Subvocalization of Japanese Students when Listening to English

Akio Suzuki
Ikuko Okuno

Abstract

This paper attempts to explore how Japanese learners of English as a foreign language at the beginner level process auditory input. The hypothesis was developed based on the working memory model: Japanese learners of English at the beginning level subvocalize acoustic information in the phonological loop in the working memory. Therefore, articulatory suppression would affect their performance on listening comprehension. The experiment was conducted with twenty-nine high school students, who were all required to read and listen in English and Japanese both, with and without articulatory suppression, and to answer multiple-choice questions. The results failed to support our hypothesis and the findings of the previous research. The necessity to examine the phonological representation of the learners' mental lexicon is suggested.

The difference of phonetic sounds between Japanese and English

It is often pointed out that Japanese learners of English face difficulties in learning the phonological aspects of the English language. In particular, they face problems with respect to the perception of not only phonemes but also sounds in fast speech and prosody (Takei, 2002). Some researchers have also found that Japanese learners are unable to discriminate between certain English contrasts even if they live in English-speaking environments for several years (e.g., Brown, 2000 ; Riney & Flege, 1998 ; Yamada, 1995).

One of the probable reasons that Japanese learners of English face difficulties in comprehending auditory information is that the Japanese

phonological system is different from that of English. For instance, the Japanese language has only five vowels, while English is said to have twenty-two vowels (Kawagoe, 1999) or at least fourteen distinct stressed vowels (Celce-Murcia, Brinton, & Goodwin, 1996). Another reason is that more than ninety percent of the Japanese syllables are composed of a consonant sound (C), followed by a vowel sound (V), while other syllables have a CVC structure (Kawagoe, 1999). On the other hand, English has a propensity for consonant clustering in the form of CCV, CCCV, VCC, and CVCCC (Celce-Murcia et al., 1996).

Linguists have investigated how the sounds of a second language (L2) are perceived and acquired by the learners of that language. Miyawaki, Strange, Verbrugge, Liberman, Jenkins, and Fujimura (1975) conducted an experiment involving Japanese participants and demonstrated that phonemes are perceived in terms of the listener's native categories. Since then, researchers have been making attempts to investigate how L2 sounds are mapped onto native categories. Best (1994, 1995) addresses the relationship between a listener's native phonological system and the perception of L2 sounds and claims that L2 sounds are assimilated into the listeners' native categories based on their respective articulatory similarities. The perception of L2 sounds is classified into three patterns: L2 segments are either assimilated into a particular native category as an exemplar, assimilated into a native phonological space as an uncategorizable speech sound, or not assimilated into a native phonological space. Therefore, the ability of a listener to perceive an L2 contrast depends on the extent to which the L2 contrast is assimilated into native categories.

Flege (1995) notes that foreign accents are common in the speech of nonnative speakers when conversing in English, and suggests that this is due to the inaccurate perception of English sounds. He claims that nonnative speakers of English do not perceive English sounds in exactly the same way that native speakers of English do and that English learners are more likely to equate English sounds with those of their native language. This is because the phonetic system is frequently stabilized before English learning begins (Flege, 1992).

Hancin-Bhatt (1994) attempts to explain how L2 sounds are mapped onto native phonological categories based on the feature prominence. She claims that features do not have discrete values; rather, they are continuous. She further states that each language has a different hierarchy of the feature prominence. Features and feature patterns that are used more frequently in the language's inventory will be more prominent than the ones that are used less frequently. Moreover, the features with higher

prominence in native phonology will tend to have a greater influence on how L2 sounds are mapped onto the existing native language (L1) categories. For instance, since continuant is in relatively strong competition with the other features in the Japanese inventory, Japanese speakers are more likely to misperceive the interdental as a sibilant rather than a stop.

Brown (2000) attempts to investigate the interrelation between speech perception and the phonological system and how the speech perception mechanism operates in the native speaker. From the experimental study, she found that learners with the same native language face different difficulties in perceiving and acquiring L2 contrasts, based on the presence or absence of contrastive features such as voice, manner, and place of articulation in their native phonological system. If the learner's native phonological inventory lacks the feature that discriminates a given phonological contrast, he or she will not be able to accurately perceive the contrast and the acoustic signals, because the phonemes will be mapped onto a single phonological category. On the other hand, the presence of contrastive features in an L1 phonological system will facilitate the classification of acoustic signals for the phonemes, leading to them being mapped onto two distinct phonological categories.

From the aforementioned literature, it appears unlikely that the phonological system of the target language is established in L2 learners, especially in the initial stages of L2 learning. Therefore, the acoustic information in English that Japanese learners of English usually receive in classrooms is arguably processed through the phonological system of their native language and mapped onto their native categories. In this manner, the phonological representations of English are constructed in the interlanguage of Japanese learners.

When Japanese learners of English at the beginner level perceive the speech produced by native or near-native speakers of English, we posit that learners need to match the information to their mental phonological representations of English in their interlanguage in order to confirm the acoustic inputs. Consequently, we assume that learners subvocalize phonetic sounds when they match the acoustic inputs with their mental phonological representations. This paper investigates whether Japanese learners of English at the beginner level subvocalize phonetic sounds when they process auditory information produced by native or near-native speakers of English, by employing the model of working memory. In the following section, we will briefly review the previous research on working memory.

Working Memory

Psychologists have attempted to investigate how human beings acquire and comprehend language from the perspective of human memory. Working memory and long-term memory (LTM) are considered to serve as two functionally separable cognitive systems. According to Baddeley and Logie (1999), working memory refers to "multiple specialized components of cognition" and its main role is to retain information while performing complex cognitive tasks such as language comprehension (p. 29).

Baddeley and Hitch (1974) originally proposed a working memory model, and subsequently, Baddeley and his colleagues developed the model (e.g., Baddeley, 1999; Baddeley, 2000; Baddeley et al., 1998; Baddeley & Logie, 1999). The model consists of four components: the central executive, two subsidiary systems, and the episodic buffer. The central executive is a supervisory attentional system that coordinates the subsidiary systems and activates as well as integrates representations in LTM. The episodic buffer is a temporary storage system with a limited capacity and can integrate information from various sources. The visuospatial sketchpad, one of the slave systems, maintains and manipulates visual images. The other slave system is the phonological loop, which retains verbal information for brief periods of time. It comprises a passive phonological store and an active rehearsal process; the phonological store holds information in phonological form, which fades with time, and the rehearsal process refreshes the decaying representations within the phonological store.

A number of studies have been conducted on patients suffering from brain damage and normal children and adults in order to discover the mechanism and functions of working memory and its multiple components. When the subjects were normal children or adults, the technique of articulatory suppression was utilized, which prevented subjects from rehearsing subvocally by requiring them to repeat an irrelevant word or count numbers. The results of these studies provided evidence for some hypotheses. For instance, it was found that auditory and visual inputs are processed differently. Auditory input is analyzed and directly transferred to the phonological store, while the written input of verbal material is visually analyzed and transferred to the phonological store through subvocal rehearsal that translates visual input into phonological code. The subjects could not feed visual information to the phonological store if they suppressed articulation, because visually presented material did not get translated into a phonological code by subvocal rehearsal.

Some studies involving Japanese learners found that articulatory

suppression impedes reading comprehension. Kadota (1997) conducted a study involving Japanese university students to examine the interference of articulatory suppression on the reading and listening comprehension of English texts. The students were divided into four groups, and each group participated in one of four types of tests: English listening, English reading, Japanese listening, and Japanese reading. Two groups silently read or listened to 20 English sentences that were presented one at a time, and the remaining groups were required to suppress articulation by repeating "ta ta ta..." at a 200-ms inter-beat-interval (IBI) while reading or listening to the English sentences. They transcribed as much of the sentence as they could remember and wrote its meaning in Japanese. The results revealed that although articulatory suppression affected their performance in processing the visually presented material, it did not interfere with the memorizing of auditorily presented sentences. Kadota (1987) also conducted a study involving Japanese university students in order to investigate if articulatory suppression disrupts reading comprehension in Japanese as well. The results showed that articulatory suppression did reduce the level of performance in the reading test. Therefore, Kadota (1987) concluded that articulatory suppression interferes with reading comprehension in Japanese.

Based on the literature reviewed above, we utilized the technique of articulatory suppression that was developed on the basis of the working memory model in order to investigate how Japanese learners of English at the beginner level process auditory information. More precisely, we hypothesize the following:

1. Articulatory suppression affects both English and Japanese reading, as revealed by previous studies.
2. It does not interfere with Japanese listening, as shown by previous empirical studies.
3. It affects English listening because Japanese learners of English at the beginner level subvocally rehearse auditory information in the phonological loop when they process speech input.

The first and second hypotheses serve as control predictions, and this research focuses on the last hypothesis.

Methodology

Participants

Twenty-nine first-year high school students participated in this study. They were aged between 16 and 17 years. They all had three years of formal instruction in the English language, mainly provided by Japanese teachers and occasionally by native speakers of English. Hence, we considered them as English learners at the beginner level.

Materials

Eight types of materials were prepared for English listening, Japanese listening, English reading, and Japanese reading comprehension tests. Both English and Japanese listening tests were drawn from the previous 3rd Grade level of the Society for Testing English Proficiency tests. Two texts were selected for each test type. The texts for the English listening comprehension tests were recorded by an English native speaker from the U.S. using a SONY TCM-1390, while those for the Japanese listening comprehension tests were recorded by one of the authors, a native speaker of Japanese. Further, the texts for the reading comprehension tests were selected from the available educational resources. Eight multiple-choice questions were developed for each text. The multiple-choice questions were utilized because we decided that considering the level of participants' English, it would be difficult for them to answer more advanced questions such as open-ended questions and cloze tests.

Procedure

The four types of tests were administered to small groups in the audiovisual room of the high school. A SONY CFD-5230 was utilized for the listening comprehension tests. Each test type was administered twice. First, the participants listened to or read one of the texts before answering the questions. Subsequently, students suppressed subvocalization by counting the numbers one to five at a 200-ms IBI while listening to or reading one of the other texts. They listened to the texts and answered the questions within 2 minutes. For the English reading tests, they were allotted 2 minutes and 15 seconds to read and 2 minutes to answer the questions. As for the Japanese reading tests, the participants were allotted 1 minute and 15 seconds to read and 2 minutes to answer the questions. The scoring procedure was dichotomously based on correctness. Two scorers objectively marked the students' answer sheets against the key.

Results

Table 1 shows the average number of correct points (out of the 8 possible) for the multiple-choice comprehension test for each group. This experiment used a mixed design that included 3 within-subject factors: 2 types of task conditions (with or without articulatory suppression), 4 types of language-mode conditions (English listening, English reading, Japanese listening, and Japanese reading), and 8 types of materials. In the English listening condition, for example, half of the participants were asked to listen to Material A with articulatory suppression and then to Material B without the suppression. On the other hand, the remaining participants were asked to listen to Material B with articulatory suppression and then Material A without the suppression. An analysis of variance (ANOVA) with task type (with or without suppression) and material type as within-subjects factors yielded significant differences for material type in English reading condition ($F(1, 54) = 16.36; p < .01$) and for task type in the Japanese listening condition ($F(1, 54) = 8.14; p < .05$) and the English reading condition ($F(1, 54) = 4.52; p < .05$). Although task type in the English reading condition yielded a significant difference, this was due to the difference of material type, which also reached a significant level.

Our hypothesis, which claims that articulatory suppression affects English listening because Japanese learners of English at the beginner level subvocally rehearse auditory information in the phonological loop when they process speech input, was not supported. The English listening was not disrupted by articulatory suppression. Furthermore, the control predictions were not supported as well. This implies that both the English and Japanese reading were not disrupted, while the Japanese listening was disrupted by articulatory suppression. This is inconsistent with the previous research.

Table 1: Summary of mean correct points

	English Listening		Japanese Listening		English Reading		Japanese Reading	
Suppression	NS	S	NS	S	NS	S	NS	S
Correct Points	4.21	3.62	6.66	5.69	4.41	3.31	6.34	5.90
SD	1.45	1.15	1.04	1.39	1.72	1.56	1.34	1.29

Note. NS = without articulatory suppression, S = with articulatory suppression

Discussion and Conclusion

The research question in this study pertains to whether or not auditory information is subvocalized in the phonological loop. In order to answer this research question, we employed the method of articulatory suppression. The results obtained from the experiment failed to answer our research question since the control predictions were not supported. We will discuss the reasons for the same in the following section.

The scores obtained in both the English listening and the English reading conditions are low, which can be referred to as the floor effect. This is partly because the participants in this study are at the initial stages of English learning. The result obtained in the English listening condition is almost parallel to that of previous research on articulatory suppression, which indicates the absence of articulatory suppression in the listening task. As for the English reading condition, since the difference between the passages reached a significant level, that between the scores obtained cannot be simply attributed to the articulatory suppression.

Although according to previous research such as Kadota (1987, 1997), there should be articulatory suppression for the Japanese reading condition, we failed to observe a significant difference. From informal observation, we interpreted this result as follows. The participants in the Japanese reading condition were allowed to adjust the time spent on reading and answering the questions without restrictions, while in the Japanese listening condition, they were not allowed this adjustment. The effect of articulatory suppression observed in the Japanese listening condition is inconsistent with that of the previous research.

When we consider the reason why the effect of articulatory suppression was observed in the Japanese listening condition, it is assumed that this difference was yielded not so much by articulatory suppression as by the dual task.

Watters (1985) maintains that producing sounds disrupts reading comprehension, not as articulatory suppression but as a distractor of the participants' attention from information processing. It limits their capacity to process information. In this study, the counting of numbers was utilized as articulatory suppression in order to prevent the participants from subvocalizing. However, because the participants were required to count numbers at a fixed rate, their attention may have been directed toward performing this task.

In the present study, the participants were asked to count numbers at a 200-ms IBI. We decided to employ the rate based on previous research (e.g., Kadota, 1997), according to which the rate does not seem to interfere

with the cognitive process of language comprehension such as reading. However, from the informal observation of our experiment, the participants directed their attention toward counting the numbers at the fixed rate. As a result, they were required to perform the dual task, namely, comprehending the texts and counting numbers at the fixed rate.

This study attempted to examine the process in which Japanese learners of English at the beginner level match acoustic information to their mental lexicon by using the technique of articulatory suppression based on the working memory model. However, the results obtained from the experiment did not support our hypothesis. Therefore, future research should focus on investigating the phonological representation of the learners' mental lexicon constructed through learning English as a foreign language in Japan.

References

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

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

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Phonological Review*, *105*, 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. *The psychology of learning and motivation: Advances in research and theory*, *8*, 47–89.

Baddeley, A., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 28–61). Cambridge: Cambridge University Press.

Best, C. (1994). The emergence of native-language phonological influence in infants: A perceptual assimilation model. In J. C. Goodman and H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). Cambridge, MA: The MIT Press.

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 171–204). Baltimore: York Press.

Brown, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second language acquisition and linguistic theory* (pp. 4–63). Oxford: Blackwell Publishers Inc.

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

Cambridge: Cambridge University Press.

Flege, J. E. (1992). Speech learning in a second-language. In C. A. Ferguson, L. Menn, & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, implications* (pp. 565–624). Parkton, MD : York Press.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 233–277). Baltimore: York Press.

Hancin-Bhatt, B. (1994). Segment transfer: A consequence of a dynamic system. *Second Language Research*, *10*, 241–269.

Kadota, S. (1987). Nihongo no dokkai to shinritekionseikagensho. *Hishoten joshi tanki daigaku kiyou*, *13–14*, 83–102.

Kadota, S. (1997). Shikaku oyobi choukaku teijibun no shori ni okeru onseitekikanshoukadai no eikyou. *Kotoba to komyunikeishon*, *1*, 32–44.

Kawagoe, I. (1999). *Eigo no onsei wo kagaku suru*. Tokyo: Taishukanshoten.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A., Jenkins, J. & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, *18*, 331–340.

Riney, T. J., & Flege, J. E. (1998). Changes over time in global foreign accent and liquid identifiability and accuracy. *Studies in Second Language Acquisition*, *20*, 213–243.

Takei, A. (2002) *Eigo risuningu ron: kiku tikara to sidou wo kagaku suru*. Tokyo: Kagensha.

Watters, G. S. (1985). The effects of concurrent tasks on reading: Implications for phonological recording. *Journal of Memory and Language*, *24*, 27–45.

Yamada, R. (1995). Age and acquisition of second language speech sounds perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 305–320). Timonium, MD : York Press.