1. Introduction

Processing of speech sounds for meaning is generally divided into three types—bottom-up, top-down, and interactive. Early research proposed a *bottom-up processing* model in which listeners build understanding starting with the smallest differentiable units of the acoustic message (that is, phonemes), which are combined into words, phrases, clauses, and sentences (Carrel, 1988; Flowerdew and Miller, 2005; Rost, 1990).

Others argue that effective listeners use *top-down processing*. In this model, listening is viewed as an active process in which the listener must seek necessary information (Ross, 1975; Rubin 1994) rather than just passively analyzing speech as a series of sounds, as in the bottom-up model.

More recently, the idea has gained support that listening comprehension is an interplay of information at different levels, and listeners have been argued to employ both bottom-up and top-down processes simultaneously (O’Malley and Chamot, 1990). A number of researchers have argued that effective listeners most often use this *interactive processing* method (Buck, 2001; Field, 2008; O’Malley, Chamot, and Kupper, 1989).

A number of studies have argued for the superiority of the interactive processing model. Many of these have also argued for the necessity of listening comprehension training to promote both bottom-up and top-down skills (Ellsworth, 2006; Flowerdew and Miller, 2005; Higuchi, 1998; Orii-Akita, to appear), or what will be referred to as the *interactive teaching method* in this paper.

However, while numerous studies have supported the potential of interactive teaching, few have examined or theorized the underlying nature of the listening process. A good theory of listening would be able to shape the instructional principles of listening instruction in a way that would allow better pedagogical methods and listening instruction to be derived from them (Ellsworth, 2006: 78).

This paper examines how human cognition functions in the processing of speech in first language (L1) and second language (L2) listening. Based on the cognitive theory that I will elaborate, I then present the instructional principles of a listening course, based on the argument that interactive teaching yields better results than other teaching methods.

To begin with, in section 2, we will consider where humans extract the knowledge needed to process language in L1 and L2 listening. The use of *declarative memory*, part of long-term memory, is the mechanism of the slow, conscious processing that takes place in L2 listening. I will argue that shadowing training
and prosody instruction will benefit L2 learners by allowing them instead or in addition to use *procedural memory* (which is an automatic and unconscious process) to process language, as L1 listeners do. Section 3 elaborates the progression of information processing in the working memory model. In section 4, I argue that a cooperative environment of top-down processes and bottom-up processes is essential for linguistic information processing; rather than a single direction only, data processing must occur in both bottom-up and top-down directions. In section 5, it is suggested that prosody instruction, along with shadowing training, need to be included in listening comprehension teaching to promote bottom-up processing; and in section 6, it is suggested that listening strategy instruction is effective in promoting top-down processing. A short conclusion sums up.

2. Long term memory and L2 listening processes

First, I will give a general outline of how various stimuli are transmitted to the brain. Next, I will examine the contents of long-term memory and the types of such content. Iwata (2011, part 3, chapter 2) and Fukunaga (2009, chapter 2) serve as references regarding the structure of neurons and their transmission process.

Information that involves an electrical signal called an action potential that is sent from the hippocampus, which is one of the two major components that comprise the limbic system, stimulates neurons in the cerebral cortex. Neurons are independent units, and very small gaps are known to exist between them. A neuron is composed of a central part that is called the soma and two types of projections that are called the dendrites and the axon. The soma contains the nucleus, which is the core of the neuron and which houses genetic information. The dendrites, which extend from the soma in large numbers like tree branches, receive stimuli and information from other neurons. Also extending from the soma is a single axon, which sends information from the soma to other neurons. The terminus of the axon, which is the sender of information, and the tips of the dendrites, which are the receivers, meet in an area called the synapse. The role of the synapse is to receive information. Neurons are separated from each other by a 20–50-nanometer gap that is called the synaptic cleft.

Neurons are both filled with and surrounded by a solution containing various ions. A neuron is enveloped in a cell membrane that is impermeable to external ions. However, when a neuron receives information, that stimulus will open a channel (hole) in the membrane that transmits certain ions, thus changing the membrane into a state that permits the flow of ions inside and outside the neuron. At this time, an electrical signal called an action potential occurs; information is transmitted inside the neuron in the form of an electrical signal. However, as previously stated, neurons are independent units; the action potential electrical signal that is transmitted by the axon cannot cross over to the next neuron. Therefore, the form of the information is changed, and it is transmitted as a chemical substance. The information is then reconverted into an electrical signal in the next neuron and transmitted inside that neuron. The connection responsible for transmitting information from one neuron to the next is the previously described synapse. A chemical substance is released from the synapse, which is on the tip of the axon of the neuron that is the sender of information. This substance is transmitted to the receptor of the neuron that is acting as the receiver, thus forming a new connection between the neurons. As these synaptic connections stabilize, they are thought to remain as memories.
Next, I will discuss the classifications of long-term memory that are based on the such memories, and examine the long-term memory knowledge which L1 and L2 listeners turn to when processing speech.

Long-term memory stores information over a long period; in some cases, this information may be stored semi-permanently. On the basis of the nature of the information that is stored, long-term memory is divided into two major types called declarative memory and procedural memory (Squire & Zoro-Morgan, 1991).

Declarative memory, which is also known as explicit memory or memories that one is conscious of having or not having, refers to memories that can be retrieved consciously. Declarative memory is often further classified into episodic memory and semantic memory. Episodic memory (Tulving, 1972) is centered on autobiographical experiences and includes events that occur in society. Semantic memory (Collins & Quillian, 1969) includes fact-related memories and conceptual knowledge, the important parts of which form a mental lexicon of linguistic knowledge. Semantic memory has been hypothesized to retain knowledge regarding the visual characteristics of objects (Gathercole & Alloway, 2009: 13).

Procedural memory, in contrast, refers to implicit memories that one is not conscious of having or not having; this includes memories of how to do things and physically learned automated skills. For a detailed structure of long-term memory, please refer to Hakoda, Tsuzuki, Kawabata, & Ogiwara (2010, chapter 6).

Information regarding language is believed to be stored first in semantic memory. The database that is related to linguistic knowledge includes the following types of knowledge: phonetic knowledge, such as sound change; grammar; vocabulary; morphemes; context (schema); paralanguage (emotional information); body language; and so on. When speech stimuli are received and speech is perceived, the above information is elicited from long-term memory and used to advance comprehension. For example, when a second-language (foreign language) learner sees a written word and tries to remember its meaning, it is retrieved from explicit semantic memory.

However, for native speakers, linguistic knowledge is not stored in explicit semantic memory; rather, implicit procedural memory is used. As infants learn their native language, it is thought that, initially, they consciously retain the speech they have learned as semantic memory. However, as they repeat identical or similar speech acts, they begin to operate automatically without entering consciousness. When this level is reached, it is generally called technical skill, and it involves the conversion of linguistic knowledge into implicit memory. The conversion of semantic memory, which is explicit knowledge, into procedural memory, which is implicit knowledge, is generally considered to be an extremely slow process. Through repeated encounters with identical or similar information, the process of retrieving knowledge from long-term memory speeds up and eventually becomes automatic. This type of knowledge that can be retrieved and used unconsciously can be called procedural memory. In order to manipulate language freely, explicit knowledge must be converted into implicit knowledge. For native speakers who are exposed to a large volume of linguistic input for many years beginning from birth, this conversion is not so difficult. However, for second-language learners, particularly those who have few chances to encounter the target foreign language (i.e., those who have low-input volume), this conversion often does not proceed well. For example, when concepts such as the third-person singular “s” and subject-predicate agreement are learned as procedural knowledge, the grammar knowledge takes time to remember, and fluent speech is not possible. Such concepts are also often difficult to apply online, even when they are known. In order to speak a second language fluently, it is desirable that the learned semantic memory be changed into a state
where it can be used unconsciously (procedural memory). However, foreign language learners often cannot accumulate sufficient linguistic experience for semantic memory to be converted into procedural memory. In other words, foreign language information that only exists as mere knowledge cannot actually be applied in linguistic activity. Therefore, an aid is necessary for converting explicit knowledge into implicit knowledge. One such method is shadowing. Shadowing is the process of repeating a native speaker’s speech. Through repeated practice, the cognitive load that is necessary to perform shadowing gradually decreases and automation is expected to progress incrementally. When the cognitive load for the repetition process is decreased, cognitive load is expected to be allotted to the internalization of semantic comprehension and the words used.

3. Working memory and human cognitive processing

How do humans receive sensory input, such as audition and vision, from sensory organs? What types of processes process and store such information? Since the 1950’s, many memory experiments have been conducted in order to answer these questions. In addition, there have been various attempts to systematize the answers (Broadbent, 1958, etc.).

These concepts were combined by Atkinson & Shifrin (1968) into a model of short-term memory (the A & S model), which posits three stores of information. In this model, sensory input is first sent to a buffer called sensory memory, which is also called the sensory register and which retains information temporarily. At this point, the information is almost entirely unprocessed. In the case of audition, the information is sent to echoic memory, and it is retained for up to 3–4 seconds. In contrast, in the case of vision, the information is retained in iconic memory for only 0.2–0.3 seconds. Next, the core processes of rehearsal, coding, and retrieval are executed in the short-term store, which is also called short-term memory. At this point, information is retrieved from the knowledge base in the long-term store (which is also called long-term memory), and it is processed. Information is retained in the short-term stores for a short period of up to approximately 15 seconds (Loftus & Loftus, 1976). Beyond that period, only a portion of this information is sent to the long-term store as necessary, and it is considered to be retained semi-permanently.

Another model was later proposed by Baddeley & Hitch (1974). This working memory model (the B & H model) added a function that performs control tasks compared to the previously proposed retention tasks. This model was proposed as an expansion of the short-term memory portion of the A & S model into a process that has a multilayered structure called working memory. The initial working memory model consisted of a supervisory system called the central executive, or central control, and slave systems called the phonological loop and the visuospatial sketchpad. This model is a multicomponent model in which higher-level cognition is divided into verbal and visual characteristics, both of which are controlled by the central executive. It is considered a radical revision of simplified modal models, such as the A & S model (Osaka, 2008: 5).

In 2000, another slave system called the episodic buffer was added to the two previously stated slave systems, thus expanding the multicomponent model to four components (Baddeley, 2000). This model is still widely accepted today as the most comprehensive model.

Here, I would like to elaborate again on the course of information processing in the working memory model that began with Baddeley & Hitch (1974) and developed into that of Baddeley (2000). This model
consists of the central executive, which is the core that integrates information processing, and the following three slave, or subordinate, systems: the phonological loop, the visuospatial sketchpad, and the episodic buffer (Baddeley, 2000, 2002).

The phonological loop is responsible for verbal information. This system retrieves and processes necessary information from various knowledge bases in long-term memory while temporarily retaining information that can be manipulated as spoken language. Baddeley (2002: 86) has stated that the span of information that can be temporarily stored phonologically is the number of words and letters that can be repeated within two seconds.

The phonological loop is further divided into two systems: the phonological short-term store and the subvocal rehearsal. The phonological short-term store system passively retains phonological information as it is, while the subvocal rehearsal system rehearses this phonological information internally. In other words, subvocal rehearsal consciously and actively repeats the phonological information. Through this repetition, phonological information that would normally be forgotten in two seconds can be retained in the phonological loop for a longer period. With this function, deeper information processing is considered possible.

Auditory verbal information is imported into the phonological loop, and it reaches the phonological short-term store system directly. Visually perceived verbal information, such as when one is looking at words or letters, is first inputted temporarily into the subvocal rehearsal process, and it is then retained in the phonological short-term store in the phonological loop. Baddeley has stated that one reason for establishing the concept of subvocal rehearsal is to handle visual information (Baddeley, 2002: 86). Furthermore, Kadota (2007: 151) has stated that Baddeley proposed the phonological loop subsystem within working memory because of the idea that textual information is not stored visually but rather is retained by pronouncing it to oneself silently. That is, information is retained and processed on a foundation of phonological information. In his own English sentence memorization experiment of Japanese learners of English (Kadota, 1997), Kadota stated that the same results were obtained (Kadota, 2007: 154).

The visuospatial sketchpad is the system that is responsible for retaining and processing nonverbal information, such as information that cannot be coded phonologically. This system is responsible for the short-term retention and processing of diagrams and other visuospatial information.

The episodic buffer is a system that uses information of self-experience in order to retain and process information as it relates to oneself. It can be considered an interface that links information from the phonological loop, the visuospatial sketchpad, and long-term memory (Baddeley, 2002: 93).

The above three processes are integrated by the central executive. On the basis of the idea that human cognitive capacity is limited, the responsibility of the central executive is to consider and control how these cognitive resources are distributed. In particular, when simultaneous processing is necessary, the central executive performs the following functions: deciding which processing tasks to assign to which systems, controlling the cognitive capacity that is allotted to each system, and switching between processes as necessary. It has also been suggested that, when a certain task must be completed, the central executive suppresses the diversion of attention to any unnecessary processing. However, many details regarding this function are unknown (Osaka, 2008: 130). Interestingly, the switching and renewal tasks that are performed by the central executive depend upon the function of the phonological loop. According to Funabashi, Saito,
& Osaka (2003), when forced to switch between tasks (an addition task and a subtraction task), subjects tend to take longer to switch between tasks when articulation is suppressed. That is, efficient execution is made possible by the task goal retention function that is provided by the phonological loop (quoted in Kadota, 2006: 19).

Human cognitive capacity is limited in the overall functions that support the retention and processing of information. In dual tasks, which is the parallel execution of two different tasks, Baddeley & Hatch (1974) have considered that the processing resources of working memory are shared and distributed between the information that is to be retained and the execution of the tasks. If the information that is necessary for cognitive tasks requires a large amount of retention and processing, processing speed will decrease, and errors will increase. In such difficult situations, individual differences are observed in the resulting level of constraint. This level of constraint affects higher-level cognitive activity, particularly language comprehension. It can be predicted that this tendency is more marked in verbal cognitive activity in a second language or a foreign language than in one’s native language. I will take up the topic of shadowing again later in regard to this point.

Miller (1956) famously once claimed that the amount of information that can be stored in short-term memory is “the magical number seven, plus or minus two.” This is a theory that holds that the amount of information that can be retained in working memory is $7 \pm 2$ chunks. A chunk represents the unit of processing of words or numbers. The number of chunks is stipulated to be quantitatively limited, regardless of the amount of information. However, Miller’s judgment criterion is criticized today as overly optimistic. Cowan (2001) holds that the pure storage capacity of the average person who does not rely on strategies is $4 \pm 1$ chunks, and this is a belief that is widely accepted. For a comparison of studies on the retention span of working memory, please refer to Morishita & Osaka (2008).

In recent years, many researchers have used brain imaging technology to learn how information is processed in the cerebrum. In other words, the validity of working memory models is now being verified with experimental psychology data. Positron emission tomography (PET) is a computer tomography technique that can determine sites in the brain that are activated based on heightened neural activity, as well as increased metabolic rate and blood flow, at those sites. Case studies of brain-damaged patients have also advanced verifications with PET (for details, see Osaka, 2002, chapter 7). On the basis of the results of these studies, the central executive can be said to be the most evolved part of the brain. Since Baddeley (1986) indicated that the central executive lies in the prefrontal cortex, there have been many neuroimaging studies that have explored the relationship between the prefrontal cortex and the central executive. Through these studies, it has been learned that the functions of the central executive are divided between the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC). In examinations of PET images of sites that are activated during dual task performance, which requires greater attention control than a single task, the DLPFC has been shown to be associated with the procurement, distribution, and maintenance of attention resources. That is, the DLPFC is associated with proper functioning while monitoring the information that is retained by working memory (Osaka, 2002: 138–140). However, the ACC, which has recently been an area of increasing focus (Osaka, 2002: 160), appears to be responsible for suppressing the processing of excess stimuli when multiple stimuli are present.

The phonological short-term store is located in the supramarginal gyrus, which is in the right inferior
area of the parietal lobe (Funabashi, 1997). The supramarginal gyrus is part of Wernicke’s area, which is also known as the sensory speech center and which is in the left hemisphere; damage to this region is known to cause aphasia. According to Valler & Papagno (1995), subvocal rehearsal, or articulation rehearsal, is a functional system that extends from the temporal lobe to the parietal lobe and that is centered particularly in the parietal lobe (cited from Kadota, 2006: 15–16). However, there is also the opinion that processing occurs in Broca’s area, which is in the inferior frontal gyrus (Paulesu, Frith, and Frackowiak, 1993). In addition, Baddeley (2007: 154) has stated that, while perhaps using only a specific area, the episodic buffer may be deeply related to the frontal lobe. Furthermore, Wernicke’s area and the surrounding areas are considered to be responsible for long-term memory, which handles verbal processing (Osaka, 2002). In addition, the area that is considered to be responsible for the visuospatial sketchpad is in the right hemisphere on the opposite side from where the phonological loop is located in the left hemisphere (Jonides, Smith, Koepp, Awh, Monoshima, & Mintun, 1993).

Neuroscientists currently advocate a dual-stream model that holds that there are two routes in the recognition and rehearsal of speech (see Ward, 2010: 228–230). The first route is called the ventral acoustic route (Ward, 2010: 228). Auditory information that enters the ear is first recognized as sound when it is transmitted to the cerebral sensory cortex. First, auditory information is received by (1) the primary auditory area/cortex in the left temporal lobe. This information is then sent to (2) the auditory association area, which recognizes and assesses auditory information. Next, (3) the temporal association area, which lies in the anterior position of the temporal lobe and which is responsible for integrating visual and auditory information, processes vocabulary and meaning. Finally, for vocalization, the information is sent to (4) Broca’s area, which is the motor area for speech. Broca’s area contains an accumulation of various muscle and motor memories that are related to speech production. On the basis of these memories, plans for the beginning of and procedures of movement are made in the motor association area (premotor area). Muscle movement commands are then issued in the motor area, resulting in vocalization. Due to this route’s inclusion of semantic comprehension, it is also known as the “what route.” When repeating sounds, it is inferred that the ventral route can only handle actual words. This is due to the fact that the ventral route involves semantic comprehension (Ward, 2010: 229, Kadota, 2012: 140).

However, the route that is used for subvocal rehearsal in the phonological loop is called the dorsal motoric route. As with the first route, auditory information is first recognized as sound by the primary auditory area/cortex. This information is then sent to Wernicke’s area, which is in the left-hemisphere temporal area, where auditory stimuli are recognized as language. This information is phonologically coded in the left parietal lobe, which is believed to be involved in linguistic and cognitive activity, while subvocal rehearsal is performed. Furthermore, a number of findings have shown that this route is also linked through the motor area to Broca’s area, which is related to phonological production (Kadota, 2012: 140). This route, which is also known as the “how route,” is used in the verbatim repetition, including nonverbal repetition, of words as they are heard (Ward, 2010: 230). Shadowing, which uses the latter route, is the repetition of speech immediately as it is spoken, without semantic comprehension, although it is not precluded completely.
4. Data Access (Top-Down and Bottom-Up Processing)

How is stored information elicited? Generally, the initial models were serial models in which the following process is repeated: all knowledge is accessed and searched in stages, and perceived stimuli are processed (e.g., Forster, 1976). However, the current leading theory holds that the brain performs parallel processing, like a computer. In parallel processing, processing that has a single purpose is divided into multiple processes that are performed simultaneously. In other words, the above information searches are believed to not be conducted serially or in stages but concurrently or with them mutually acting upon each other. Following the logogen model (Morton & Patterson, 1980), the cohort model of spoken language processing (Marslen-Wilson, 1987) was ultimately proposed.

In the cohort model, when a speech stimulus, such as a word, is inputted, words that share the first few phonemes are immediately activated from the mental lexicon, thus forming a cohort. This task utilizes the bottom-up processing of phonological knowledge. The top-down processing of context is also utilized concurrently in order to aid cognitive comprehension.

Working memory studies in autism, which is a developmental disorder, have offered important clues concerning human cognitive processes. According to Koshino (2008), processing in patients with autism tends to depend on low-level perceptual processes. Brain imaging studies have shown that patients with autism strongly tend to favor the visuospatial sketchpad over the phonological loop. In addition, functional coupling between the frontal and occipital regions in patients with autism is lower than in individuals without autism, and the central executive (for which the frontal area is responsible) does not function adequately. Therefore, in tasks that require more complex high-level linguistic cognitive function, top-down or feedback information, which goes from high level to low level, is observed to be inadequate. When the central executive functions normally, cognitive division and switching proceed smoothly, and the cooperative environment of top-down and bottom-up processes functions properly. According to Koshino (2008: 182–183), for successful speech processing, “an environment which promotes the bottom-up integration of low-level information into high-level information, as well as the top-down process of high-level information integrated into low-level information with low-level processing.” is essential. However, in patients with autism, high-level to low-level top-down processing is weak; therefore, the “environment” is considered to not function smoothly. That is, it has been concluded that the cooperative environment of top-down and bottom-up processes does not function properly in patients with autism. In other words, rather than relying on a single direction for processing, data processing must occur in both directions.

Next, I will examine the relationship between top-down and bottom-up processes in listening comprehension. One difference between native speaker listening (L1 listening) and second-language (foreign language) learner listening (L2 listening) is that the judgment and identification of sounds are slower in L2 listening than in native speakers. Because L1 listening is in the native language, the identification of speech sounds is automatic and instantaneous. In L2 listening, however, sounds are grasped consciously while knowledge is retrieved from semantic memory; thus, processing is slower than in L1 listening. As in the previously discussed cohort model, upon hearing the first few phonemes, a number of candidate words are selected. At the same time, the top-down processing of context and other information is also performed, thereby advancing speech comprehension. On the basis of the above, the following is necessary
for effective L2 listening: (1) the ability to execute automatic, instantaneous, and bottom-up processing in order to judge fast sounds, and (2) the ability to execute efficient top-down processing of context and other information.

In addition, an eye movement study by Rayners (2001) has found that, in sentence comprehension, native English speakers focus their eyes on all words when reading through a sentence. In other words, sentence comprehension is not a top-down process in which the eyes focus only on the important words. Thus, bottom-up processing is not omitted but rather is performed instantaneously and automatically. On the basis of the results of this study, Rayners (2001) has suggested the importance of bottom-up processing as a prerequisite for top-down processing.

In the next section, I will discuss the nature and effects of shadowing, which is a form of listening practice that aids the acceleration of bottom-up processing. In addition, in section 6, I would like to discuss the instruction of a listening strategy, which is a top-down processing drill that enables learners to efficiently perform top-down processing.

5. Promoting bottom-up processing: shadowing and its effects

Shadowing is defined as the process of immediately repeating the speech of a native English speaker without pauses (Kadota, 2006: 14). According to Kadota (2006: 55), shadowing is the practice of the rehearsal of English speech that is integrated in the phonological loop in working memory without semantic comprehension. In the perception and comprehension of speech, shadowing involves the performance of perception alone.

Training in shadowing, which does not permit breaks, promotes the automation of speech perception, which is considered to incidentally result in the automation of speech knowledge retrieval (Kadota, 2006: 74).

Baddeley (2002: 86) has stated that the span of information that can be temporarily stored is the number of words or letters that can be repeated within two seconds. Although there are individual differences, it is generally considered that native speakers can store a large amount of information, whereas the span of a second-language learner shrinks relative to their proficiency level. Therefore, training in the subvocal rehearsal function in the phonological loop is intended to improve the number of words or letters that can be retained in the phonological loop. Shadowing is the repetition of speech exactly as it is heard, and it can also be thought of as the vocalization of subvocal rehearsal in the phonological loop in working memory. Once the learner grows accustomed to subvocal rehearsal, the cognitive load that is required for it decreases. It is thought that the resulting surplus cognitive capacity can be utilized in the semantic comprehension of the rehearsed speech and the internalization of new information (efficient memory of vocabulary and sentence structures) (Kadota, 2012: 122).

Next, we will examine the superiority of shadowing by comparing it to other instruction methods. Tamai (2005) has compared the results of shadowing and dictation by comparing Japanese high school students’ performances in classes that conducted shadowing and dictation. It was found that, in the mid-proficiency group and the low-proficiency group, listening ability was significantly improved more with shadowing than with dictation.

Mochizuki (2004) has also compared shadowing with traditional instruction by conducting 13 experi-
mental classes, each lasting 20 minutes, with junior high school students. Performance in the experimental group, for which classes were conducted with shadowing, was compared with performance in the traditional group, for which classes were conducted with traditional instruction. Traditional instruction classes were conducted as follows: first, new words were introduced; after practicing pronunciation, students listened to a passage and practiced repeating. It was found that the experimental group who received shadowing instruction demonstrated significantly greater improvements in listening ability.

Hori (2008a, b) has reported the results of an experiment that examined the effects of shadowing. Participants showed even greater improvement in pronunciation speed, listening span, and oral reading test scores.

Miyake (2009) has shown that accelerated articulation is an effect of shadowing. In an examination of vocalization times for shadowing and phrase unit repeating,9 the accumulated practice of shadowing gradually reduced vocalization times; however, with repeating, vocalization times showed a tendency to grow slightly longer. With repeating, the learner attempts to store words and structures in English sentences in memory and comprehend the content. In shadowing, however, the learner must follow along and pronounce words one after another in order not to fall behind the continuous speech. The learner must also pronounce the ongoing speech as quickly as possible and be ready for the next speech. It is thought that vocalization times are shortened as a result. Kadota (2012: 177) has stated that, when a learner is placed in a situation that requires constant concentration and processing, “the learner will proceed in a direction which allows them to somehow reduce the cognitive load required for speech input,” which is “a prerequisite for the automation of speech perception.”

Shadowing is continuously devoted to the perception, rather than the comprehension, of speech input. Therefore, words and sentence structures cannot be stored consciously, even if the learner attempts to do so. However, it is thought that this information will be learned unconsciously without the learner being aware of it. Through the repeated practice of shadowing, it is considered that words and sentence structures will become established in unconscious latent memory, which is procedural memory (Kadota, 2012: 78).

A number of studies have reported that shadowing is particularly effective in mastering prosody. Hori (2008) has reported that shadowing instruction results in increased pitch variation. In addition, a study on the mastery of Japanese prosody has reported that shadowing drills are effective in mastering Japanese pitch accent (Ogiwara, 2007). Another report has shown that dual training in shadowing and oral reading significantly improved Japanese learners’ English speech prosody; when reading aloud, learners were able to use the English accent and rhythm to stress the portions of English speech that should be stressed (Mori, 2011: 17). It has also been stated that auditory priming effects10 are displayed more in prosody (rhythm and intonation) than in segments (vowels and consonants). By repeatedly encountering specific prosodic structures, one can partition speech into chunks, which are the units of semantic coherence. However, Kadota (2012: 261) has indicated that explicit instruction in rhythm and intonation is uncommon; therefore, auditory priming effects are seldom elicited in Japanese learners of English who do not possess explicit semantic knowledge that is related to prosody. In other words, an explicit knowledge of speech prosody is a necessary prerequisite for the conversion of semantic knowledge into procedural memory through shadowing.
6. Promoting top-down processing: listening strategy training

Previous studies suggest that the use of listening strategies may aid L2 learners’ top-down processing and improve their listening abilities. O’Malley, Chamot, and Kupper (1989) and Vandergrift (1997) have argued that it is important for L2 learners to be exposed to a variety of listening strategies in the following three categories: cognitive strategies (strategies used to understand what one has heard), metacognitive strategies (strategies to plan, monitor and evaluate one’s understanding), and socioaffective strategies (strategies which “either involve other people in our effort to understand, or which we use to encourage ourselves to understand” [Lynch, 2009: 79]).

In this listening strategy training, L2 learners are encouraged to use their background knowledge to make predictions, and to listen for overall understanding, and selectively attend to necessary information. Goh and Taib (2006) and Vandergrift and Tafaghodtari (2010) argued that implicitly raising learners’ metacognitive awareness by guiding them through the listening process (i.e., process-based instruction) is more effective than explicitly teaching individual strategies.

Goh and Taib (2006) conducted a small-scale study of process-based metacognitive instruction with ten primary school pupils in an English immersion school in Singapore. Pupils’ final reports indicated a deeper understanding of strategic knowledge for coping with comprehension difficulties. In addition, the authors compared in-house listening test scores before and after the experimental classes, and observed improvement in all participants except one. The authors did not report the statistical significance; nonetheless, this study shows the possibility that process-based instruction can raise strategic knowledge in young learners who are still developing cognitively.

Similarly, Vandergrift and Tafaghodtari (2010) conducted a larger-scale study with adults Canadians learning French at the university level. The experimental group went through thirteen guided-listening classes to improve underlying metacognitive processes. The authors found that this process-based approach was effective, especially for less skilled learners.

On the other hand, Field (2008: 308–309) argues that strategies should be explained individually and explicitly. They argue that strategies should be demonstrated their value in advance, and practiced in controlled listening tasks. As familiarity with the strategies improves, learners should be encouraged to consciously choose among them in a less controlled context.

The efficacy of listening strategy instruction has also begun to be verified in neuroscience experiments. Oishi (2006) has used optical topography to observe intracranial mechanisms in English language processing. She performed listening and reading tests on 20 university and graduate school students with different proficiency levels (based on TOEFL scores). In beginning and intermediate learners, excessive attention is focused on word cognition; cerebral blood flow increases excessively in Broca’s area and other sites that do not require activation (hyperactivation). However, in advanced learners, only Broca’s area is selectively activated (selective activation). However, the increase in blood flow is less than that of intermediate learners, and the appearance of the progressing automation of language processing is observed. With the same methods, subjects who were given content-related information (in the form of keywords presented before listening) were compared to those who were not given such information. Comprehension test scores increased in all proficiency levels when content-related information was given. However, in learners with
hyperactivation and selective activation, the proportion of increased blood flow to Broca’s area inversely decreased. The author has suggested that schema activation brings the state of brain activation closer to a state of automatic processing.

7. Conclusion

In this paper, I argued that a cooperative environment of top-down processes and bottom-up processes is essential for linguistic information processing: rather than relying on a single direction for processing, data processing must occur in both directions. I further argued for the necessity of first enhancing explicit knowledge: semantic knowledge that is related to prosodic structures may promote smooth bottom-up processing, and automate the retrieval of semantic knowledge to a level where it becomes unconscious. I introduced shadowing, which is the vocalization of subvocal rehearsal in the phonological loop, as an automation drill.

This action results in the vocalization of the subvocal rehearsal of English sounds that are integrated in the phonological loop in working memory without semantic comprehension. Through the repeated practice of shadowing, the cognitive load required for subvocal rehearsal decreases, thereby promoting the automation of speech perception. In other words, shadowing is an attempt to more closely approach the situation of a native speaker, in whom the retrieval of phonological information is automated. Thus, the anticipated effect is a conversion from explicit declarative memory to unconscious latent memory. Furthermore, a number of studies have reported that shadowing is particularly effective in mastering prosody. However, it has been indicated that enhancing semantic knowledge, which is the explicit knowledge of the phonological structure of the target language, is necessary to elicit the maximum effects of shadowing.

It was also suggested that the use of listening strategies may aid L2 learners’ top-down processing and improve their listening abilities. The efficacy of listening strategy instruction was discussed based on experimental studies including a neuroscience experiment.

Notes
1 This work was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) [24520660].
2 Peterson & Peterson (1959) have stated that, although 90 percent of this information is forgotten within 15 seconds, the retention period can be extended by rehearsing the information.
3 Although the multicomponent B & H model is currently widely accepted primarily in the West, many other models have also been proposed, and many comparisons have been attempted. The Americans Just and Carpenter et al. (1992) have proposed the CAPS model, which posits that working memory is a system that supports the parallel activation of the retention and processing of information that is necessary for improving higher-level cognition. Other proposed models include a focalization model (Cowan, 1999) in which the function of working memory is thought to bear the focus of attention that is activated by long-term memory. For a critique of this model, please refer to Baddeley (2007, 8.1.2). Baddeley (2007, chapter 1) and Osaka (2008, chapter 1) have covered various models of working memory and the history of the concept in detail.
4 The amount of information that can be repeated is expected to change due to various factors. One such factor is individual differences in the amount that can be repeated in a given time limit. Another is whether the words are known or whether the list consists of related words. In the case of second language acquisition, proficiency level is
also considered to be deeply involved. This point as it relates to shadowing and verbal repetition ability is covered later.

5 Subvocal rehearsal is also known by many other names, such as phonological rehearsal, articulatory rehearsal, and articulatory control process.

6 So-called articulation suppression is an experimental method. In this method, the rehearsal function that is normally performed in order to store or retain a stimulus (the subvocal rehearsal function in the phonological loop) is hindered by forcing subjects to mutter unrelated sounds or words or a repeating series of numbers. The act of pronouncing sentences or phrases after they have been spoken is called repeating. There is a pause before repeating begins, providing time for the retrieval of vocabulary and other information. Therefore, it is possible that, rather than imitate and pronounce the speech as is, the learner performs a sort of substitution that is based on imprecise phonological information. It may be the case that, while rehearsing speech input that is retained in the phonological loop, this input is unconsciously converted into speech representations with Japanese pronunciations and rhythms, with these pronunciations being adopted.

7 It has been reported that this sort of attention control is particularly displayed during Stroop interference (Bush, Luu, and Posner, 2000). The Stroop effect refers to the phenomenon of interference that results in a longer response time when the word “green” is presented written in red compared to when “green” is presented written in green or “red” is presented written in red.

8 Other parallel processing models have also been proposed, and these include the interactive model, which focuses on written language processing (McClelland & Rumelhart, 1981).

9 Pronouncing spoken sentences or phrases after they have been uttered is called repeating. There is a pause before repeating begins, and this provides time that allows for the retrieval of vocabulary and other information. Therefore, the speech is not received, imitated, and pronounced as it is. Rather, the learner may perform a sort of substitution that is based on imprecise phonological information: the input is unconsciously converted into speech representations with Japanese pronunciations and rhythms.

10 Tulving (1972) holds that memories are formed more quickly and precisely with perceptual priming effects, which arise from repeated perception.

References


Erlbaum Associates.


