chapter 2, Kadota (2007, 2012) advocates the effect of shadowing on enhancement of the subvocal rehearsal process; hence, this result provides indirect evidence for his claim. To be specific, as illustrated by Figure 2.7 in Chapter 2, Kadota claims that shadowing helps learners to achieve the acceleration of subvocal rehearsal, resulting in the improvement of listening skills as well as the internalization of formulaic sequences (or phrasal knowledge). Most of these effects are associated with speech perception rather than parsing and utilization, indicating that shadowing is effective especially for improving speech perception ability.

Acquisition of the formulaic sequences can be accomplished through development of the ability to process the aural input holistically. This ability is associated with Item 14 (to try to figure the phrase chunks), where less than 30% of the participants in the SH group responded "*agree*" to this item. However, the percentage of those who provided positive feedback ("*agree*" + "*mildly agree*") to the item almost reached 80% (79.3%), suggesting that majority of the SH group felt the improvement of the holistic sound-processing ability. This seems to echo the finding of O'ki (2012b) that learners exploited their phrasal knowledge to shadow English passages.

Unlike in the RPD-R, significant score improvement was demonstrated only by the SH group in the CNT-R. This result is unexpected given that accelerated speech dictation shares slightly more cognitive processes with the CNT-R (Items 11, 13, and 14) than does shadowing (Items 7 and 13). In other words, the CNT-R should also have detected the improvement of the ASD group. The reason may be twofold. First, unlike the RPD-R, the CNT-R does not require participants to write down words (but only to count them in their minds); thus, the test materials were not kept in their memory until the post-test. Second, the CNT-R is not as cognitively demanding as the RPD-R, so that lower-level learners can experience the practice effect more

easily. In sum, participants can complete the CNT-R without much effort, and even lower-level learners can easily demonstrate higher performance on the post-test. For this reason, the CNT-R may be more appropriate for evaluating the performance of lower-level learners than that of upper-level learners. In fact, the SH group demonstrated the lowest performance in all five tests (i.e., the proficiency test and the pre- and post-tests of the two measurement tasks), indicating that they had the poorest listening ability. To confirm this speculation, an item analysis of the CNT-R based on Item Response Theory (IRT) needs to be conducted in the future research.

Interpretation of the Control Group Improvement in the RPD-R. Besides the insignificant improvement of the ASD group in the CNT-R, the analysis revealed another unexpected result: the significant improvement of the control group on the RPD-R. This is probably a practice effect caused by using the same materials in both the pre- and post-tests, even though there was a two-month gap between the tests. In written reproduction tasks (or dictation tasks), practice effects can emerge because the learner's memory of the words heard before can be reinforced by transcribing them on paper.

Furthermore, as discussed in the previous section, the participants tended to utilize semantic and grammatical processing in the RPD-R. According to Craik and Lockhart (1972) and Craik and Tulving (1975), attention to meaning allows deeper language processing and thus contributes to the retention of linguistic information. These are less likely to happen in the CNT-R because learners do not have a chance to process the target words visually during the pre-test, making it difficult to utilize higher-level processing and maintain the linguistic information in memory for two months. For the same reasons stated above, accelerated speech dictation is expected to increase spelling knowledge and enhance the recognition skill of written words as well as spoken words.

The practice effect can also be strengthened by greater proficiency of the control group.

The pre-test revealed that the control group possessed significantly better listening ability than the two experimental groups (see Table 7.4). This possibly indicates that they also had better "learning" ability, so that they could memorize the test materials in the pre-test and recall them in the post-test.

7.4 Summary of the Findings: Chapter 7 (Study 5-1)

7.4.1 Answer to RQ5-1

RQ5-1: Will long-term training using accelerated speech dictation and shadowing improve learners' speech perception ability?

A series of analyses on the pre- and post-tests revealed that the two kinds of speech perception training yielded different results as per the measurement task. Specifically, those who were engaged in accelerated speech dictation demonstrated the greatest reproduction rate improvement in the RPD-R, while only those who trained through shadowing showed a significant score increase in the CNT-R. The strategy survey results obtained in Studies 4 and 5 were compared to interpret these results, and the analysis revealed what effects can be expected from the two speech perception exercises, as summarized in Table 7.10. Most of the effects are associated with speech perception ability.

As can be seen in the table, almost identical effects can be expected from the two training exercises. However, accelerated speech dictation is considered to facilitate a wider variety of processing because it is bi-modal. Specifically, while jotting down the words they hear, learners become attentive to semantic information and word forms as well as the phonological information of the training materials. Therefore, accelerated speech dictation may be effective for general listening ability as well as speech perception ability. In contrast, shadowing is characterized especially by its effect on enhancement of the articulatory rehearsal process in the working memory, which supports the claim made by Kadota (2007, 2012).

Table 7.10

The Effects of Accelerated Speech Dictation and Shadowing

Types of Training	Skills Expected to Improve
Accelerated Speech Dictation	(1) input decoding (phonemic decoding, holistic sound-processing) (2) recognition of spoken words as well as written words (3) parsing of the speech input to understand propositions (4) forming meaning and discourse construction
Shadowing	(1) input decoding (phonemic decoding, articulatory rehearsal, holistic sound-processing) (2) recognition of spoken words (3) parsing of the speech input to understand propositions (4) forming meaning and discourse construction

Note. The effects unique to each training are underlined.

7.4.2 Answer to RQ5-2

RQ5-2: Will the training develop learners' strategy use in listening?

Irrespective of the strategy categories (i.e., cognitive and metacognitive), similar results were obtained for the two experimental groups, with a few differences. Concerning the eight cognitive strategies, most of which were associated with top-down processing in listening, the majority of the two groups recognized the development of all the strategies. The development took place because the participants improved their skills for bottom-up listening through long-term speech perception training, thus, they acquired a better command of top-down listening strategies. However, differences between the groups lay in the strategies that referred to translation of the passage, extraction of keywords, and subvocal rehearsal of the input. The ASD group responded more positively to the first two strategies, while the SH group agreed that the last strategy had improved the most. These results corroborate the previous assumptions that dictation can stimulate the use of meaning processing while shadowing is efficacious for the enhancement of subvocal rehearsal.

Concerning the eight metacognitive strategies, nearly all participants of both groups

acknowledged the development of the three strategies employed in the perception of various linguistic components (i.e., perception of morphemes, words, and chunks). These results provide additional evidence in support of the impact of both speech perception training on the development of speech perception ability. It was noticeable, especially for the ASD group, that more participants acknowledged the development of the strategies to understand the passage's meaning. Given the findings stated earlier, it can be concluded that dictation activities, including accelerated speech dictation, can sensitize learners to passage meaning and, thus, foster wider skills and knowledge.

7.4.3 Study Limitations

This study has two limitations. First, the pre-tests revealed significant differences in the speech perception ability of the three groups. Especially, the control group demonstrated significantly higher performance in the proficiency test than the experimental groups. Moreover, their speech perception ability, measured by the two measurement tasks, surpassed the other two groups. As studies by Tamai (2005) and Suzuki (2007) suggested, the effects of shadowing and dictation can differ depending on the learner's proficiency level, i.e., lower-level learners are more likely to benefit from the training. Therefore, the level of the participants should be equal between the groups.

Second, the two measurement tasks developed for this study (the RPD-R and the CNT-R) primarily focused on accuracy in speech perception. However, the degree to which learners can process the input efficiently is also important in listening because of the transient and elusive nature of aural language. This study indicated that those who trained through shadowing enhanced their ability to perform subvocal rehearsal. Enhancement of working memory function can lead learners to become efficient in speech perception. For these reasons, accuracy and efficiency in speech perception should be investigated to evaluate the training effect, using

a different approach focused on the reaction time of word recognition.

Chapter 8

Study 5-2: Long-Term Training Effects of Accelerated Speech Dictation and Shadowing on Speech Perception Ability (Qualitative Analyses)

8.1 Study Goal

The previous chapter reported whether the participants' speech perception ability had improved as a result of the two-month training through either accelerated speech dictation or shadowing by referring to their performances on the two measurement tasks (i.e., the RPD-R and CNT-R). Statistical analyses yielded an interactive effect between the training type and the measurement task. Specifically, the ASD group demonstrated the best improvement on the RPD-R, while only the SH group showed significant improvement on the CNT-R. It was speculated that the mixed results could be attributed to the similarity between the cognitive processes involved in the training and those involved in the measurement task. Namely, attention toward word forms and passage meaning is likely to be raised during accelerated speech dictation and the RPD-R, while subvocal rehearsal (i.e., unvocalized repetition of received input) tends to be active in shadowing and the CNT-R. Some evidence for this assumption was obtained from the participants' responses to the listening strategy survey. Accordingly, this chapter reports the results of text analyses with regard to their weekly journals and the questionnaire to find qualitative evidence for the previous assumption.

Based on these backgrounds, this chapter presents an investigation of the following RQs:

RQ6-1: What kinds of effects did the participants feel throughout the training?

RQ6-2: What kinds of cognitive processes were involved in accelerated speech dictation and shadowing?

8.2 Method

The results reported in this chapter are some of the outcomes obtained by Study 5; hence, the method is the same as described in the previous chapter. To obtain qualitative evidence of the training effect, text analyses using KH Coder (https://khcoder.net/) were carried out on the weekly journals and the participants' responses to a free writing item of the questionnaire, in which they reflected on the whole training period. For each experimental group, three kinds of analyses were conducted through the software. First, to figure out what words or expressions were frequently used, a list of the most frequent 40 words was produced. Second, to extract concepts or ideas from their responses, a co-occurrence network of the most frequent 60 words was computed. As Fujii, Kosugi, and Lee (2005) pointed out, text analyses based on computation require careful inspection of the original texts because the meanings of words drawn out by computation cannot be interpreted without the contexts in which they are embedded. For this reason, some of the participants' responses written in Japanese are quoted as examples in footnotes. Third, to investigate how the usage of words changed in the course of training, a correspondence analysis plot was created for the relationship between the frequent words and the training period (i.e., 2nd through 5th weeks vs. 6th through 10th weeks).

8.3 Results and Discussion

8.3.1 Text Analyses of the Responses by the ASD Group

As a result of the study, a total of 288 responses were obtained from the ASD group. The text mining using KH Coder extracted 10,684 tokens and 940 types (see Appendix 8A for the most frequent 150 words),[1] out of which 4,242 tokens and 733 types were left after

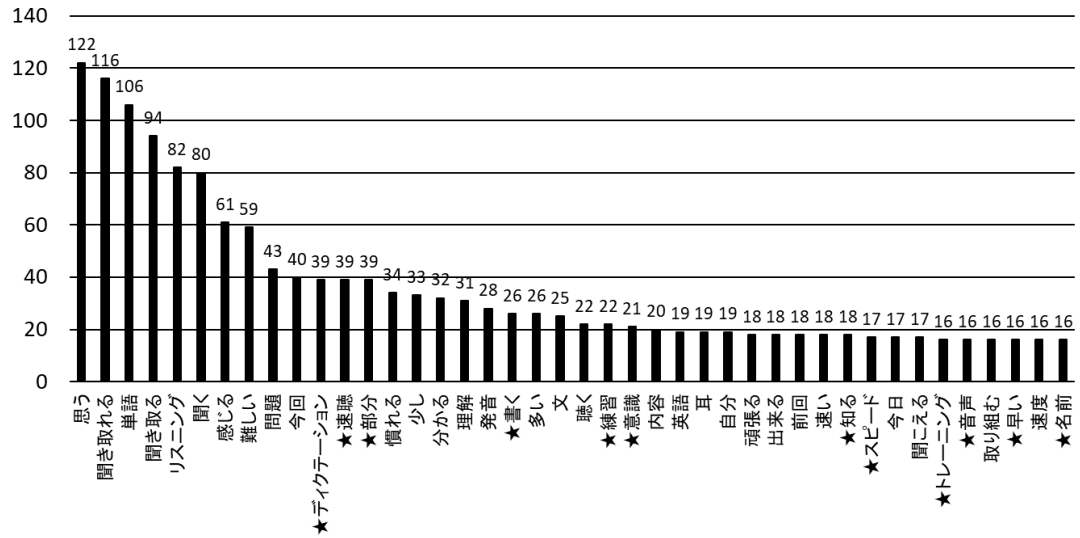
[1] KH Coder recognized "速聴" ("accelerated speech listening") as two separate words (速 and 聴); thus, the word was inputted into the forced pick-up so that it was extracted as one word.

excluding several parts of speech that carried little meaning such as Japanese particles (e.g., “は” and “の”) and then used for the subsequent analyses.

Figure 8.1 illustrates the 40 words that appeared most frequently in the participants’ responses. The words were ranked from second to sixth including “聞き取れる” (“to be able to perceive”), “単語” (“word”), “聞き取る” (“to perceive”), “リスニング” (“listening”), and “聞く” (“to listen”), which are related to perception of words, indicating that a number of responses had something to do with word perception. Moreover, there was “難しい” (“difficult”), “分かる” (“to understand”), “慣れる” (“to become familiar”), “頑張る” (“to make efforts”), “出来る” (“to be able to do something”) and “取り組む” (“to be engaged”). These words imply that, although the participants first felt the training to be difficult, they gradually got used to it and became able to understand the English speech better as their training proceeded. There are also words mentioning the speed of the passage such as “速聴” (“accelerated speech listening”), “速い/早い” (“fast”), “速度/スピード” (“speed”), suggesting that the participants had difficulty in catching up with the speed of the input.

Figure 8.1

Top 40 Words Observed in the Responses by the ASD Group



Note. Words with a star mark were not listed in the top 41 words of the SH group.

To examine whether these assumptions were correct, a co-occurrence network among the most frequent 60 words was computed while looking for some responses that represent assumed concepts. As illustrated by Figure 8.2, several clusters were extracted. The cluster A, which consists of 11 words with large bubbles, seems to suggest that quite a few participants found it difficult or impossible to perceive words during the training or that there were many parts (“部分”) they could not hear. The same concept can also be found in the cluster B, which consists of “聞き取る” (“to perceive”), “難しい” (“difficult”), and “感じる” (“to feel”). The difficulty was partly due to the word spellings;[2] however, the main cause was clearly the passage speed. As indicated by the inner circle in cluster C, where such words as “速い” (“fast”), “速度” (“rate”), “スピード” (“speed”), “通常” (“usual”), and “音声” (“sounds”) co-occurred, many participants found the training materials to be unusually fast. Examination of all the responses revealed that 21 responses referred to the excessive speech rate. For example, a participant confessed on the first date (i.e., 2nd week) that he could not write anything at all because the passages were too fast for him.[3]

As mentioned repeatedly in this dissertation, acceleration of passage speed turns on the holistic-processing system, preventing learners from analytically processing the aural input. Subsequently, it was revealed that a total of 40 responses referred to the difficulty not only in perceiving sounds that did not stand out in the sentences such as prepositions and contracted forms but also in segmenting a set of words that sounded like a chunk.[4] Notably, spoken

---

[2] 「聞き取れても書き取りがなかなか上手くいかないと感じた。(3rd week)」「聞き取れてもスペルがわからないことが多かったのでもっと単語を勉強したいと思った。(3rd week)」

[3] 「速すぎて何を言っているのか分からず全然書けなかった。(2nd week)」

[4] 「音声が速くて書き取りがついていけず、難しかった。前置詞が特に聞き取れなかった。(2nd week)」 「I've などの短縮した表現の聞き取りが難しい。(2nd week)」 「1.3 倍速で聞き取るのは大変だった。部分部分では聞き取れるが繋がって発音してるような所は聞き取りづらかった。(2nd week)」

English is characterized by sound change phenomena such as weak forms (e.g., *them* /em/); elision (e.g., *that child* /ðætʃaild/); assimilation (e.g., *this year* /ðiʃiər/); and linking (e.g., *in an hour* /inənauər/) (Kayatama, Nagase, & Joto, 1996). These characteristics cause difficulty in word segmentation, which is a major problem in speech perception according to Anderson (2005). Hence, learners must be familiar with how words are actually pronounced in connected speech. However, learners who have had little exposure to aural English often end up having wrong images about word pronunciations. Speech perception can be very troublesome for these learners; for instance, some participants mentioned in their journal they could not figure out what was being said because the pronunciation was different from what they had thought.[5] In this sense, the study uncovered the role of dictation to make learners realize the discrepancy between their image about word pronunciations and the actual pronunciation. This is also evident from the cluster D, which consists of "自分" ("myself"), "発音" ("pronunciation"), "違う" ("different"), and "分かる" ("to understand").

In addition to pronunciations, the participants became aware of the lack of their knowledge about word spellings through the training. This idea can be observed from the cluster B, which has a word combination with "スペル" ("spelling"), "名前" ("name"), and "難しい" ("difficult"). These words were extracted from responses referring to the difficulty in spelling out names of people.[6] There were also some cases where participants could not jot down some words even though they understood them.[7] These responses revealed that this experience

---

[5] 「何度も聞き返しても、分からないものもあったので、自分が思ってる発音の仕方と違うのかなと思いました。(7th week)」 「would like to と have asked for の聞き取りが難しかったです。思っていた発音と少し違いました。(9th week)」

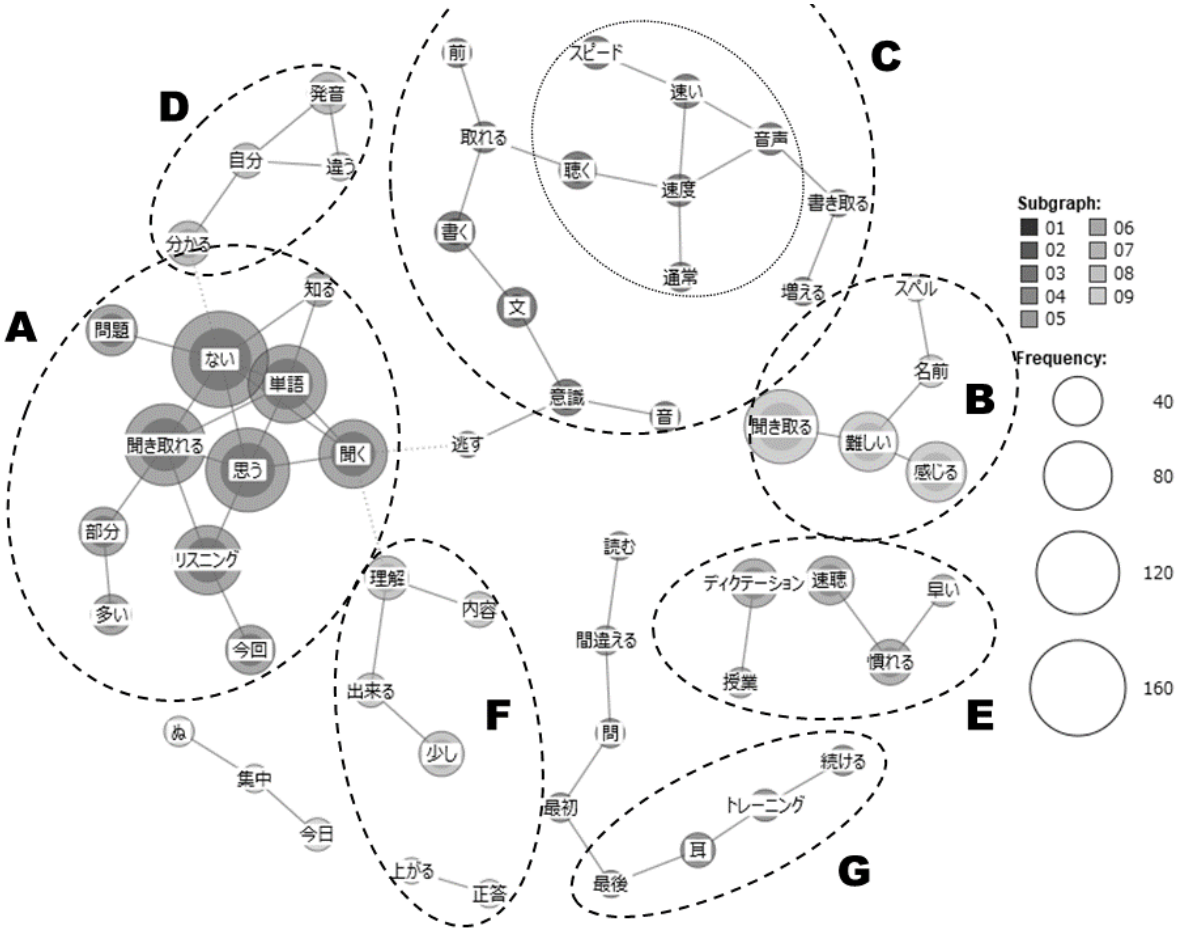
[6] 「人の名前のスペルや be 動詞が濁って聞こえたところが難しかったです。(4th week)」

[7] 「聴こえても綴りがわからなかったり、意味がわからなかったりすると文全体の内容を把握すること難しいため、語彙力を高める必要があることを実感した。(2nd week)」「(前略) 聞き取れてもスペルがわからないことが多かったのでもっと単語を勉強したいと思った。(3rd week)」

enhanced their motivation toward learning about word spellings. Moreover, some participants wished to reduce the number of spelling errors.[8] As such, these responses suggest the role of accelerated speech dictation to focus learners' attention on word spellings. In the previous chapter, it was speculated that the reason for the ASD group demonstrating the greatest improvement in the RPD-R was that they enhanced their sensitivity toward word spellings through the training. This assumption has been supported by the responses presented above.

Figure 8.2

Co-occurrence Network of the Most Frequent 60 Words (ASD Group)



[8] 「書き取れる問題数が増えてきました。ですがはやく書こうとしすぎてスペルミスが目立つようになりました。(7th week)」「(前略) スペルミスが多く、特に簡単な単語での間違えが多かったので意識して学習したいです。(後略) (7th week)」

As the result of the training, many participants improved their training performance. The cluster C, which has the combination of “書き取る” (“to write down”) and “増える” (“to increase”), seems to mean that the participants became able to transcribe more words successfully as the training proceeded. Similarly, the network consisting of “文” (“sentences”), “書く and 取れる” (“to write down”), and “前” (“before”) indicates that the participants became able to take dictation of more sentences than before. In fact, several participants reported that their dictation performance gradually improved.[9] Examination of all the responses revealed that 146 responses were related to the improvement of speech perception ability or to greater familiarity with the training.[10] The latter is backed up by the cluster E, which includes “速聴” (“accelerated speech listening”), “ディクテーション” (“dictation”), “早い” (“fast”), and “慣れる” (“to become familiar”). Out of the 146 responses, 99 of them were observed during the later period of training (i.e., 6th to 10th weeks). Therefore, it can be concluded that the participants recognized the growth of their speech perception ability as a consequence of the long-term training.

Furthermore, some participants felt that their comprehension ability improved as well. This is observed in the cluster F, where such words as “内容” (“content”), “出来る” (“to be able to do”), “理解” (“to understand”), and “少し” (“a little”) constitute a concept meaning that they became able to understand the content a little (better). The improvement of comprehension ability can be explained by two reasons, one of which is the improvement of their speech

---

[9] 「前回よりは長い文を続けて書き取れるようになりました。(4th week)」 「文字を早く書くコツがわかってきました。なので書き取れる問題数が増えてきました。(6th week)」 「初回はとても速くて聞くことで精一杯でしたが、だんだん書き取れる問題が増えてきました。(10th week)」

[10] 「前回より聞き取りやすくなっている気がします。リスニングも高校の時と比べてぜんぜんできているので楽しくなってきました。(3rd week)」 「リスニングの正答率も上がり、少しずつ聞き取れるようになってきたので、この調子で頑張りたいです。(4th week)」

perception ability. The last chapter revealed that the ASD group showed the greatest improvement in the RPD-R. As Kadota (2007) advocates, the better perception ability learners have, the more they can focus on passage meaning. This is corroborated by the responses of several participants who mentioned that they had felt more comfortable with listening to the materials at the original speed as they were exposed to the accelerated speech.[11] The other reason is that their awareness toward listening strategies was raised. For example, a participant referred to the importance of inferencing the passage meaning based on the words they had understood.[12] Another participant reported that verbs are crucial for understanding the general passage meaning.[13] Moreover, a participant realized that proper nouns such as store names had impeded her meaning processing but that she became able to understand them based on the context.[14] These responses clearly show that the participants reflected on their cognition during the activity, although they were not given instructions about these listening strategies during the course. Therefore, development in listening strategy use is crucial for learners to be advanced listeners, as Vandergrift and Goh (2012) put it: “Strategies help them improve comprehension, retention, and recall of information; and, at the same time, they assist in planning for overall listening development as part of their language learning effort” (p. 89). With this claim in mind, it is speculated that accelerated speech dictation can not only develop

---

[11] 「速聴を 2 回聞いたあと、通常のものを聞いてみたら、とてもゆっくりに感じ、内容をよく理解することができたと思う。(4th week)」 「早い速度の音声に慣れることで、実際の音源がゆっくりに感じて、落ち着いて聴けるようになった気がします。正答率も少し上がりました。(9th week)」

[12] 「以前よりも聞き取れる単語が増えたが、聞き取れなかった単語をひきずってしまって、聞き逃していたことがあったので、次回からは、聞き取れなくても切り替えて、会話の全体を聞いてなんの話をしているのか推測できるようにしたいです。(3rd week)」

[13] 「(前略) 分からない単語も多々あったが動詞の意味さえ分かれば大体の内容を把握することができることがわかった。(4th week)」

[14] 「以前は、知らないお店の名前とかホテルの名前とかで、建物の名前と気づかないことがあって、問題がわからなかったですが、会話の流れでわかるようになりました。(6th week)」

learners' strategy use but also aid the sustainable development of their listening ability.

Finally, it is also observed that the participants who recognized the effect of the training using accelerated speech dictation were willing to continue the training. This assumption is derived from the cluster G, which consists of such words as "最後" ("the end"), "耳" ("ears"), "トレーニング" ("training"), and "続ける" ("to continue"). In fact, quite a few participants mentioned that they would like to continue the training to improve their "ears."[15] It is noteworthy that some of them expressed their enthusiasm to study outside class or after the end of the training, meaning that they recognized the importance of extended study for improving their listening ability as well as the effect of training using accelerated speech dictation.

In sum, the analyses yielded the following findings: (1) a number of participants first felt the training to be very difficult, especially with regard to the passage speed, making it difficult to find word boundaries; (2) however, they gradually overcame the difficulty and were able to reproduce more words as the training proceeded, while being aware of the paucity of their knowledge about pronunciation and spelling; (3) they also improved their comprehension ability as a byproduct of the development in their speech perception ability and strategy use; and finally (4) they wished to continue the training outside class or even after completion of the training.

Figure 8.3 shows the result of the correspondence analysis and illustrates how the word usage in participants' responses changed over time. As it indicates, during the earlier period of

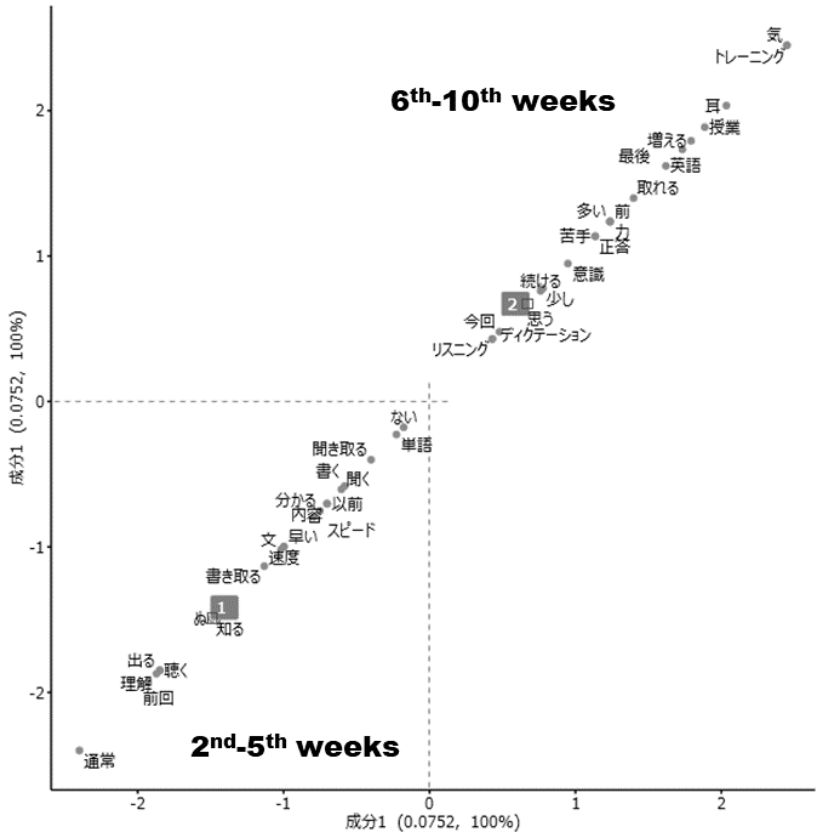
---

[15] 「初めに聞いた際に、早くて聞き取れなかったところも、このトレーニングで耳慣らしをすることによって、聞き取りやすくなりました。初めて聞くリスニングにおいても、しっかりと聞き取ることができるように、たくさんトレーニングを積み重ねていきたいと思います。(8th week)」「段々と耳がなれ、130%の音声でも聞けるようになってきました。100%の音声を聞いた時に、しっかりと聞き取ることができるよう、これからもトレーニングを続けていきたいと思います。(9th week)」 「いつでも聞き取りやすくするためには、単語の理解だけでなく、耳のトレーニングを続けていく必要があると感じた。今後も続けて練習したいと思う。(10th week)」

training (i.e., 2nd to 5th weeks), words that refer to the fast speech rate—in other words, “速度” (“rate”), “スピード” (“speed”) and “早い” (“fast”)—can be observed. In addition, verbs like “聞く/聞き取る” (“to listen”), “書く/書き取る” (“to write down”), “知る” (“to know”), and “分かる” (“to understand”) appeared near negative particles such as “ない/ぬ” (“not”). However, the later period (i.e., 6th to 10th weeks) can be characterized by more positive words such as “続ける” (“to continue”), “増える” (“to increase”), “多い” (“many”), and “取れる” (“to catch”). These can be thought of as supplementary evidence for the claims stated so far.

Figure 8.3

Correspondence Analysis Plots of the Most Frequent 40 Words Used by the ASD Group in Relation to the Training Period (2nd to 5th weeks vs. 6th to 10th weeks)

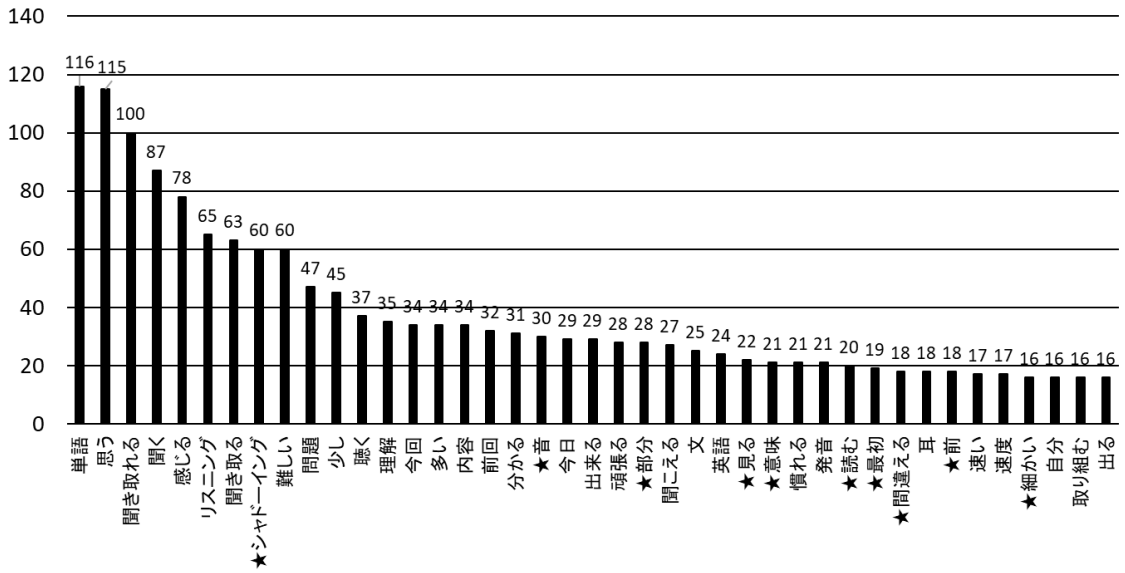


8.3.2 Text Analyses of the Responses by the SH Group

As a result of the data collection, a total of 346 responses were obtained from the SH group. The text mining using KH Coder extracted 11,307 tokens and 1,009 types (see Appendix 8B for the most frequent 150 words), out of which 4,563 tokens and 798 types were used for the subsequent analyses after excluding several parts of speech that carried little meaning. Figure 8.4 shows the 41 words that appeared most frequently in responses of the SH group. Except the 10 words with the star mark, the other 31 words overlapped with those in the list of the ASD group. Remarkably, 9 out of the top 10 words were identical between the two groups, implying that both groups underwent similar experiences through the training.

Figure 8.4

Top 41 Words Observed in the Responses by the SH Group



Note. Words with a star mark were not listed in the top 40 words of the ASD group.

Figure 8.5 shows a co-occurrence network of the 60 words that were most frequently used by the SH group. As the circles drawn by the author indicate, five clusters were observed. The cluster A, which consists of 18 words that are almost identical with those in the clusters A

and B of Figure 8.2, indicates that quite a few participants had difficulty perceiving words. Examination of all the responses revealed that 77 of them were related to the perception difficulty, primarily because the participants failed to find word boundaries.[16] This result corresponds to the discussion held in the previous section that accelerated speech dictation, which is a cognitively demanding activity for learners and can cause difficulty in word segmentation because fast speech rates compel learners to holistically process the aural input. In fact, this is probably what happened to the participants engaged in shadowing too. Specifically, shadowing is also such a difficult activity in which learners must simultaneously listen and speak that the participants could not focus on the input in detail. The result corroborates a finding of O'ki (2012b) that learners' phrasal knowledge had a crucial impact on their shadowing performance in that learners need to recognize and reproduce the incoming speech as efficiently as possible. Based on this finding, O'ki hypothesized that shadowing can increase phrasal knowledge. His claim seems to be supported by a participant's response that shadowing enhanced her awareness toward the word chunks.[17]

For learners to detect word boundaries, they must know how words are pronounced when they appear in sentences. Some responses obtained from this study showed that shadowing helped the participants update their pronunciation knowledge. For example, a participant wrote that she was surprised to know the pronunciation of a word was very different from what she had known.[18] Another participant admitted that she mistook an unfamiliar word

---

[16] 「単語の語尾が次の単語と繋がっているようなところが聞き取りづらかった。(2nd week)」「簡単な単語でも聞き取るのが難しかったです。つなげて発音されるところも多いのでそれを聞き取れるように頑張りたいです。(3rd week)」 「やはりというか、is や a などの短い単語が繋げられてより聞き取り辛くなっているのが聞き取りでの特に難しく感じる所である。(4th week)」「音が繋がっていたりほとんど発音していなかったりするところがまだ聞き取れないのでそれらに注意して取り組んでいきたいと思います。(7th week)」

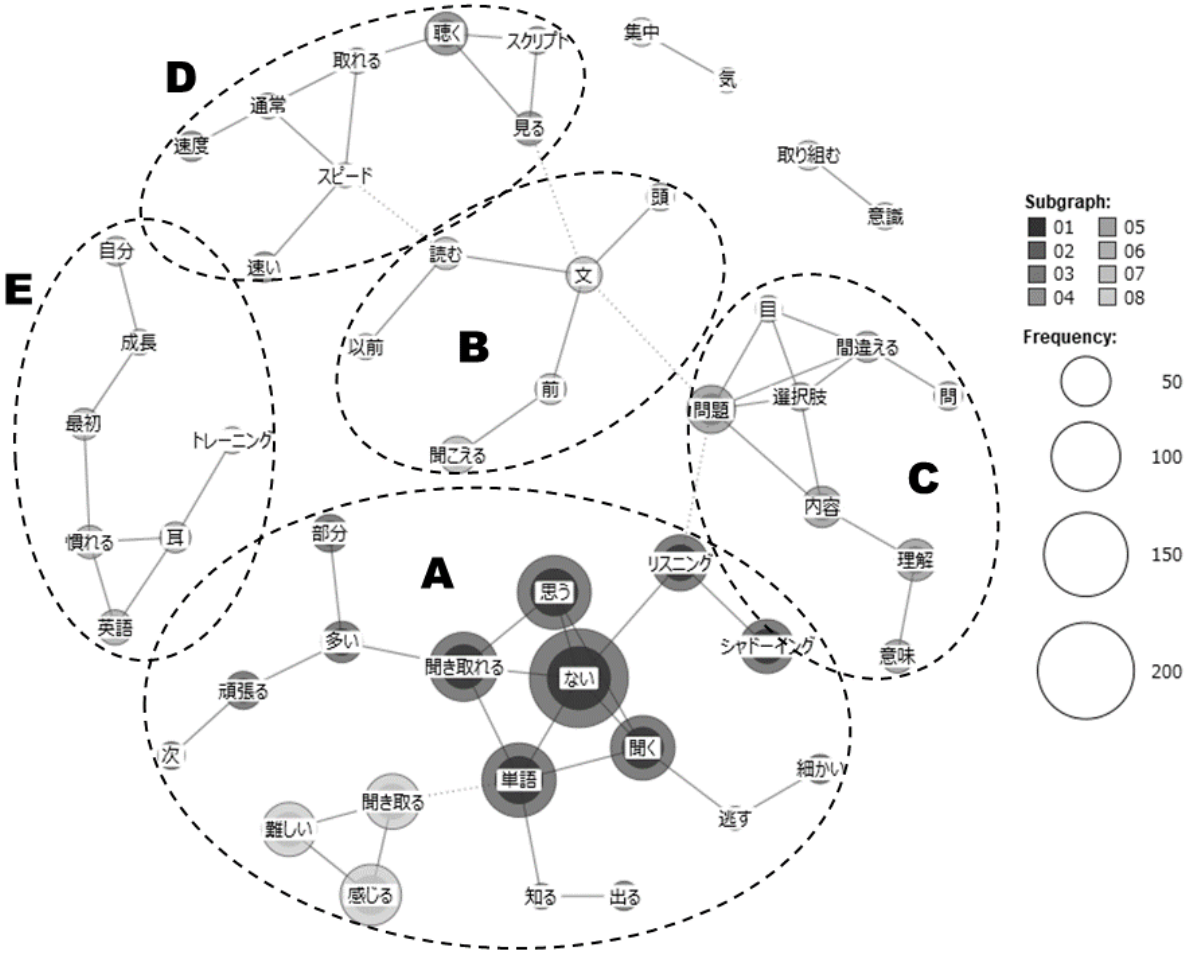
[17] 「単語のつながり、まとまりを意識する力が少し付いた気がします。(10th week)」

[18] 「(前略) この単語はこういう風に発音するんだ！という発見もありました。(4th week)」

for a more familiar one with similar pronunciation.[19] There was also a participant who reported that repetition of the materials in mind had enabled him to focus on the sounds in detail.[20] These responses correspond to the claim of Kadota (2007, 2012) that learners can develop an adequate phonological representation through shadowing by repeating the input over and over.

Figure 8.5

Co-occurrence Network of the Most Frequent 60 Words (SH Group)



[19] 「あまり使わない単語を聞くと、聞いたことのある似ている単語と勘違いをしてしまいそうになった。(3rd week)」

[20] 「何回も何回も音を聞き、頭の中で繰り返すことによって、より細かい部分にまで注意深く聞くことが出来るようになったと感じた。(10th week)」

Kadota (2007, 2012) and Tamai (2005) refer to another effect of shadowing involving repetition of the input which enhances the efficiency of a working memory function called subvocal rehearsal. Subvocal rehearsal—which refers to the silent repetition of language input—is believed to promote the retention of language information (e.g., Baddeley, 1999; Kawasaki, 2005; Osaka, 2002). The cluster B, which includes such words as "頭" ("head or mind") and "文" ("sentence"), shows that the participants were actually repeating the sentences in their minds.[21] Moreover, such words as "以前" ("before") and "読む" ("read") indicate that the participants became more competent than before at catching up to the speed of materials read by the computer.[22] This is probably why only the SH group showed significant improvement on the WCT in the previous chapter. In fact, a participant straightforwardly expressed that shadowing enabled him to count words better.[23] In spite of these effects, it should be kept in mind that repetition of the input may be challenging for some learners because several participants had trouble with articulation of the phonological input.[24] As such, this could happen even when learners recognize what has been said because, as a participant described,[25] shadowing is a dual task in which learners must perform listening and speaking in parallel.

The cluster C, which consists of eight words, is associated with the way that the

---

[21] 「聴き取る時に、頭の中で問題文や例文をリピートして、情報を整理しながら取り組むようになった。(7th week)」 「シャドーイングを行ったことで、聞いたことを言うため、頭に残りやすく感じた。(10th week)」 「何回も何回も音を聞き、頭の中で繰り返すことによって、より細かい部分にまで注意深く聞くことが出来るようになったと感じた。(10th week)」

[22] 「以前よりも問題を読むスピードについていけるようになった気がします。(5th week)」 「以前よりコンピュータの読むスピードについていけるようになった気がします。(6th week)」

[23] 「シャドーイングの活動に取り組んで、単語の語数が分かるようになった気がする。」

[24] 「通常速度でも思ったより口が回らなかった。(2nd week)」 「初めは早く話せなかったり大変でした。次回はもう少し上手く話せるようにしたいです。(3rd week)」

[25] 「通常のスピードでも速いと感じてしまった。聴くと話すを同時にすることが難しいと感じた。(2nd week)」

participants worked on the comprehension questions. Such words as “問題” (“question”), “選択肢” (“answer option”), “目” (“eye”), and “間違える” (“to make a mistake”) indicate that some participants selected wrong answers because they could not look over the answer options before listening to the materials.[26] In contrast, there were also other participants who were able to understand (“理解”) the content (“内容”) or passage meaning (“意味”) because they could afford to read the answer options in advance.[27] Responses like these show that the participants became aware of usefulness of the strategy or that they became capable of using the strategy. Their development can be explained by two reasons. First, throughout the course, the participants were repeatedly advised to preview the answer options before listening to each passage in order so they could guess its meaning. The feedback proved that the instruction exhibited an effect. Second, the improvement of their speech perception ability allowed them time to preview the answer options. In fact, some participants referred to this in their journal.[28]

The cluster D also showed the development of a strategy use to exploit the written script. Such words as “スクリプト” (“script”), “見る” (“to look”), “聴く” (“to listen”), and “取れる” (“to catch”) indicate that the participants were able to recognize words by looking at the script. Actually, some participants mentioned that the scripts were useful when they failed to find word boundaries,[29] while another participant reported that the script made him aware that he had missed a few words.[30] Interestingly, other participants developed their own learning

---

[26] 「選択肢に目を通す時間が足りなかったので間違えた問題があった。(3rd week)」

[27] 「選択肢を事前に目を通しておくことで内容の理解が早かった。(6th week)」

[28] 「聞き取りに余裕がある分回答にも余裕があったため、次の問題の選択肢にも目を通すことが出来て、理想的な形で取り組めました。(4th week)」「前よりも選択肢に目を通す余裕が出来たと思う。はじめより単語がスラスラ入ってきている感じがした。(9th week)」

[29] 「前回よりも音のつながっているところがよく聞けた。聞けなかったところはスクリプトを見てもう一度聴いたり聞けるようになるように努めた。(2nd week)」 「英語を続けて早く言われると聞き取れないところもあったけど、そういう所はスクリプトを見て確認しました。(4th week)」

[30] 「内容はおおむね理解できるが、スクリプトを見た時に、意外と聞き取れなかった単語や未知

strategies. For example, a participant acknowledged the effectiveness of overlapping, an activity in which learners read the script aloud in parallel with the aural input.[31] Although the author introduced this activity to the students at the beginning of the course, it was their own judgement whether and when to do overlapping.

Additionally, learning strategies related to adjustment of the speech rate were used by many participants as well. For instance, a participant reported that reduction of speech rate allowed her to shadow the input better.[32] Unexpectedly, several participants attempted to use the accelerated materials prepared for the ASD group, thereby making them feel the original speed was slow.[33] Furthermore, a participant reported that an alternate use of the accelerated materials as well as the original ones allowed her to perceive every word without the script.[34] Accordingly, participants' feedback about the effect of using accelerated materials are reflected in the word network composed of "通常" ("ordinary"), "速度" ("rate"), "スピード" ("speed"), and "速い" ("fast"). Subsequently, an examination of all the responses revealed that 14 responses were associated with deployment of this learning strategy. As explained in the previous chapter, the participants were allowed to adjust the speech rate at any level—as long as it was not too easy for them. In other words, the participants spontaneously selected the speeds. Therefore, for the sake of an effective speech perception training through shadowing, it is necessary for teachers to give their learners the right to choose or manipulate the speech rate.

語が少しだけあるので、それらを徹底的に潰していきたいと思います。(4th week)」

31 「シャドーイングをしてからスクリプトをみてオーバーラッピングをするという一連の動作がとても有効だと思った。(2nd week)」

32 「0.75 倍速でシャドーイングをしたら、大体話せたと思う。あと、一回スクリプトを読んだ後は、文の意味を理解しながらできた。(3rd week)」

33 「1.3 倍を数回聴いたのちに、通常の速度で聞いてみると、非常に遅く感じた。(3rd week)」
「1.3 倍速で聴くことにより、(中略) 原形の音を聴いた時に、遅く感じることができているので、少しずつですが成長ができていると感じた。(6th week)」

34 「倍速で読んで標準で読むの繰り返しをするとスクリプトなくても細かい単語まで聴き取れるようになった。(5th week)」

So, because of the training using these strategies, the participants gradually became accustomed to the training and developed their speech perception ability. This is evident from cluster E, which includes such words as “耳” (“ear”), “慣れる” (“to become familiar”), and “成長” (“development”). For example, several participants referred to their growth at the early stage of the training (i.e., 2nd to 5th weeks);[35] however, descriptions like these seem to be observed more frequently at the later stage of the training (i.e., 6th to 10th weeks).[36] To confirm this, all the responses were examined. It was then revealed that a total of 152 responses were concerned with the participants’ development, while two thirds of them (i.e., 101 responses) were obtained during the last five weeks.

This is also discernible in the correspondence analysis plot (see Figure 8.6). The early stage of the training is characterized by such words as “難しい” (“difficult”), “スクリプト” (“script”), and “復習” (“review”), indicating that the participants evaluated the usefulness of checking the scripts and reviewing unknown vocabulary to overcome the difficulty of shadowing.[37] In contrast, some words that are distinctive for the later stage are “成長” (“growth”) and “出来る” (“to be able to do something”). Moreover, such words as “意識” (“consciousness”) and “取り組む” (“to work on”) show what the participants were focused on

---

[35] 「リスニングで間違えたところを多くシャドーイングするようにしました。最初は難しかったけど少しずつ慣れることができました。(2nd week)」「簡単な単語であれば通常のスピードで聴き取れるようになった。最初よりも成長していると自分で実感している。(4th week)」 「聞き取れない単語も沢山あるが、成長していると感じる。(後略) (5th week)」

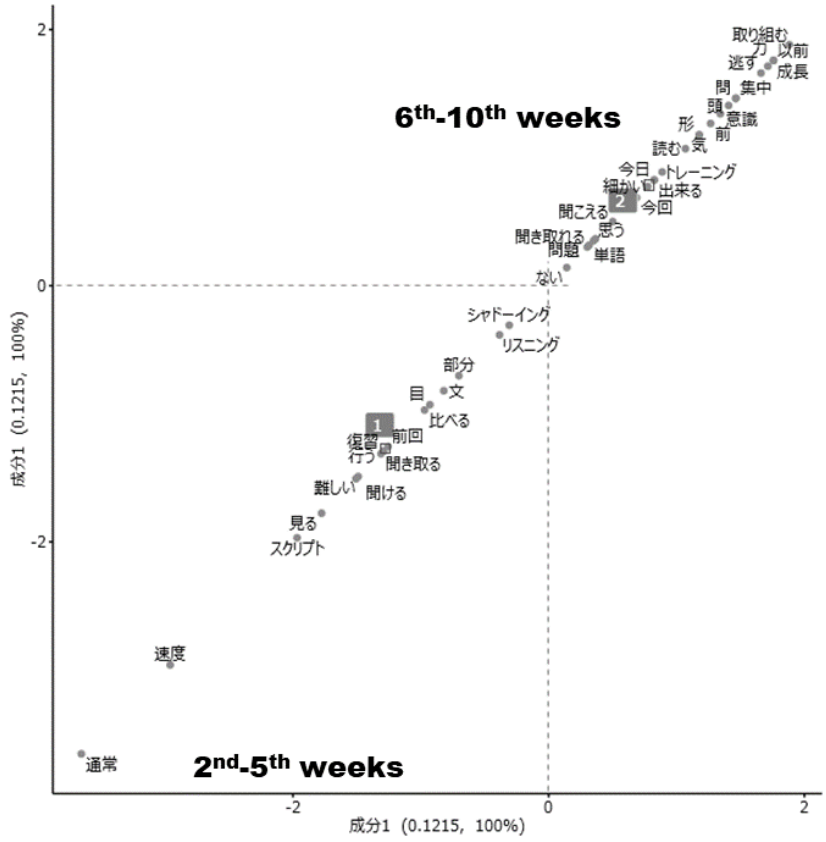
[36] 「最初より成長していると自分で感じている。(7th week)」 「聞きながら頭の中で文をまとめることが出来た。それによって解くことが簡単になったと感じました。全問正解は出来ませんでしたが、最初の時よりは成長しているかと思います。(8th week)」 「最初の頃に比べたら単語一つ一つが聞き取れるようになった。リスニングは一つ一つの単語を繋げて読むから聞き取り辛かったけれど、それも慣れてきた。(9th week)」

[37] 「意味のわからない単語とかが出てきたりしたのでそこの復習をしっかりしたい。(2nd week)」

while working on the training.[38] In the questionnaire administered during the 10th week, quite a few participants admitted the effect of training and expressed their enthusiasm to continue it even after its completion.[39]

Figure 8.6

Correspondence Analysis Plots of the Most Frequent 40 Words Used by the SH Group in Relation to the Training Period (2nd to 5th weeks vs. 6th to 10th weeks)



[38] 「音が繋がっていたりほとんど発音していなかったりするところがまだ聞き取れないのでそれらに注意して取り組んでいきたいと思います。(7nd week)」 「今回は『聞き取ろうとする』ではなく『意味をとろうとする』ということに意識して取り組みました。(8th week)」

[39] 「毎週取り組むことで耳が英語に慣れていった感覚があったので、今後も少しずつ続けていきたいと思いました。(10th week)」「初めの頃は音声に付いていくので精一杯、または置いていかれていたけれど、回を重ねるごとに音声に付いていくことが出来るようになったので、自分の成長を感じられたので良いトレーニングだった。(10th week)」

In sum, the analyses yielded the following findings: (1) a number of participants first felt the training to be very difficult, especially because they could not detect the word boundaries; (2) however, since they developed their learning strategies (i.e., confirmation with written scripts, application of overlapping, and manipulation of the speech rate) and the function of subvocal rehearsal, they gradually overcame the difficulty and shadowed more words as the training proceeded while updating their pronunciation knowledge; (3) they also felt that their listening ability (or listening scores) probably got better because of the improvement of their speech perception ability and strategy use; and finally (4) they wished to continue the training outside class or even after completion of the training.

These findings are almost identical with those obtained from the ASD group. However, one crucial difference is that, while many responses related to word spellings were observed for the ASD group, only one such response was observed for the SH group.[40] Instead of word spellings, the attention of the SH group was solely directed toward articulation (or subvocalization) of word sounds. This difference probably led to the contrastive results between the two measurement tasks (i.e., the RPD-R and the CNT-R). Specifically, the ASD group demonstrated the greatest improvement on the RPD-R where the participants needed to spell out words, while only the SH group showed significant improvement on the CNT-R where the participants needed to rehearse the input in mind to count the number of words.

8.4 Summary of Findings: Chapter 8 (Study 5-2)

8.4.1 Answer to RQ6-1

RQ6-1: What kinds of effects did the participants feel throughout the training?

The text analyses on the participants' responses to the weekly journal and the

[40] 「重要と思われるキーワードを聞くものの、頭で連想するスペルが違うと内容がわからなくなってしまった。(4th week)」

questionnaire revealed that the two training groups (i.e., the ASD and SH groups) had experienced very similar effects—of which there were four kinds. First, despite the training difficulty at the beginning, a majority of the participants became accustomed to the training and felt that their performance on the training gradually improved. Notably, as the week passed, the ASD group was able to transcribe more words, while the SH group was capable of repeating more words. Also, many felt that their ability to perceive words had improved. Second, the training made the participants aware of the gap between their knowledge about word pronunciations and how these words are actually pronounced. It was especially challenging for the participants to figure out the word boundaries because English speech is often accompanied with various phonological changes. Checking out written scripts was an effective way for the participants to fill the gap and realize exactly what words were indistinguishable for them. Third, many participants mentioned that their listening ability (or their listening scores) had improved because of the training. The improvement can be attributed to development of the participants' speech perception ability and listening strategy use. Many participants from the SH group developed their own learning strategies such as adjusting the speech rate and brushing up shadowing skills through overlapping. Finally, many participants became willing to do the training outside class, or even after completion of the course, because they learned the importance of speech perception ability and the effectiveness of the training.

8.4.2 Answer to RQ6-2

RQ6-2: What kinds of cognitive processes were involved in accelerated speech dictation and shadowing?

The finding that the two groups received similar effects from the training indicates that they had also gone through similar cognitive processes. The series of analyses revealed that both groups had found it very difficult to figure out word boundaries. This was caused by the

circumstance under which the participants could not analytically process the input because of the excessive passage speed (for the ASD group) and the difficulty derived from simultaneously performing listening and speaking (for the SH group). Nonetheless, most participants ruminated over their cognitive processes during the training and found the particular problems that faced them. In addition, to compensate for the lack of their knowledge about pronunciation and vocabulary, they explored their own learning strategies. Such enhancement of their metacognition contributed to the improvement of their speech perception ability, resulting in greater training performance and comprehension ability.

Despite these similarities in the way that the two groups approached the training, there was a crucial difference in their cognitive processes, which probably led to the contrastive results for the two measurement tasks (i.e., the RPD-R and the CNT-R). The ASD group, whose goal was to reproduce the spoken input in the written format, had been attentive to both spellings and pronunciations of the words throughout the training. It was then assumed that they were able to make the greatest progress on the RPD-R in which spelling knowledge is important. In contrast, the SH group exclusively concentrated on mimicking the input without paying attention to word spellings. For this reason, they could not make as much progress on the RPD-R as the ASD group but showed significant improvement on the CNT-R where learners need to subvocalize the input to count the number of target words.

8.4.3 Study Limitations

To confirm plausibility of the rationales stated above, future research needs to reveal two things. First, it is necessary to investigate whether accelerated speech dictation will contribute to enrichment of the spelling knowledge; specifically, paradigms that can measure the ability to spell out words based on their pronunciation need to be implemented. Second, there is also a need to reveal whether shadowing will increase the efficiency of subvocalization.

Listening span tests, used in working memory research (e.g., Daneman and Carpenter, 1980; Ishiou & Osaka, 1994), can be used for this purpose.

Chapter 9

General Conclusion

9.1 Overview of Findings

9.1.1 Influence of Speech Rate on Perception (Study 1)

Based on the literature, the author hypothesized that 200 wpm is a threshold level at which most L2 listeners have great difficulty in perception. In Study 1, the reproduction rates of the upper- and lower-level groups gradually decreased as the speech rate increased, suggesting that the faster the speech, the more difficult its perception, regardless of the learner's proficiency level. The study failed to provide positive evidence for the hypothesis; however, the perception of the upper-level learners listening to the speech faster than 200 wpm was as poor as that of the lower-level learners listening to 135-wpm speech. This result indicates that English speech faster than 200 wpm is cognitively demanding even for advanced listeners; thus, dictation using materials of this speed is expected to engage learners in the perception of aural materials. Error analyses revealed that several phonemes, as well as unstressed function words, tended to be obstacles in perception for Japanese English learners.

9.1.2 Short-Term Effects of Accelerated Speech Dictation and Shadowing (Study 2)

The purpose of Study 2 was to compare the short-term effects of accelerated speech dictation (dictation using English speech faster than 200 wpm) and shadowing on speech perception ability. Tests using a written reproduction task administered before and after a 15-minute training session revealed that the two training groups showed significant parallel score increases. This result may indicate that accelerated speech dictation and shadowing were equally efficacious. Still, it was indecisive due to several limitations related to the study design, such as the number of participants, the absence of a control group, the length of training, the

difficulty of the test material, and a possible practice effect caused by using the same material in the pre-and post-tests. In addition, there was a question about the test task since written reproduction tasks involve writing.

9.1.3 Measurement Tasks for Speech Perception Ability (Studies 3 and 4)

Written reproduction tasks (dictation tasks as a testing method) have been commonly used to measure speech perception ability in the literature. Still, learners' performance in those tasks reflects not only their perception ability due to their skill-integration feature. Studies 3 and 4 thus attempted to validate an original non-integrative task, named the word count task, as a measurement task for speech perception ability. To this end, Study 3 compared learners' scores on a word count task (CNT) with those on a written reproduction task (RPD); however, both the validity (external aspect analyzed by correlation analyses) and reliability (internal consistency measured by Cronbach's alpha) of the CNT were found to be questionable.

Study 4, based on suggestions of the previous study, developed revised versions of the two measurement tasks (CNT-R and RPD-R) and analyzed the participants' scores using the same procedures as Study 3, revealing a remarkable improvement in both indices. Moreover, participants' responses to the listening strategy survey suggested that phonological processing was more dominant in both tasks than meaning processing. However, it also turned out that the two tasks measure somewhat different aspects of listening ability; that is, the CNT-R promotes phonological processing at the levels of words and chunks, while the RPD-R demands closer attention to sounds as well as to the global meaning of the passage. Therefore, it was concluded that rather than choosing one task, using both tasks would more adequately measure speech perception ability.

9.1.4 Long-Term Effects of Accelerated Speech Dictation and Shadowing (Study 5)

Employing the two measurement tasks and the listening strategy survey developed in Study 4, Study 5 (Studies 5-1 and 5-2) reexamined the effectiveness of accelerated speech dictation and shadowing in the long term. The participants engaged in a two-month training program for either exercise, during which they also kept weekly journals on the training effect. Both quantitative and qualitative analyses were performed to test the question thoroughly. Moreover, the experimental groups were compared with a control group engaged in no training.

Statistical analyses of the two measurement tasks and the listening strategy survey revealed that both exercises were effective in improving speech perception ability overall. They are expected to help develop skills such as phonemic decoding, holistic sound processing, and spoken word recognition. However, accelerated speech dictation, which includes both listening and writing, can be distinguished by its effects on meaning processing and written word recognition; thus, it may be effective in improving general listening ability as well as spelling knowledge. By contrast, while shadowing engaged the learners in little higher-level processing, as claimed in the literature, it improved their articulatory rehearsal process in working memory.

The text analyses also revealed common effects of the two trainings. First, although both groups initially found the exercise (especially the task of finding word boundaries) very challenging, they gradually overcame this difficulty and felt progress in their performance. Second, the participants identified their perception problems and learned how English words sound in actual utterances. Third, many participants referred to the improvement of their listening ability, which can be attributed to the development of their perception ability and strategy use. Finally, a number of participants, regardless of the training type, expressed enthusiasm to continue the training outside class.

9.2 Pedagogical Implications

9.2.1 Learning Effects Expected From Each Exercise

The findings of the present research suggest that a two-month training with accelerated speech dictation or shadowing can result in the development of speech perception ability. As the strategy surveys showed, the cognitive processes involved in the two exercises are not necessarily identical, thus they may contribute to the development of different sub-skills. Accelerated speech dictation, in which learners need to transcribe all words correctly, will help learners acquire more elaborate input decoding skills than shadowing. Moreover, the process of spelling out words prompts learners to be more attentive to word forms and to keep them in memory, leading to an expansion of their lexical knowledge. This knowledge may transfer to other language skills; therefore, as the literature suggests, accelerated speech dictation can be incorporated into the language classroom as a general exercise.

In contrast, since shadowing is an online task where learners must listen and speak simultaneously, learners' focus on phonological features of the input is not as intense as that in accelerated speech dictation. However, shadowing is distinctive in that it enhances learners' working memory through the process in which they subvocalize the input. Working memory plays a crucial role in the decoding process because listeners need to hold linguistic information in memory to parse it and comprehend the message. In addition, researchers have argued for the importance of working memory in language learning; therefore, shadowing may also contribute to developing other language skills.

9.2.2 Instructional Tips

Although the two exercises are not necessarily identical in their cognitive features, the same instructional tips are useful for L2 teachers, of which three are adduced here. First, since both exercises are inherently challenging for learners, teachers should allow learners to choose

materials of an appropriate level or adjust their difficulty when they are too hard. To this end, the listening class should be conducted in an environment where learners can access audio data of their desirable materials. Giving them control on speech rate is a useful option, but it is important to remind them not to make it too slow. This kind of instruction seems to contradict accelerated speech dictation. However, as Study 1 indicated, perception difficulty derived from passage speed increases incrementally as its rate rises, rather than soaring when the rate exceeds 200 wpm; thus, the moderate speed level is different from learner to learner.

Second, teachers should make written scripts available for learners to refer to when they find it impossible to perceive (or reproduce) any more words. As claimed by many researchers, difficulty in perception is often due to the inability to segment connected speech. Primarily, this problem arises from characteristics of spoken English, such as phonological modification; in many cases, learners do not realize what went wrong. Participants' feedback in the weekly journals suggested that checking written scripts enabled them to become aware of their perception problems, allowing them to fill the gap between their pronunciation knowledge and actual pronunciation. To encourage this process, explicit instruction on the prosodic features of spoken English is also effective.

Third, it is also effective to have learners reflect on their training. The results of Study 5 suggest that keeping weekly journals helped learners realize not only their problems but also their progress. This will enable them to keep track of their learning and maintain their motivation toward learning. Moreover, as shown by some participants' reports on how they had worked on the training, their awareness of listening strategies may be raised. Asking learners to share their learning experience with the class is a useful instructional task as well because it can be an indirect metacognitive experience, which may result in better strategy use by the other learners.

9.3 Limitations and Suggestions for Future Research

Although the current study offers new insights into the effectiveness of dictation and shadowing, it also has several limitations. This final section presents suggestions for future research circumventing these limitations.

First, the effectiveness of the two exercises, when conducted for a short time, which Study 2 aimed to reveal, is still open to question because the follow-up study (i.e., Study 5) was conducted much longer. As repeated use of the same test for both pre- and post-tests with a short interval may well lead to a practice effect, the study should employ a design where participants do not take the same test repeatedly while counterbalancing the difficulties of the pre-and post-test materials.

Second, the impact of learners' proficiency level on the training effect could not be examined. The literature suggests that the effectiveness of dictation and shadowing may vary depending on the learner's listening ability or the difficulty of training materials. In Study 2, the size of each training group was too small (accelerated speech dictation group = 14, shadowing group = 13) for them to be divided into proficiency groups. In Study 5, the control group, which was recruited from outside the course, outperformed the experimental groups on all three pre-tests. To deal with the cases where proficiency gaps were found, a comprehension test was implemented that used the scores as a covariate; however, the data did not fulfill the assumptions of the covariate analysis. In future research, learners' proficiency levels at the beginning should be controlled among all groups.

Third, the measurement tasks used in this research (i.e., written reproduction tasks and word count tasks) are aimed at accuracy in perception. However, the degree to which learners can process the input efficiently is also important in listening because of adverse listening conditions (e.g., perception is hampered by spoken English features). The present research results suggest that the phonological loop of the shadowing group developed, while the

accelerated speech dictation group gradually overcame the difficulties arising from the speed of the material. This hints that the learners improved their efficiency in speech perception. For these reasons, future research needs to investigate whether the two exercises improve learners' perception efficiency using a study paradigm focused on, for example, how fast learners can recognize words. According to Jiang (2012), research paradigms aimed at measuring reaction time provide useful information that cannot be obtained by studies focusing on accuracy rates because they employ online tasks.

Finally, although this study has provided positive evidence for improving speech perception ability, it did not investigate whether this improvement contributes to better comprehension ability. To test this, along with the tests for speech perception ability, listening comprehension tests need to be administered in the pre-and post-tests, analyzing whether the improvement of speech perception ability constitutes a significant predictor of the improvement of comprehension ability.

References

Afsharrad, M., & Benis, A. R. S. (2014). The effect of transcribing on beginning learners' dictation. *Theory and Practice in Language Studies, 4(10)*, 2203-2208. https://doi.org/10.4304/tpls.4.10.2203-2208

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). Worth Publishers.

Audacity. [Computer software] https://www.audacityteam.org/

Baddeley, A. (1992). Working memory. *Science, 255,* 556-559.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417-423.

Baddeley, A. D. (1999). *Essentials of human memory.* Psychology Press.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford University Press.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63*, 1-29. https://doi.org/10.1146/annurev-psych-120710-100422

Baker, W., Trofimovich, P., Flege, J. E., Mack, M., & Halter, Randall. (2008). Child-adult differences in second-language phonological learning: The role of cross-language similarity. *Language and Speech, 51(4)*, 317-342. https://doi.org/10.1177/0023830908099068

Beatty, M. J., Behnke, R. R., & Goodyear, F. H. (1979). Effects of speeded speech presentations on confidence-weighted and traditional comprehension scores. *Communication Monographs, 43*, 147-151.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp.

167-224). The MIT Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro & O. Bohn (Eds.), *Language experience in second language second language speech learning: In honor of James Emil Flege* (pp. 13-45). John Benjamins.

Blau, E. K. (1990). The effect of syntax, speed, and pauses on listening comprehension. *TESOL Quarterly, 24,* 746-753.

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(4)*, 2299-2310.

Breitenstein, P. H. (1972). Readers' letters. *English Language Teaching, 26*, 202-203.

Brown, H. D., & Abeywickrama, P. (2010). *Language assessment: Principles and classroom practices* (2nd ed.). Pearson Education.

Brown, J. D., & Hilferty, A. (1986). The effectiveness of teaching reduced forms of listening comprehension. *RELC Journal, 17*(2), 59-70.

Brown, S. (2011). *Listening myths: Applying second language research to classroom teaching.* The University of Michigan Press.

Buck, G. (2001). *Assessing listening.* Cambridge University Press.

Buck, G., & Tatsuoka, K. (1998). Application of the rule-space procedure to language testing: Examining attributes of a free response listening test. *Language Testing, 15*(2), 119-157.

Cai, H. (2012). Partial dictation as a measure of EFL listening proficiency: Evidence from confirmatory factor analysis. *Language Testing, 30(2)*, 177-199. https://doi.org/10.1177/0265532212456833

Carroll, J. B. (1972). Defining language comprehension: Some speculations. In J. B. Carroll &

R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge* (pp.1-29). V. H. Winston & Sons.

Cebrian, J., & Carlet, A. (2014). Second language learners' identification of target language phonemes: A short-term phonetic training study. *Canadian Modern Language Review, 70(4)*, 474-499.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge University Press.

Chino, J. (2006). Effects of the combination of dictation and shadowing practices on listening comprehension skill. *Language Education & Technology, 43*, 95-109. https://doi.org/10.24539/let.43.0_95

Clark, H. H., & Clark, E. V. (1977). *Psychology and language: Introduction to psycholinguistics.* Harcourt.

Cohen, J. (2015). The effectiveness of using dictation to develop listening comprehension. *Kinki University, Center for Liberal Arts and Foreign Language Education Journal, 6(2)*, 43-56.

Conrad, L. (1989). The effects of time-compressed speech on native and EFL listening comprehension. *Studies in Second Language Acquisition, 11,* 1-16.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104(3)*, 268-294. https://psycnet.apa.org/doi/10.1037/0096-3445.104.3.268

Cross, J. (2009). Diagnosing the process, text, and intrusion problems responsible for L2 listeners' decoding errors. *The Asian EFL Journal Quarterly, 11*(2), 31-53.

Cummins, J., Swain, M., Nakajima, K., Handscombe, J., Green, D., & Tran, C. (1984). Linguistic interdependence among Japanese and Vietnamese immigrant students. In C. Rivera (Ed.), *Communicative competence approaches to language proficiency assessment: Research and applications* (pp. 60-81). Multilingual Matters.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Davis, P., & Rinvolucri, M. (1988). *Dictation: New methods, new possibilities.* Cambridge University Press.

Dupoux, E., & Green, K. (1997). Perceptual adjustment to highly compressed speech: Effects of talker and rate changes. *Journal of Experimental Psychology: Human Perception and Performance, 23(3)*, 914-927. https://doi.org/10.1037//0096-1523.23.3.914

Educational Testing Service. (2012). *TOEIC test shin koshiki mondaishu vol. 5.* The Institute for International Business Communication.

Ellis, R. (2008). *The study of second language acquisition* (2nd ed.). Oxford University Press.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications Inc.

Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal, 57*(4), 325-334.

Field, J. (2008). *Listening in the language classroom.* Cambridge University Press.

Field, J. (2013). Cognitive validity. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: Research and practice in assessing second language listening* (pp. 77-151). Cambridge University Press.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-

developmental inquiry. *American Psychologist, 34(10)*, 906-911. https://doi.org/10.1037/0003-066X.34.10.906

Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice.* Cambridge University Press.

Fujii, M., Kosugi, K., & Lee, J. W. (Eds.). (2005). *Fukushi, shinri, kaigo no text mining nyuumon* [An introduction to text mining in the fields of welfare, psychology, and nursing]. Chuou Houki.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds "L" and "R." *Neuropsychologia, 9(3)*, 317-323.

Grabe, W., & Stoller, F. L. (2011). *Teaching and researching reading* (2nd ed.). Pearson Education.

Green, A. (2014). *Exploring language assessment and testing: Language in action.* Routledge.

Griffiths, R. (1990). Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning, 40,* 311-336.

Griffiths, R. (1992). Speech rate and listening comprehension: Further evidence of the relationship. *TESOL Quarterly, 26,* 385-390.

Habibi, P., Nemati, A., & Habibi, S. (2012). The role of listening comprehension in dictation. *Indian Journal of Science and Technology, 5(8)*, 3208-3210. https://doi.org/10.17485/ijst/2012/v5i8.18

Hamada, Y. (2017). *Teaching EFL learners shadowing for listening: Developing learners' bottom-up skills.* Routledge.

Henning, G., Gary, N., & Gary, J. (1983). Listening recall: A listening comprehension test for low proficiency learners. *System, 11*, 287-293. https://doi.org/10.1016/0346-251X(83)90046-5

Hirai, A (Ed.). (2017). *Kyoiku-shinrikei kenkyu no tameno data bunseki katsuyouhou* (2nd ed.) [Introduction to data analysis for educational and psychological research]. Tokyo Tosho.

Hirai, A., Fujita, R., & O'ki, T. (2013). The influence of the Center Listening Test on listening learning motivation: An analysis focusing on university type, admission type, and major. *JACET (The Japan Association of College English Teachers) Journal, 57*, 59-81.

Hughes, A. (2003). *Testing for language teachers* (2nd ed.). Cambridge University Press.

Ishiou, A., & Osaka, M. (1994). An approach to measure a listening span for preschool children. *Japanese Journal of Educational Psychology, 42*, 167-173.

Jafarpur, A., & Yamini, M. (1993). Does practice with dictation improve language skills? *System, 21(3)*, 359-369.

Jiang, N. (2012). *Conducting reaction time research in second language studies.* Routledge.

Kadota, S. (2007). *Shadowing to ondoku no kagaku* [Science of shadowing and oral reading]. Cosmopier.

Kadota, S. (2012). *Shadowing, ondoku to eigo shutoku no kagaku: Input kara output e* [Science of shadowing, oral reading, and English acquisition]. Cosmopier.

Kadota, S., & Tamai, K. (2004). *Ketteiban eigo shadowing* [The definitive book on English shadowing]. Cosmopier.

Kakehi, H., Suenobu, M., Young, R., Kanazaki, K., & Yamane, S. (1981). An analysis of perceptual error: Learning process. *JACET (The Japan Association of College English Teachers) Journal, 12*, 133-144.

Katayama, Y., Nagase, Y., & Joto, A. (1996). *Eigo onseigaku no kiso: Oto henka to prosody wo chushin ni* [Fundamentals of English phonetics: A special reference to sound changes and prosody]. Kenkyusha Shuppan.

KATE Research Promotion Committee (Kanto Koshinetsu Eigo Kyoiku Gakkai Kenkyu

Suishin Iinkai). (2018). To what extent do Japanese high school students acquire what they have learned in junior high school? *KATE (Kantokoshinestu Association of Teachers of English) Journal, 32*, 115-128.

Kato, K. (2009). *L2 speech learning: Perception and production of English by Japanese native speakers.* VDM Verlag Dr. Müller Aktiengesellschaft & Co. KG.

Kawasaki, E. (Ed.) (2005). *Kotoba no jikken shitsu: Shinri gengogaku eno approach* [Introduction to psycholinguistics]. Brain Shuppan.

Kelch, K. (1985). Modified input as an aid to comprehension. *Studies in Second Language Acquisition, 7,* 81-90.

KH Coder. [Computer software] https://khcoder.net/

Kiany, G. R., & Shiramiry, E. (2002). The effect of frequent dictation on the listening comprehension ability of elementary EFL learners. *TESL Canada Journal, 20(1)*, 57-63.

Kohno, M. (1993). Perceptual sense unit and echoic memory. *International Journal of Psycholinguistics, 9,* 13-31.

Kohno, M. (2001). *Onsei gengo no ninshiki to seisei no mechanism: Kotoba no jikan seigyo kikou to sono yakuwari* [The mechanism of recognition and production of oral language: Time control system of language and it role]. Kinseido.

Kohno, M (Ed.). (2007). *Explorations into the mechanism of language and cognition.* Sanseido.

Komori, K. (2010). *2010 nendo center shiken listening test no bunseki to taisaku* [An analysis of the National Center Listening Test in 2010 and its preparation]. G.C.D. *Eigo Tsushin, 47*, 16-17.

Kougo, H., & Kubono, M. (2004). Eigo II: Intake of the text and insight. *Eigo Kyoiku, 53 (September)*, 40-42.

Koya, T. (2017). *Shokyu eigo gakushusha no choukai ni ataeru hatsuwa sokudo chousei no*

kouka (The effect of speech rate adjustment on listening comprehension of novice English learners). *Housei Daigaku Koganei Ronshu, 13*, 11-30. http://doi.org/10.15002/00013995.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MIT Press.

Levelt, W. J. M. (1993). Language use in normal speakers and its disorders. In G. Blanken, J. Dittmann, H. Grimm, J. C. Marshall & C.-W. Wallesch (Eds.), *Linguistic disorders and pathologies* (pp. 1-15). De Gruyter.

Lynch, T., & Mendelsohn, D. (2010). Listening. In N. Schmitt (Ed.), *An introduction to applied linguistics* (2nd ed.) (pp. 180-196). Hodder Education.

Marzban, A., & Abdollahi, M. (2013). The effect of partial dictation on the listening comprehension ability of Iranian intermediate EFL learners. *International Research Journal of Applied and Basic Sciences, 5(2)*, 238-244.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics, 9*, 283-303.

Mohammed, B. S. (2015). Using dictation in teaching college students. *Global Journal on Technology, 8*, 205-215. (An online journal)

Morris, S. (1983). Dictation: a technique in need of reappraisal. *ELT (English Language Teaching) Journal, 37*(2), 121-126.

Munby, J. (1978). *Communicative syllabus design: A sociolinguistic model for defining the content of purpose-specific language programmes.* Cambridge University Press.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking.* Routledge.

Nunan, D. (1997, October 3-5). Approaches to teaching listening in the language classroom [Conference session]. Korea TESOL Conference, Kyoung-ju, South Korea. Retrieved from: https://koreatesol.org/content/kotesol-proceedings-1997

Obunsha. (2019). *Nanoka-kan kansei eiken 2 kyu yosou mondai doriru* (5th ed.) [Drill questions

for Grade 2 of the Eiken tests to get ready in seven days].

Oki, T. (2010a). Investigating the role of shadowing for facilitating bottom-up processing. *Tsukuba Eigo Kyoiku (Tsukuba Review of English Language Teaching), 31*, 1-22.

Oki, T. (2010b). The role of latency for word recognition in shadowing. *ARELE (Annual Review of English Language Education in Japan), 21*, 51-60. https://doi.org/10.20581/arele.21.0_51

O'ki, T. (2011). Shadowing kaishiki ni okeru gakushusha no fukushou strategy no bunrui [Classification of learners' repetition strategies in shadowing at the beginning stage]. *KATE (Kantokoshinetsu Association of Teachers of English) Journal, 25*, 33-43.

O'ki, T. (2012a). ESL/EFL listening ni okeru hatsuwa sokudo no yakuwari: 'kotoba no jikan seigyo kikou' ni motoduita saikou [The role of speech speed in ESL/EFL listening; A reconsideration based on the "temporal control system of language"]. *Hakuoh Journal of the Faculty of Education, 6*(1), 91-112.

O'ki, T. (2012b) Word repetition in EFL shadowing: The roles of phrasal knowledge, context, and proficiency. *ARELE (Annual Review of English Language Education in Japan), 23*, 45-60. https://doi.org/10.20581/arele.23.0_45

O'ki, T. (2014). Kurikaeshi no shadowing eno kouka: Gakushusha wa yori ookuno gowo fukushou dekiruka? [Effectiveness of repetition on shadowing: Can learners repeat more words?]. *Hakuou Daigaku Ronshu (The Hakuoh University Journal), 28*(2), 169-187.

O'ki, T., & Izumi, Y. (2015). Kurikaeshi no shadowing eno kouka [Effectiveness of repetition on shadowing: Can learners repeat more words?] *Hakuoh Daigaku Ronshu (The Hakuoh University Journal), 28*(2), 169-187.

Oller, J. W. Jr. (1971). Dictation as a device for testing foreign language proficiency. *ELT Journal, 25(3)*, 254-259.

Oller, J. W. Jr., & Streiff, V. (1975). Dictation: A test of grammar-based expectancies. *ELT Journal, 30*(1), 25-36.

Osaka, M. (2002). *Nou no memochou: Working memory* [Working memory: The sketchpad in the brain]. Shinyo-sha.

Oshima, H. (2003). Bunpou yakuokushiki kara Onsei juushie: Intake wo huyasu shadowing [A paradigm shift from focus-on-grammar to focus-on-sounds: Shadowing for enhancing the intake process]. *Eigo Kyoiku, 51*, 40-42.

Oyama, Y. (2009). Effect of English text speed reading training on English listening comprehension for Japanese university students: Comparison to dictation training. *Japan Journal of Educational Technology, 32(4)*, 351-358. https://doi.org/10.15077/jjet.KJ00005353783

Pallier, C., Sebastian-Gallés, N., Dupoux, E., Christophe, A., & Mehler, J. (1998). Perceptual adjustment to time-compressed speech: A cross-linguistic study. *Memory & Cognition, 26(4)*, 844-851. https://doi.org/10.3758/BF03211403

Papagno, C., & Vallar, G. (1992). Phonological short-term memory and the learning of novel words: The effects of phonological similarity and item length. *Quarterly Journal of Experimental Psychology, 44A*, 47-67.

Papagno, C., Vallentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*(3), 331-347.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance, 31(6)*, 1315-1330. https://doi.org/10.1037/0096-1523.31.6.1315

Rahimi, M. (2008). Using dictation to improve language proficiency. *The Asian EFL Journal,*

10(1), 33-47.

Raphael, L. J. (2005). Acoustic cues to the perception of segmental phonemes. In D. B. Pisoni, & R. E. Remez (Eds.), *The handbook of speech perception* (pp. 182-206). Blackwell Publishing.

Richards, J. C. (1983). Listening comprehension: Approach, design, procedure. *TESOL Quarterly, 17(2)*, 219-240. https://doi.org/10.2307/3586651

Richards, J. C., & Schmidt, R. (2010). *Longman dictionary of language teaching & applied linguistics* (4th ed.). Pearson Education.

Rivers, W. M. (1966). Listening comprehension. *The Modern Language Journal, 50*(4), 196-204. https://doi.org/10.1111/j.1540-4781.1966.tb01817.x

Rivers, W. M., & Temperley, M. S. (1978). *A practical guide to the teaching of English as a second or foreign language.* Oxford University Press.

Rost, M. (2016). *Teaching and researching listening* (3rd ed.). Routledge.

Rost, M., & Wilson, JJ. (2013). *Active listening.* Pearson Education.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal, 78(2)*, 199-221.

Ryalls, J. (1996). *A basic introduction to speech perception.* Delmar, Cengage Learning.

Sakai, H. (2009). Text length and language for recall: Their influence on L2 listening test-taking strategies. *JACET (The Japan Association of College English Teachers) Journal, 48*, 67-79.

Sano, F. (1995). The oral approach. In K. Tasaki (Ed.), *Gendai eigo kyojuhou souran* [Introductory Book to Modern English Teaching] (pp. 59-69). Taishukan Shoten.

Sato, T., & Nakamura, N. (1998). Shadowing no kouka to gakushusha no ishiki [The effects of shadowing and the students' feedback]. *Tsukuba Kokusai Daigaku Kenkyu Kiyou, 4*, 47-

57.

Sawaki, Y., & Nissan, S. (2009). *Criterion-related validity of the TOEFL® iBT listening section.* Princeton, NJ: ETS.

Sebastian-Gallés, N., Dupoux, E., Costa, A., & Mehler, J. (2000). Adaptation to time-compressed speech: Phonological determinants. *Perception & Psycholinguistics, 62(4)*, 834-842. https://doi.org/10.3758/BF03206926

Shi, L., & Farooq, N. (2012). Bilingual listeners' perception of temporally manipulated English passages. *Journal of Speech, Language, and Hearing Research, 55*, 125-138. https://doi.org/10.1044/1092-4388

Shiffrin, R. M., & Shneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices: An analysis of reproduction rate and types of reproduced words. *ARELE (Annual Review of English Language Education in Japan), 21*, 81-90. https://doi.org/10.20581/arele.21.0_81

Shizuka, T. (2019). *Nihongo neitibu ga nigatena eigo no oto to rizumu no tsukurikata ga ichiban yokuwakaru hatsuon no kyoukasho* [The best pronunciation textbook for Japanese native speakers to learn English sounds and rhythm]. TOEFL Seminar.

Siegel, J., & Siegel, A. (2015). Getting to the bottom of L2 listening instruction: Making a case for bottom-up activities. *Studies in Second Language Learning and Teaching, 5(4)*, 637-662. https://doi.org/10.14746/ssllt.2015.5.4.6

Stanovich, K. E. (1980). Toward an interactive -compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly, 16*(1), 32-71.

Suenobu, M., Young, R., Kanzaki, K., & Yamane, S. (1982). Ananalysis of perceptual error: Effect of learning and mechanism of hearing. *JACET (The Japan Association of College English Teachers) Journal, 13*, 83-97.

Sugawara, Y. (1999). Dictation and listening comprehension: Does dictation promote listening comprehension? *Language Education & Technology, 36*, 33-50. https://doi.org/10.24539/llaj.36.0_33

Sugito, M. (2012). *Nigongo no accent, Eigo no accent* [Japanese accents and English accents]. Hitsuji Shobou.

Suzuki, H. (2007). Shadowing wo mochiita eigoryoku koujou no shidou ni tsuiteno kenshou [An investigation into the instructions for fostering English listening comprehension using shadowing]. *STEP Bulletin, 19*,112-124.

Suzuki, J., & Kadota, S. (Eds.). (2018). *Eigo listening shidou handbook* [Handbook of listening teaching]. Taishukan Shoten.

Takayama, Y. (2007). Shadowing skill wa eigo unyou nouryoku no shihyou to nariuruka [Can the shadowing skill be a measure for practical English ability]. *Eigaku Ronkou, 36*, 11-23.

Takebayashi, S., & Saito, H. (2008). *Eigo onseigaku nyumon* [An introduction to English phonetics]. Taishukan Shoten.

Takei, A (Ed.). (2002). *Eigo listening ron [The theory of English listening].* Kagensha.

Takeuchi, O. (1997). Dictation: Is it really effective for language teaching? *Kansai Daigaku Shichoukaku Kyoiku, 20*, 55-63.

Tamai, K. (1992). Follow-up-no choukairyoku koujou-ni oyobosu kouka oyobi follow-up nouryoku-to choukairyoku-no kankei [The effect of “follow-up” on development of listening ability and the relationship between “follow-up” ability and listening ability]. *STEP (Society for Testing English Proficiency) Bulletin, 4*, 48-62.

Tamai, K. (1997). Shadowing-no kouka-to choukai process-ni-okeru ichiduke [The effect of shadowing and its role for listening comprehension]. *Jiji Eigogaku Kenkyu, 36,* 105-116.

Tamai, K. (2005). *Listening-shidohou-to-shiteno shadowing-no koka-ni kansuru-kenkyu.* [Research on the effects of shadowing as a listening instruction]. Kazama Shobo.

Tatham, M., & Morton, K. (2006). *Speech production and perception.* Palgrave Macmillan.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11,* 90-105.

Taylor, L. (2013). Introduction. In A. Geranpayeh & L. Taylor (Eds.), *Examining listening: research and practice in assessing second language listening* (pp. 1-35). Cambridge University Press.

Templeton, H. (1977). A new technique for measuring listening comprehension. *ELT (English Language Teachers) Journal, 31(4)*, 292-299. https://doi.org/10.1093/elt/XXXI.4.292

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native Korean adults and children. *Journal of Phonetics, 33*, 263-290. https://doi.org/10.1016/j.wocn.2004.10.002

Ur, P. (1984). *Teaching listening comprehension.* Cambridge University Press.

Vandergrift, L. (2003). Orchestrating strategy use: Toward a model of the skilled second language listener. *Language Learning, 53*, 463-496. https://doi.org/10.1111/1467-9922.00232

Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* Routledge.

Weir, C. (1993). *Understanding & developing language tests.* Prentice Hall.

Wingfield, A., & Nolan, K. A. (1980). Spontaneous segmentation in normal and in time-

compressed speech. *Perception & Psychophysics, 28,* 97-102.

Word Level Checker. http://someya-net.com/wlc/index_J.html

Yanagihara, K. (1995). Eigo Chokairyoku no shidouhou ni kansuru jikkenteki kenkyu: Shadowing to dictation no kouka nit suite [A study of teaching methods for developing English listening comprehension: The effects of shadowing and dictation]. *Language Laboratory, 32*, 73-89. https://doi.org/10.24539/llaj.32.0_73

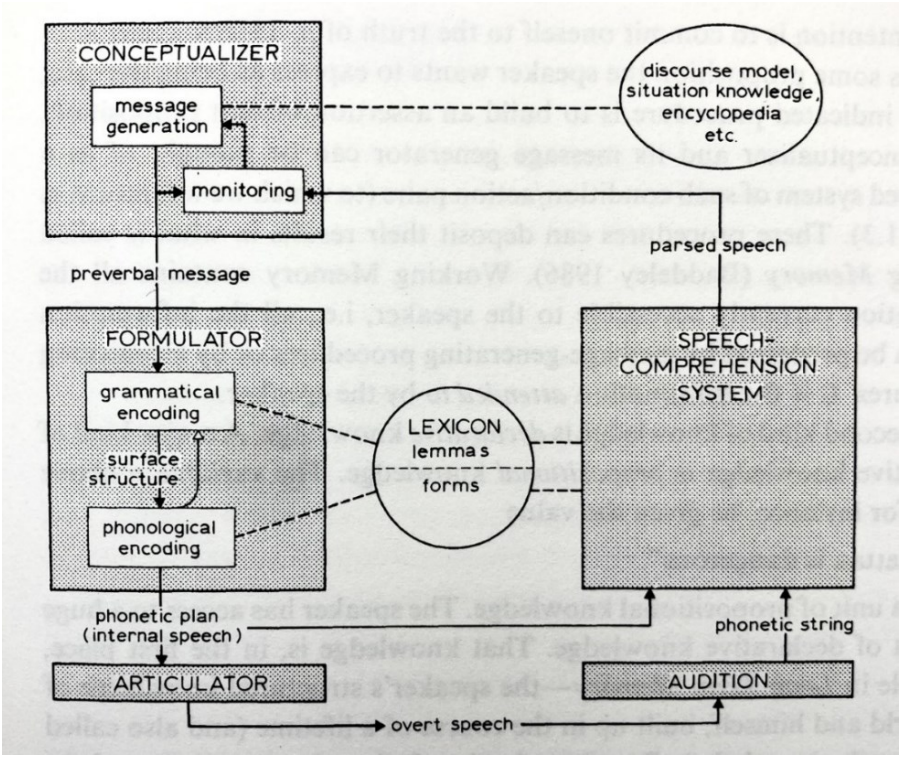
Yashima, T. (1988). Tsuyakukunren-no eigo-kyouiku-eno ouyou I: Shadowing [Application of an interpretation training into English education: Shadowing]. *Eigaku (Heian Jogakuin Tanki Daigaku Eigakukai), 21*, 29-37.

Yonezaki, H. (2014). Is dictation effective at improving listening comprehension of Japanese high school learners of English?: An empirical study focusing on those who excel at reading and grammatical competence. *CELES (The Chubu English Language Education Society) Journal, 43*, 43-50. https://doi.org/10.20713/celes.43.0_43

Zhao, Y. (1997). The effects of listeners' control of speech rate on second language comprehension. *Applied Linguistics, 18,* 49-68.

Appendices

Appendix 2A. A Blueprint for the Speaker (Levelt, 1989, p. 9)



Appendix 2B. Micro-Skills in Conversational Listening (Richards, 1983, pp.228-229)

1. ability to retain chunks of language of different lengths for short period
2. ability to discriminate among the distinctive sounds of the target language
3. ability to recognize the stress patterns of words
4. ability to recognize the rhythmic structure of English
5. ability to recognize the functions of stress and intonation to signal the information structure of utterances
6. ability to identify words in stressed and unstressed positions
7. ability to recognize reduced forms of words
8. ability to distinguish word boundaries
9. ability to recognize typical word order patterns in the target language
10. ability to recognize vocabulary used in core conversational topics
11. ability to detect key words (i.e., those which identify topics and propositions)
12. ability to guess the meanings of words from the contexts in which they occur
13. ability to recognize grammatical word classes (parts of speech)
14. ability to recognize major syntactic patterns and devices

15. ability to recognize cohesive devise in spoken discourse
16. ability to recognize elliptical forms of grammatical units and sentences
17. ability to detect sentence constituents
18. ability to distinguish between major and minor constituents
19. ability to detect meanings expressed in differing grammatical forms/sentence types (i.e., that a particular meaning may be expressed in different ways)
20. ability to recognize the communicative functions of utterances, according to situations, participants, goals
21. ability to reconstruct or infer situations, goals, participants, procedures
22. ability to use real world knowledge and experience to work out purposes, goals, settings, procedures
23. ability to predict outcomes from events described
24. ability to infer links and connections between events
25. ability to deduce causes and effects from events
26. ability to distinguish between literal and implied meanings
27. ability to identify and reconstruct topics and coherent structure from ongoing discourse involving two or more speakers
28. ability to recognize markers of coherence in discourse, and to detect such relations as main idea, supporting idea, given information, new information, generalization, exemplification
29. ability to process speech at different rates
30. ability to process speech at different rates
31. ability to make use of facial, paralinguistic, and other clues to work out meanings
32. ability to adjust listening strategies to different kinds of listener purposes or goals
33. ability to signal comprehension or lack of comprehension, verbally and non-verbally

Appendix 2C. Micro-Skills in Academic Listening (Richards, 1983, pp. 229-230)

1. ability to identify purpose and scope of lecture
2. ability to identify topic of lecture and follow topic development
3. ability to identify relationships among units within discourse (e.g., major ideas, generalizations, hypotheses, supporting ideas, examples)
4. ability to identify role of discourse markers in signaling structure of a lecture (e.g., conjunctions, adverbs, gambits, routines)
5. ability to infer relationships (e.g., cause, effect, conclusion)
6. ability to recognize key lexical items related to subject/topic
7. ability to deduce meanings of words from context

8. ability to recognize markers of cohesion
9. ability to recognize functions of intonation to signal information structure (e.g., pitch, volume, pace key)
10. ability to detect attitude of speakers toward subject matter
11. ability to follow different modes of lecturing: spoken, audio, audio-visual
12. ability to follow lecture despite differences in accent and speed
13. familiarity with different styles of lecturing: formal, conversational, read, unplanned
14. familiarity with different registers: written versus colloquial
15. ability to recognize irrelevant matter: jokes, digressions, meanderings
16. ability to recognize functions of non-verbal cues as markers of emphasis and attitude
17. knowledge of classroom conventions (e.g., turn taking, clarification requests)
18. ability to recognize instructional/learner tasks (e.g., warnings, suggestions, recommendations, advice, instructions)

Appendix 2D. Macro- and Micro-Concepts Related to Speech Perception (Munby, 1978, pp. 123-126)

1. Discriminating sounds in isolate word forms:
    - 1.1 phonemes, especially phonemic contrasts
    - 1.2 phoneme sequences
    - 1.3 allophonic variants
    - 1.4 assimilated and elicited forms (esp. reduction of vowels and consonant clusters)
    - 1.5 permissible phonemic variation

3. Discriminating sounds in connected speech:
    - 3.1 strong and weak forms
    - 3.2 neutralisation of weak forms
    - 3.3 reduction of unstressed vowels
    - 3.4 modification of sounds, esp. at word boundaries, through
        - 3.4.1 assimilation
        - 3.4.2 elision
        - 3.4.3 liaison

5. Discriminating stress patterns within words:
    - 5.1 characteristic accentual pattern
    - 5.2 meaningful accentual patterns
    - 5.3 compounds

7. Recognising variation in stress in connected speech:
   - 7.1 variation of word accentual patterns for rhythmic considerations (e.g., accent shift in 'level-stress' words)
   - 7.2 variation of word accentual patterns for meaningful prominence
   - 7.3 non-stressing of pronouns
   - 7.4 differentiating phrases from compounds

9. Recognising the use of stress in connected speech
   - 9.1 for indicating information units:
     - 9.1.1 content words and form words
     - 9.1.2 rhythmic patterning
   - 9.2 for emphasis, through location of nuclear accent
   - 9.3 for contrast, through nuclear shift

11. Understanding intonation patterns: neutral position of nucleus and use of tone, in respect of
   - 11.1 falling tone with declarative/moodless clauses
   - 11.2 falling tone with interrogative clauses beginning with a question-word
   - 11.3 falling tone with imperative clauses
   - 11.4 rising tone with 'yes/no' interrogative clauses
   - 11.5 rising tone with non-final clauses
   - 11.6 fall-rise tone with any clause type
   - 11.7 rise-fall tone with any clause type
   - 11.8 multi-nuclear patterns
   - 11.9 tones with question-tags
   - 11.10 others

13. Understanding intonation patterns: interpreting attitudinal meaning through variation of tone or nuclear shift, viz.
   - 13.1 rising tone with declarative/moodless clauses
   - 13.2 rising tone with interrogatives beginning with a question word, having the nucleus in
     - 13.2.1 end position
     - 13.2.2 front position
   - 13.3 same as 11.2 but nuclear shift to front position
   - 13.4 rising tone with imperative clauses
   - 13.5 falling tone with 'yes/no' interrogative clauses

13.6 same as 11.4 but nuclear shift to front position

13.7 others

19. Deducing the meaning and use of unfamiliar lexical items

19.1 understanding word formation:

19.1.1 stems/roots

19.1.2 affixation

19.1.3 derivation

19.1.4 compounding

19.2 contextual clues

Appendix 3A. Test Sheet Used in Study 1

※あなたは**セット 1**です。

【課題①】

自分のセット（1〜3）の音声 01 (Normal) を聞いて、聞こえてくる英語を書きとってください。

時間は 10 分間です。はじめに次の注意事項を読み（　　）に✓をいれてください。

（　　）字ははっきりと書いてください

（　　）音声を途中で止めたり、繰り返し聞いたりして構いません

（　　）指示があるまで次の音声は聞かないでください

終わった人も、指示があるまで、次のページには進まないでください

【課題②】

自分のセット（1〜3）の音声 02（30%UP）を聞いて、聞こえてくる英語を書きとってください。時間は 10 分間です。はじめに次の注意事項を読み（　　）に✓をいれてください。

（　　）字ははっきりと書いてください
（　　）音声を途中で止めたり、繰り返し聞いたりして構いません
（　　）指示があるまで次の音声は聞かないでください
（　　）前のページには戻らないでください

終わった人も、指示があるまで、次のページには進まないでください

【課題③】

自分のセット（1〜3）の音声 03（60%UP）を聞いて、聞こえてくる英語を書きとってください。時間は 10 分間です。はじめに次の注意事項を読み（　　）に✓をいれてください。

（　　）字ははっきりと書いてください
（　　）音声を途中で止めたり、繰り返し聞いたりして構いません
（　　）指示があるまで次の音声は聞かないでください
（　　）前のページには戻らないでください

終わった人も、指示があるまで、次のページには進まないでください

Appendix 3B. Reproduction Rate of Each Word in Two Faster Conditions (30% UP and 60% UP) in Study 1

Passage A: Employee Meeting (n = 14 + 14)

Good	morning,	everyone.	Thanks	for	coming	to	this
27/96%	26/93%	28/100%	26/93%	20/71%	27/96%	22/79%	24/86%
employee	meeting.	Unfortunately,	there	was	an	accident	last
12/43%	24/86%	8/29%	17/61%	27/96%	13/46%	25/89%	27/96%
night	in	the	restaurant's	kitchen.	One	of	the
27/96%	28/100%	24/86%	7/25%	20/71%	26/93%	26/93%	22/79%
cooks	burned	his	hand	badly	when	a	pot
18/64%	10/36%	20/71%	24/86%	19/68%	10/36%	10/36%	8/29%
of	hot	soup	was	knocked	over.	I	just
14/50%	19/68%	20/71%	13/46%	1/4%	23/82%	26/93%	28/100%
want	to	remind	you	all	to	follow	our
15/54%	24/86%	10/36%	15/54%	4/14%	17/61%	17/61%	3/11%
safety	rules	at	all	times.	We	want	you
18/64%	6/21%	21/75%	25/89%	21/75%	28/100%	26/93%	14/50%
to	work	quickly,	but	safety	is	the	most
28/100%	27/96%	25/89%	20/71%	23/82%	23/82%	22/79%	23/82%
important	thing.						
23/82%	22/79%						

Passage B: Business Program (n = 14 + 12)

Ted	studies	business	at	college.	For	him,	the
24/92%	18/69%	20/77%	22/85%	14/54%	26/100%	26/100%	22/85%
most	interesting	thing	about	the	program	is	that
26/100%	24/92%	20/77%	24/92%	11/42%	20/77%	20/77%	19/73%
he	sometimes	gets	to	work	in	real	companies.

24/92%	22/85%	21/81%	23/88%	25/96%	21/81%	14/54%	13/50%
He	is	learning	a	lot	about	how	companies
24/92%	20/77%	22/85%	23/88%	22/85%	21/81%	20/77%	7/27%
work.	He	thinks	this	will	be	a	valuable
24/92%	24/92%	23/88%	25/96%	19/73%	23/88%	14/54%	5/19%
experience	for	his	future.	He	has	to	work
25/96%	25/96%	25/96%	24/92%	26/100%	25/96%	23/88%	25/96%
very	hard,	though,	and	he	does	not	have
23/88%	26/100%	13/50%	22/85%	20/77%	21/81%	23/88%	22/85%
much	time	to	relax	at	home	anymore.	
26/100%	26/100%	21/81%	21/81%	21/81%	22/85%	23/88%	

Passage C: Snow Noise (*n* = 12 + 14)

Many	people	find	falling	snow	very	beautiful.	But
26/100%	26/100%	26/100%	9/35%	26/100%	26/100%	26/100%	18/69%
animals	in	the	ocean	may	find	it	annoying.
22/85%	10/38%	12/46%	17/65%	21/81%	25/96%	7/27%	1/4%
Researchers	have	discovered	that	snowflakes	hitting	the	ocean’s
7/27%	23/88%	9/35%	15/58%	5/19%	3/12%	10/38%	2/8%
surface	create	a	noise.	For	animals	under	the
22/85%	10/38%	2/8%	22/85%	24/92%	24/92%	24/92%	25/96%
surface,	this	sound	can	be	very	loud.	The
24/92%	26/100%	26/100%	25/96%	24/96%	26/100%	23/88%	24/92%
researchers	do	not	yet	know	for	sure	whether
8/31%	21/81%	24/92%	23/88%	16/62%	8/31%	7/27%	2/8%
the	noise	harms	or	disturbs	the	animals,	but
16/62%	20/77%	14/54%	7/27%	5/19%	16/62%	20/77%	17/65%
they	know	the	animals	can	hear	it.	
0/0%	5/19%	10/38%	19/73%	11/42%	9/35%	12/46%	

Appendix 4A. Test Sheets Used in the Pre- and Post-Tests of Study 2

【Pre-test (All)】

3 番の音声を聞いて、聞こえてくる**英語**を書き取ってください。各文の先頭の語だけ示してありますので、それに続けて書いてください。制限時間は **5 分**です。わからない箇所があるときは、**途中で止めて繰り返し聞いても構いません**。**字ははっきり**書いてください。

It _____

_____

Many _____

_____

A _____

_____

According _____

_____

In _____

_____

However, _____

_____

In _____

_____

This _____

_____

_____

指示があるまで、次のページに進まないでください。

【Training (Accelerated Speech Dictation Group)】

4 番の音声を聞いて、聞こえてくる**英語**をディクテーションしてください。各文の先頭の語だけ示してありますので、それに続けて書いてください。活動の時間は **15 分**です。わからない箇所があるときは、**途中で止めて繰り返し聞いても構いません**。

It _____

_____

Many _____

_____

A _____

_____

According _____

_____

In _____

_____

However, _____

_____

In _____

_____

This _____

_____

_____

指示があるまで、次のページに進まないでください。

【Training (Shadowing Group)】

4 番の音声を聞いて、聞こえてくる**英語**をシャドーイングしてください。各文の先頭の語だけ示してあります。活動の時間は **15 分**です。わからない箇所があっても、**途中で止めずに、最後まで通して**シャドーイングしてください。時間内にできるだけ何回も練習してください。

難しいと感じる人は、**4 番(やや遅い)**の音声を使って練習し、おおむね言えるようになったら 4 番の音声を再び使用してください。

It

Many

A

According

In

However,

In

This

指示があるまで、次のページに進まないでください。

【Post-test (All)】

5 番の音声を聞いて、聞こえてくる**英語**を書き取ってください。各文の先頭の語だけ示してありますので、それに続けて書いてください。制限時間は **5 分**です。わからない箇所があるときは、**途中で止めて繰り返し聞いても構いません**。**字ははっきり**書いてください。

It _____

_____

Many _____

_____

A _____

_____

According _____

_____

In _____

_____

However, _____

_____

In _____

_____

This _____

_____

_____

【Questionnaire (All)】

手順 4 で行った活動(シャドーイング or ディクテーション)は聞き取り能力の向上に、どれくらい効果的だと思いますか？次の基準のうち当てはまる数字 1 つに○をつけてください。

ほぼない	あまりない	どちらともいえない	ややある	とてもある
1	2	3	4	5

今回行った活動について自由に感想を書いてください。

ご協力ありがとうございました。

Appendix 4B. Feedback of the ASD Group on the Training

A) 3 番の音声と比べて、速度が圧倒的に速いが、文の構成がつかみやすく聞き取れない単語が聞き取れるようになった。

B) 手順 3（プレテストのこと）で内容を少し覚えていたので 3 よりも今回の方が多く書けた。

C) 細かい the などは聞き取りにくかった。（早い方）

D) 最初にゆっくりなスピードでリスニングをしたため、速いスピードで聞いた時がさほど聞き取れないことはなかった。しかし、遅いスピードで聞き取れない単語はスピードが速くなったとたんに埋もれてしまった。

E) 単語が全部くっついて聞こえたので、理解するのが大変だった。

F) 速い方が聞き取りやすく感じた（たぶん手順 3 番でも聞いたからかも）。

G) 速いのも普通のも止めながら出来るので、そんなに差が無かった気がする。

H) ディクテーションはとても神経を使うので疲れた。でも続ければ身になると思った。

I) 一度聞いた文章なのでスピードが上がっても意味は理解できる。全体的に文章を聞けたのでよかった。

J) 速いと聞き取りにくいが、逆にポイントとなる単語が聞き取りやすかった。

K) ディクテーションをやってみて、自分がどれだけ聞き取れないのかわかりました。今後の勉強に生かしていきたいと思います。

L) 聞いていくうちになんとなく聞き取れるようになりました。

M) 1 回聞いた文なので少しは把握することができたが、かなりの速さだったのでとても難しく感じた。

N) 難しかった。何回か聞くうちに話す速さに少しずつ慣れていったと思う。

Appendix 4C. Feedback of the SH Group on the Training

O) シャドーイングは言った英文を聞き取って自分で繰り返し言うことが必要になってくるので、英語を聞き取ろうと努力するので聞き取る力がつくと思います。

P) 3 番はスピードが速くてところどころしかついていけなかった。

Q) 3 番だと速くてついていけず 4 番でずっとやっていた。4 番でも若干速さを感じるがよく集中すれば聞き取れた。知らない単語があると分からなくなってしまう。

R) ディクテーションをするにつれて分からなかった細かい部分が聞き取れるようになってきた。

S) 3 番の後に 4 番を聞くと比較的シャドーイングしやすかった。

T) 交互にやっていたが、言えるような段階まではいかなかった。

U) 4 番を結構聞いていたがシャドーイングは追いかけるのがすごく大変であったが何とかついていけた。

V) 回数を重ねていくうちに文を自然に覚えられた。また、次第に次に来る文を予測することもできた。

W) いまいち聞き取れないところがあった。

X) 速度の違いにより、4 から 3 を聞くと少し聞き取れるようになった。

Y) 文が長くて追いつけなかった。

Z) 無回答

AA) 知っている語彙でのディクテーションはやりやすいが、知らない単語がある中でのディクテーションは分かりにくい。

Appendix 5. Test Sheet Used in Study 3 (with answers)

【サンプル問題】 監督者の指示を聞き、(1) (2)に答えなさい。

(1) A man painting the gate . (3)語

(2) A woman is watering a plant .

A. 英文を聞き、下線部に入る語の数を数えて、その数字を( )に書き入れなさい。

(1) He's working on a car. (4)語	61 名	(67.0%)
(2) She's painting in a studio. (4)語	61 名	(67.0%)
(3) One man is writing on a notepad. (4)語	51 名	(56.0%)
(4) The man is grasping a door handle. (4)語	40 名	(44.0%)
(5) Some tables are shaded by umbrellas. (3)語	65 名	(71.4%)
(6) Trees separate the lake from some buildings. (5)語	27 名	(29.7%)
(7) A man is adjusting a piece of equipment. (5)語	36 名	(39.6%)
(8) Bicycles are parked along a painted line. (5)語	45 名	(49.5%)
(9) Pots and pans have been piled in a drying rack. (5)語	22 名	(24.2%)
(10) Overhead wires are suspended near an unfinished structure. (5)語	41 名	(45.1%)

B. 英文を聞き、下線部に入る語を書き入れなさい。

(1) He's cooking food at the stove.	(83.3%)
(2) The women are looking at an o-pen bin-der.	(58.8%)
(3) Some chairs have been set out-side.	(52.0%)
(4) A man is using a control panel.	(85.7%)
(5) There are windows on one side of a passageway.	(62.9%)
(6) A vehicle's rear door is raised.	(12.5%)
(7) Pottery is being exhibited on a shelving unit.	(19.0%)
(8) A bridge extends towards a domed building.	(28.8%)

(Educational Testing Service, 2012, pp. 6-8, 92-93)

Appendix 6A. Revised Version of the Written Reproduction Task (RPD-R)

文章を 2 度ずつ聞き、下線部に語句を書き入れてください。「 / 」のところにそれぞれ 5 秒のポーズ（無音時間）があります。空いているスペースにメモをとっても構いません。

練習

Last summer, Kumiko went to New York _____ her uncle. / They went to _____ together. / On the last day, they _____, / and she bought _____ for her parents. /

本番

No. 1 Everyone, thank you for all your hard work _____. / You _____ practiced hard, / and all the parents said that _____ the music last night. / I think it was _____ successful concert ever. / I know that you're _____ practicing every day, / but please don't forget that we're _____ the city music festival next month. /

No. 2 Nancy went _____ this weekend. / When she returned home, she _____ telephone messages / and was surprised to find a message _____ at her part-time job. / He had called to find out why Nancy _____ in to work on Saturday. / Nancy realized that she had forgotten to ask him for _____. / She feels bad, so she is going _____ and apologize. /

No. 3 As well as being _____ for humans, / fish are used to make agricultural products _____ fertilizer. / As a result, several fish _____ threatened by overfishing. / Fish Fight is a U.K. campaign trying to _____. / The campaign hopes to reduce _____ overfished species / by encouraging consumers to buy species _____ familiar with. / Fish Fight also hopes to _____ of fish in commercial fishing. / Disposal happens because fishing boats _____ exceeding quotas. / This means dead fish are _____ into the water despite being edible. /

Appendix 6B. Revised Version of the Word Count Task (CNT-R)

文章を 2 度ずつ聞き、空所に入る単語の数を（　　）内に書いてください。1 度目のみ「 / 」のところにそれぞれ 3 秒のポーズ（無音時間）があります。空いているスペースにメモをとっても構いません。

練習

Next week, students from France will (　　). / My class will (　　) around the school. / I'm (　　) them. / I also want (　　) speaking French. /

本番

No. 1　On Sunday, the morning weather report said it would (　　), / so Peter decided to go (　　)./ He went without (　　) / and (　　) enjoying the sunshine. / However, in the afternoon, (　　), / and Peter began (　　) very cold. / Next time (　　) hiking, / he will take a jacket, (　　). /

No. 2　Paul wanted to take a short trip to (　　) in California, / but he was (　　) at work. / One day, he talked to his co-worker Maria about (　　). / She offered to do his work for (　　). / Paul was able to take (　　), / and he bought Maria (　　) to thank her for helping him. /

No. 3　Max is (　　) an amateur theater group called the Mapleton Players. / They recently (　　), / and she has helped the group members (　　) their acting. / She also has many good ideas for (　　) more interesting. / Recently, the sizes of the audiences have (　　). / Max hopes that, in the future, the group may be able to perform (　　). /

No. 4　There is (　　) of shark called the thresher shark. / In 2010, scientists discovered that they have an (　　) for hunting. / Thresher sharks have long, (　　), / and they use them to (　　). / The fish cannot move after they (　　), / so the shark can easily catch (　　). / Thresher sharks are not dangerous to people, however, and are afraid (　　). /

No. 5　Guam is a beautiful (　　) island, / but it is not a good place for those with (　　). / After being accidentally brought on (　　) around 60 years ago, / brown tree snakes spread (　　), / eating native birds and (　　). / Many kinds of spider (　　) thriving / because the native birds that ate them were made extinct (　　). / Now, the snakes frequently (　　) into people's homes, / where (　　) nasty bites. / They also regularly (　　), causing blackouts. /

Appendix 6C. Listening Strategy Survey Used in Study 4

(1)〜(16)のそれぞれはテストを受けたときのあなたのリスニングの仕方にどれくらい当てはまりますか。【　　】内の基準にもとづき、あてはまる数字 1 つに○をつけてください。

【 1 あてはまらない　2 あまりあてはまらない　3 どちらとも言えない　4 ややあてはまる　5 あてはまる 】

(1) 理解したことをもとに話の展開を推測しながら聞いた。　〔 1　2　3　4　5 〕

(2) 聞こえた単語や語句を頼りに、内容を想像しながら聞いた。　〔 1　2　3　4　5 〕

(3) トピックに関して自分が持っている知識を使いながら聞いた。　〔 1　2　3　4　5 〕

(4) 映像化しながら聞いた。　〔 1　2　3　4　5 〕

(5) 英文の内容の要点をまとめながら聞いた。　〔 1　2　3　4　5 〕

(6) 英文の意味を日本語に訳しながら聞いた。　〔 1　2　3　4　5 〕

(7) 英文をシャドウイング(頭のなかで繰り返すこと)しながら聞いた。　〔 1　2　3　4　5 〕

(8) キーワードを探しながら聞いた。　〔 1　2　3　4　5 〕

(9) それぞれの英文を正確に理解しようとして聞いた。　〔 1　2　3　4　5 〕

(10) 英文の細かい内容まで理解しようとして聞いた。　〔 1　2　3　4　5 〕

(11) 大まかな内容を理解しようとして聞いた。　〔 1　2　3　4　5 〕

(12) 複数形の s や過去形の ed など細かい音にも気をつけて聞いた。　〔 1　2　3　4　5 〕

(13) 1 つ 1 つの単語に注意して聞いた。　〔 1　2　3　4　5 〕

(14) 英語の句のまとまりを意識しながら聞いた。　〔 1　2　3　4　5 〕

(15) 英文の文法構造に注意して聞いた。　〔 1　2　3　4　5 〕

(16) 自分の理解があっているかどうかを考えながら聞いた。　〔 1　2　3　4　5 〕

Appendix 7A. Breakdown of the Participants' Majors and Grades in Study 5

Experimental Group 1 (Accelerated Speech Dictation)

	Child Education	Sports& Health	English Education	Psychology	Total
1st grade	5	6	8	0	19
2nd grade	0	1	3	0	4
3rd grade	3	1	0	0	4
Total	8	8	11	0	27

Male: Female = 12:15

Experimental Group 2 (Shadowing)

	Child Education	Sports& Health	English Education	Psychology	Total
1st grade	8	9	2	0	19
2nd grade	1	0	2	1	4
3rd grade	4	0	2	0	6
Total	13	9	6	1	29

Male: Female = 20:9

Control Group (No training)

	Child Education	Sports& Health	English Education	Psychology	Total
1st grade	6	-	1	-	7
2nd grade	4	-	1	-	5
3rd grade	0	-	9	-	9
4th grade	2	-	5	-	7
Total	12	-	16	-	28

Male: Female = 8:20

Appendix 7B. Listening Proficiency Test Used in Study 5

＜内容理解問題（第１部）＞
対話を聞き、その質問に対して最も適切なものを①～④の中から一つ選んでマークしてください。
英文はすべて一度しか読まれません。メモはとって構いません。

No. 1 ① Ask Anna to call them. ② Tell Anna when the movie starts.
③ Wait for five more minutes. ④ Go and watch the movie.

No. 2 ① She couldn’t find the swimming pool.
② She had no time to go shopping.
③ She lost her wallet.
④ She couldn’t enjoy any outdoor activities.

No. 3 ① He stayed out late. ② He kept playing soccer after 9 o’clock.
③ He left his homework at school. ④ He forgot to make dinner.

No. 4 ① Wash the dishes. ② Clean their bedroom.
③ Cook a meal. ④ Set the table.

No. 5 ① She got some advice from her colleague.
② She spent a lot of time preparing.
③ She knows a lot about making presentations.
④ She invited Linda to the presentation.

No. 6 ① She is the man’s mother.
② She lives far away from her parents.
③ She makes less money than the man.
④ She plans to go abroad during the holidays.

No. 7 ① Taking a long vacation. ② Working for her uncle.
③ Enjoy outdoor activities. ④ Starting a new business.

No. 8 ① He was studying for the science quiz. ② He woke up 20 minutes late.
③ He went back home to get his bag. ④ He was searching for this book.

No. 9 ① She was late for work. ② She took a bus later than usual.
③ She usually goes to work with Larry. ④ She missed a chance to see Larry.

No. 10 ① In the sales department. ② In a department store.
③ In the marketing department. ④ In a different company.

＜内容理解問題（第 2 部）＞
英文を聞き、その質問に対して最も適切なものを①～④の中から一つ選んでマークしてください。
英文はすべて一度しか読まれません。メモはとって構いません。

No. 16 ① She was asked to meet Sandy’s clients.
② She had some unexpected work.
③ Sandy had to take a client to lunch.
④ Sandy’s boss wanted to come along.

No. 17 ① By borrowing it from his friend. ② By buying it when it is available.
③ By making a reservation. ④ By waiting for it to be delivered.

No. 18 ① Her friend had a big family.
② She got used to the culture quickly.
③ Restaurants had menus for elderly people.
④ There was a different custom for eating.

No. 19 ① It is not acceptable in some places. ② It is popular with tourists.
③ It is grown all over the world. ④ It is known for its sour favor.

No. 20 ① By making a ball out of sand. ② By cutting and smoothing rocks.
③ By polishing rocks in a river. ④ By rolling rocks along the ground.

No. 21 ① He couldn’t reserve a hotel. ② He broke his leg.
③ His son had an accident. ④ He was busy at work.

No. 22 ① Great paintings by Prince Albert. ② The way clothes have changed.
③ Different styles of music video. ④ Famous books about Victorian art.

No. 23 ① Tell a store clerk about the boy. ② Pay for the jacket and pants.
③ Go to the service desk. ④ Visit the kids’ park.

No. 24 ① He couldn’t speak Japanese. ② He got lost while sightseeing.
③ He lost one of his belongings. ④ He couldn’t meet his friend.

No. 25 ① Study at a high school overseas.
② Stay with American families.
③ Serve Japanese food at a party.
④ Teach their host families Japanese cooking.

No. 25 に答えた人は下記の質問に答えてください。
これまでの問題を過去に学習したことがありますか？ （ ある ・ ない ）

Appendix 7C. Online Weekly Journal Used in Study 5

リスニング学習ジャーナル

成績に入りますので、毎回忘れずに入力してください。

（共有なし）

アカウントを切り替える

あなたの学籍番号は？（半角で入力）

回答を入力

あなたの氏名は？

回答を入力

あなたが選んだ活動は？ ※毎回同じ活動を行ってください。

- 速聴ディクテーション
- シャドーイング

今日の日付は？

日付

年 /月/日

今日のリスニング学習の感想を入力してください。（例：どのように活動を進めたか、聞き取りが難しかったところ、成長したと感じる点）

回答を入力

Appendix 7D. Questionnaire Given to the Experimental Groups in Study 5

アンケート

1. 授業であなたが取り組んでいた活動を○で囲んでください。

〔　シャドーイング　・　速聴ディクテーション　〕

2. 上記の活動を継続して行ったことで、どのようなリスニングの力が伸びたと思いますか？次の(1)〜(16)について、下の【　　】の基準にしたがい、最も近いと思う記号にそれぞれ○をつけてください。

【　A 伸びていないと思う　　B 少し伸びたと思う　　C 伸びたと思う　】

(1) 理解したことをもとに話の展開を推測しながら聞く力。　〔　A　B　C　〕
(2) 聞こえた単語や語句を頼りに、内容を想像しながら聞く力。　〔　A　B　C　〕
(3) トピックに関して自分が持っている知識を使いながら聞く力。　〔　A　B　C　〕
(4) 映像化しながら聞く力。　〔　A　B　C　〕
(5) 英文の内容の要点をまとめながら聞く力。　〔　A　B　C　〕
(6) 英文の意味を日本語に訳しながら聞く力。　〔　A　B　C　〕
(7) 英文をシャドウイング(頭のなかで繰り返すこと)しながら聞く力。　〔　A　B　C　〕
(8) キーワードを探しながら聞く力。　〔　A　B　C　〕
(9) それぞれの英文を正確に理解しようとして聞く力。　〔　A　B　C　〕
(10) 英文の細かい内容まで理解しようとして聞く力。　〔　A　B　C　〕
(11) 大まかな内容を理解しようとして聞く力。　〔　A　B　C　〕
(12) 複数形の s や過去形の ed など細かい音にも気をつけて聞く力。　〔　A　B　C　〕
(13) 1 つ 1 つの単語に注意して聞く力。　〔　A　B　C　〕
(14) 英語の句のまとまりを意識しながら聞く力。　〔　A　B　C　〕
(15) 英文の文法構造に注意して聞く力。　〔　A　B　C　〕
(16) 自分の理解があっているかどうかを考えながら聞く力。　〔　A　B　C　〕

3. シャドーイングまたは速聴ディクテーションでトレーニングした感想を自由に書いてください。

以上でアンケートは終わりです。

Appendix 7E. Questionnaire Given to the Control Group in Study 5

アンケート

1. 下記の(1)〜(16)はリスニングに関係のある能力です。それぞれについてあなたはどれくらい重要だと思いますか。下の【　　】の基準にしたがい、あなたの考えに最も近い記号に○をつけてください。

【　×＝重要でない　　△＝どちらとも言えない　　○＝重要だと思う　】

(1) 理解したことをもとに話の展開を推測しながら聞く力。〔　×　△　○　〕
(2) 聞こえた単語や語句を頼りに、内容を想像しながら聞く力。〔　×　△　○　〕
(3) トピックに関して自分が持っている知識を使いながら聞く力。〔　×　△　○　〕
(4) 映像化しながら聞く力。〔　×　△　○　〕
(5) 英文の内容の要点をまとめながら聞く力。〔　×　△　○　〕
(6) 英文の意味を日本語に訳しながら聞く力。〔　×　△　○　〕
(7) 英文をシャドウイング(頭のなかで繰り返すこと)しながら聞く力。〔　×　△　○　〕
(8) キーワードを探しながら聞く力。〔　×　△　○　〕
(9) それぞれの英文を正確に理解しようとして聞く力。〔　×　△　○　〕
(10) 英文の細かい内容まで理解しようとして聞く力。〔　×　△　○　〕
(11) 大まかな内容を理解しようとして聞く力。〔　×　△　○　〕
(12) 複数形の s や過去形の ed など細かい音にも気をつけて聞く力。〔　×　△　○　〕
(13) 1 つ 1 つの単語に注意して聞く力。〔　×　△　○　〕
(14) 英語の句のまとまりを意識しながら聞く力。〔　×　△　○　〕
(15) 英文の文法構造に注意して聞く力。〔　×　△　○　〕
(16) 自分の理解があっているかどうかを考えながら聞く力。〔　×　△　○　〕

2. 前回のテストから今日までリスニングの学習をしましたか？ある人はどのような学習をしていたか、例にしたがって答えてください（ない人は何も書かなくて大丈夫です)。

例：「英語コミュニケーション IA」の授業で毎週ネイティブの先生の英語を聞いたり、リスニングの問題に取り組んだりした。

以上でアンケートは終わりです。ご協力ありがとうございました。

Appendix 8A. Most Frequent 150 Words Observed in the Responses by the ASD Group

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
思う	122	前	13	動詞	7
聞き取れる	116	通常	13	メモ	6
単語	106	問	13	英	6
聞き取る	94	力	13	解ける	6
リスニング	82	違う	12	検	6
聞く	80	苦手	12	細かい	6
感じる	61	集中	12	書ける	6
難しい	59	正解	12	徐々に	6
問題	43	正答	12	全体	6
今回	40	増える	12	長い	6
ディクテーション	39	読む	12	途中	6
速聴	39	最後	11	答える	6
部分	39	出る	11	付く	6
慣れる	34	書き取る	11	文章	6
少し	33	スペル	10	無い	6
分かる	32	活動	10	I	5
理解	31	気	10	TOEFL	5
発音	28	上がる	10	もう少し	5
書く	26	逃す	10	キーワード	5
多い	26	英文	9	慣らす	5
文	25	言う	9	気づく	5
聴く	22	重ねる	9	原因	5
練習	22	重要	9	高める	5
意識	21	人	9	実際	5
内容	20	特に	9	受ける	5
英語	19	必要	9	初回	5
耳	19	復習	9	少ない	5
自分	19	テスト	8	先	5
頑張る	18	形	8	全部	5
出来る	18	行う	8	増やす	5
前回	18	今後	8	多く	5
速い	18	上手い	8	段々と	5
知る	18	成長	8	追いつく	5
スピード	17	答え	8	倍速	5
今日	17	比べる	8	A	4
聞こえる	17	聞ける	8	d	4
トレーニング	16	ミス	7	スクリプト	4
音声	16	意味	7	フレーズ	4
取り組む	16	会話	7	ポイント	4
早い	16	学習	7	安定	4
速度	16	繰り返す	7	一つ	4
名前	16	繋がる	7	一語	4
以前	15	今	7	英単語	4
取れる	15	時間	7	家	4
音	14	次回	7	課題	4
見る	14	取る	7	改善	4
授業	14	初めて	7	繋げる	4
続ける	14	大まか	7	減らす	4
間違える	13	短縮	7	考える	4
最初	13	調子	7	高校	4

Appendix 8B. Most Frequent 150 Words Observed in the Responses by the SH Group

抽出語	出現回数	抽出語	出現回数	抽出語	出現回数
単語	116	意識	13	もう少し	6
思う	115	目	13	オーバーラッピング	6
聞き取れる	100	以前	12	ディクテーション	6
聞く	87	取れる	12	英文	6
感じる	78	成長	12	解ける	6
リスニング	65	知る	12	活動	6
聞き取る	63	通常	12	苦手	6
シャドーイング	60	スピード	11	繋げる	6
難しい	60	気	11	口	6
問題	47	形	11	次回	6
少し	45	比べる	11	重ねる	6
聴く	37	聞ける	11	熟語	6
理解	35	力	11	初め	6
今回	34	練習	11	初めて	6
多い	34	音声	10	正解	6
内容	34	考える	10	正確	6
前回	32	行う	10	全然	6
分かる	31	逃す	10	答え	6
音	30	復習	10	能力	6
今日	29	会話	9	文法	6
出来る	29	語彙	9	無い	6
頑張る	28	効果	9	話	6
部分	28	時間	9	やり方	5
聞こえる	27	増える	9	テスト	5
文	25	倍	9	一つ一つ	5
英語	24	付く	9	家	5
見る	22	学習	8	過去	5
意味	21	繰り返す	8	回答	5
慣れる	21	固有名詞	8	簡単	5
発音	21	最後	8	継続	5
読む	20	重要	8	言える	5
最初	19	上がる	8	言葉	5
間違える	18	早い	8	今	5
耳	18	通す	8	取る	5
前	18	特に	8	大体	5
速い	17	入る	8	注意	5
速度	17	倍速	8	比較的	5
細かい	16	良い	8	標準	5
自分	16	解く	7	表現	5
取り組む	16	確認	7	複数	5
出る	16	感じ	7	文字	5
繋がる	15	今後	7	話せる	5
集中	15	実感	7	S	4
選択肢	15	上手い	7	あと	4
問	15	全体	7	その後	4
スクリプト	14	続ける	7	スムーズ	4
言う	14	大切	7	完璧	4
次	14	遅い	7	慣らす	4
頭	14	余裕	7	間違う	4
トレーニング	13	話す	7	機会	4