日本語話者による
単語ストレスとシュワーの処理と音の表象

Processing and representation of lexical stress and schwa by Japanese speakers:a preliminary study

杉浦香織
文化政策学部国際文化学科

Kaori SUGIURA
Department of International Culture, Faculty of Cultural Policy and Management

本稿では、母語の単語アクセントがピッチ•アクセントである日本語話者が、ストレス•アクセント単語のストレスと弱母音（シュワー）の音声処理を、ストレス•アクセント単語を母語に持つ英語母語話者と同様にできるかどうかを検討した。実験で参加者は、ストレスとシュワーの単語内での位置が対照的な無意味語（ [MIpa] vs. [miPA], [paFU] vs. [pəFU]）が連続的に音声呈示された順序を、コンピュータ上のキーを用いて復元した。単語刺激は、３語連続から５語連続へと徐々に増加し、記憶への負担が増すように呈示された。その結果、単語ストレス処理課題では、記憶への負担が増加しても、英語母語話者と日本語話者の課題におけるエラー数に統計的有意差はみられなかった。一方、シュワー課題でも両者に有意差は見られなかったが、日本語話者は英語母語話者と比べて、特に４語と５語において、エラーがより多く見られた。以上より、日本語話者は、ストレス•アクセント単語処理の際、プロソディック情報に対して英語母語話者と同程度に敏感であることが示唆された。

This study aims to see if there are any differences in terms of the processing ability of lexical stress and schwa in Japanese speakers (JS).The participants were required to perceive and reproduce contrastive stress as well as schwa and vowel /a/ contrast under different degrees of memory burden. The results showed that the participants were able to perceive lexical contrastive stress at a deeper level than the phonetic surface level in the same manner as native speakers of English (NS). As for the schwa task, it seems that distinguishing schwa from /a/ is difficult for both JS and NS. However, it was observed that both JS groups made more errors in the four- and five-word sequences in the task compared to NS. These results suggested that JS are sensitive to the prosodic information to the same degree as NS in processing of words that contain stress accent.

1. Introduction

English is one of the few languages that exploit both suprasegmental (i.e. stress) and segmental information (i.e. schwa) in perception of lexical stress (Cooper et al., 2002; Cutler & Clifton, 1984). Acquiring both suprasegmental and segmental aspects is crucial for second language (L2) learners of English. However, as language employs different suprasegmental and segmental features to make lexical distinctions, L2 learners might perceive English lexical stress differently from native speakers (NS) or misperceive it.

Using the experimental paradigm of psycholinguistics, this study was conducted to investigate whether Japanese speakers (JS), who do not exhibit lexical stress, nor the unstressed vowel schwa in their speech, can perceive contrastive stress and the contrastive pair of schwa and vowel /a/ in the same manner as native speakers of English. In addition, this study examined the differences in "processing" levels for the target sounds in JS of English and NS. Since in daily life people use their linguistic knowledge to communicate with each other, it is crucial to investigate the on-line mechanisms of learners, through which their linguistic representations are put into use. In this study, in order to see if there are any differences in terms of their processing ability of lexical stress and schwa, participants were required to perceive and reproduce contrastive stress and schwa and vowel /a/ contrast under different degrees of memory burden. The performance of NS and the two groups of JS were subsequently compared.

2. Literature background

2.1 Difference of phonetic component of lexical accent in English and Japanese

English and Japanese are different in terms of phonetic components of lexical accent. The accent in English exhibits a stress accent; on the other hand, Japanese has a non-stress accent. The English accent is realized with pitch, intensity, and duration (Beckman, 1986), while Japanese utilizes mainly pitch (Beckman, 1986; Sugito, 1969). More importantly, the Japanese accent is independent of the rhythmic aspect that is observable in the English stress accent (Haraguchi, 1977). In Japanese, the rhythmic structure is determined only with the number of moras and pauses, and there is no contribution of pitch accent to create the rhythm. However, in English the alternation of stress and unstressed syllables employing schwa in lexical words plays a

critical role in producing a stress-timed rhythm.

2.2 Difference of vowels in English and Japanese

There are differences between Tokyo Japanese vowels and American English vowels. In terms of the richness of the vowel space inventory and their phonetic natures, Tokyo Japanese vowels consist of the five vowels, /i, e, a, o, and u/, which are phonetically realized as monophones without a reduced vowel. On the other hand, American English exhibits nine phonemic vowels (Ladefoged, 1993), including /i, ɪ, ɛ, æ, e, ə, ʌ, ʊ, ɔ, and a/ in the vowel space. The Japanese vowels are distributed in peripheral areas in the vowel space, and the vowel space has an empty central area with no central vowel categories, whereas English has a mid-central vowel, /ə/. Due to the lack of a central vowel in the Japanese vowel system, it may be plausible that Japanese learners of English might have difficulty in perceiving schwa.

2.3 Perception of segmental properties in second language acquisition

Research on the perception of speech segments by non-native speakers of English has attracted considerable attention (Best, 1994; Flege, 1995). Flege (1992, 1995) has proposed the Speech Learning Model (SLM), which claims that the persistent pronunciation and perception difficulties in non-native sounds are due to the perceptual similarity between the target L2 sound and the non-native sound. The Perceptual Assimilation Model (PAM) established by Best (1994) has claimed that gestural information can encode a speech signal, proposing that if two distinct foreign sounds are processed as an identical articulatory gesture based on the native language, those two different sounds are assimilated to one segment and are difficult to discriminate. Although these theories have contributed considerably to research on L2 phonology, they were not concerned with the processing levels of speech sounds (Matthews & Brown, 1998).

Considering the degree of the sound processing levels based on Werker and Logan (1985), Matthews and Brown (1998) demonstrated the developmental stages of L2 perception. Before mentioning the results of Matthews and Brown, it would be better to explain the degrees of sound processing levels in order to better understand their study.

Werker and Logan (1985) claimed the existence of three distinct representations generated in the course of speech processing as shown in Figure 1:

(a) the acoustic representation
(b) the phonetic representation
(c) the phonemic representation

According to the researchers, the representation at each level can be found by manipulating the inter stimulus interval (ISI) in a sound discrimination task, such as a forced-choice AX discrimination task, where participants have to indicate if X is identical to A or not.

For Section (a), the acoustic representation is accessible in speech perception when the ISI is within the 250 ms. level of the task. At this level, the participants could distinguish between two sounds by making use of the fine detail cues from the speech signal, including minute distinctions between repeated utterances by the same person although the sounds were phonemically identical (e.g. [p]-[p] for native speakers). For Section (b), the phonetic representation, which can be obtained in the ISI of 500 ms., is more abstract information than the acoustic representation, but has non-language specific properties (e.g. [p]-[ph] for native speakers). For Section (c), when the ISI is longer, access to the deeper level of representation is available. At this level of speech processing, the sound features (e.g. [p]-[b] for native speakers) are language specific.

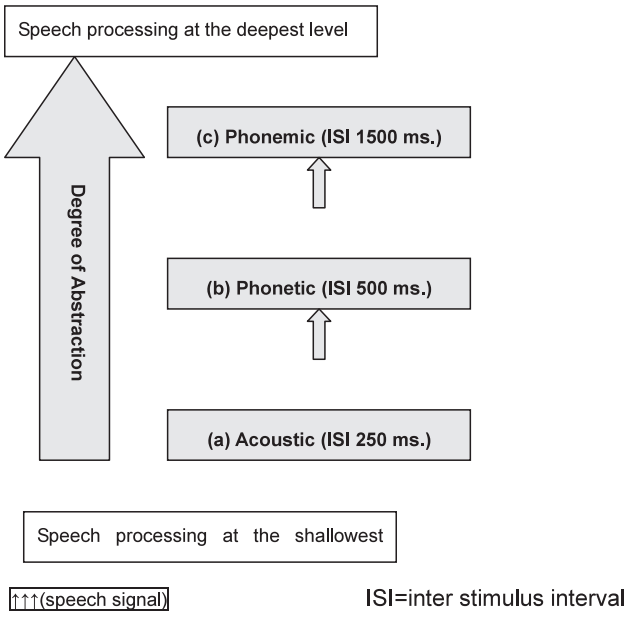

Figure 1. Types of Codes Generated in Short-term Memory during the Discrimination of Speech Sounds (Mathews & Brown, 1998, based on Werker & Logan, 1985).

Let us return to the work performed by Matthews and Brown (1998). In their study, the ISI of the stimulus pairs (ISI: 250 ms. or 1500 ms.) were manipulated in two alternative forced choice AX discrimination tasks. The findings revealed that the Japanese speakers with a low English proficiency level had more difficulty discriminating between /l/-/r/ and /s/-/θ/ than the other contrasts in the 250 ms. ISI condition, and performed poorly for all contrasts as compared to native speakers. In the 1500 ms. ISI condition, their performance on the discrimination of all the contrasts was poor compared to native speakers, suggesting that they had not yet acquired the features necessary for categorizing those contrasts at the phonemic level. Conversely, the Japanese speakers with a high English proficiency could successfully discriminate /b/-/v/ in both ISI conditions, suggesting that they had already constructed the new phonemic representation for categorizing /b/-/v/.

This evidence from speakers of both proficiency levels clearly showed that these speakers were eventually able to trigger a new phonological segment to differentiate the contrasts, although learners at an early stage of acquisition were insensitive to the L2 contrasts at a phonemic level of processing.

In sum, a large number of studies have investigated L2 acquisition and processing of segmental sounds based on the speech perception theories proposed by Best (1994) and/or Flege (1995). However, few studies in which the levels of speech processing were considered have been conducted, and the perception of phonemic contrast between the schwa and a full vowel by L2 learners has not been fully investigated. To reveal L2 learners' development of phonological representation, it is crucial to examine participants' speech processing levels in perception experiments. In addition, more studies on how L2 learners perceive the English unstressed vowel schwa are needed since it is an important vowel that frequently appears in English speech and contributes to creating English rhythm.

2.4 Perception of lexical stress in second language acquisition

To investigate lexical stress from the view of sound processing, Dupoux, Peperkamp, and Sebastian (2001) designed a sequence recall task where participants were required to hear the sequences of contrastive sound items and reproduce them. The lengths of the sequences were also manipulated to change the degree of the speech processing level. There is an important difference between the newly devised task by Dupoux et al. (2001) and a forced choice AX discrimination task, which is often used in perception experiments. The former has more possibility of assessing one's phonological representation (which relates to the deepest level of sound processing), whereas the latter might allow listeners to use acoustic cues (which are a shallower level of sound processing).

Dupoux et al. (2001) were able to demonstrate in their new experimental paradigm that the Spanish participants (whose native language has lexical stress) could distinguish stress contrast and perform better than the French participants (whose native language does not have stress contrasts). However, when Dupoux, Pallie, and Mehler (1997) employed an AX discrimination task, they failed to demonstrate the French participants' insensitivity to stress. They assumed that this failure was because the French participants were able to utilize the acoustic cues in order to flawlessly perform the stress discrimination tasks. These experiments suggest that the sequence recall task is a more robust method for determining one's ability to process sounds at the level of language-specific phonology.

Based on Dupoux et al.'s (2001) short-term memory experimental paradigm, Chan (2005) studied successful bilinguals of Cantonese and English. He examined whether the participants performed similarly to native speakers in discriminating stress contrasts. The results showed that there was no significance between bilingual and native speakers of English in distinguishing nonce words and English stress contrasts. The researcher speculated that the bilinguals of Cantonese and English might have compensated the representation of tone in Cantonese for stress; therefore, there appeared to be no difference between the performances of the two groups even though the reasoning may be different.

In light of language accent typology, Altmann and Vogel (2002) proposed the Stress Typology Model (STM), which deals with the perception of English lexical stress by L2 speakers. This model predicts that L2 speakers of non-stress first languages (such as Chinese, Japanese and Korean) are better at the perception of stress, because they do not have any positive L1 parameter setting for the model stress that can

possibly interfere with the L2 settings. On the other hand, speakers of L1 with fixed stress (such as French or Arabic) encounter considerable difficulty in acquiring new stress since they had already set several stress parameters for L1 stress, which perhaps impedes the acquisition of new L2 stress. Altmann (2006) supported the STM by systematically examining typologically different languages. These include predictable stress (i.e. fixed stress) languages, such as French, Arabic, and Turkish; non-predictable stress languages, such as Spanish, Russian, and English; and non-stress languages, such as Chinese, Japanese, and Korean. In an on-line task designed to examine the participants' processing ability, they were instructed to listen to nonce words and were then timed as they marked which syllable they felt had the most stress or prominence. As predicted in the STM, the results showed that the learners with predictable stress in L1 had problems in perceiving the location of stress. On the contrary, the learners without lexical stress or with non-predictable stress in L1 exhibited an almost perfect performance in perception.

As shown in the previous literature, the L2 learner's native language background seems to influence their performance in English stress perception. The research has shown the possibility that Japanese learners of English can perceive contrastive stress. However, it is not clear whether or not the non-stress language listeners can process stress from native speakers at the different levels of sound processing.

Based on the previous studies mentioned above, this study also examines whether or not Japanese learners process lexical stress as well as schwa/full vowel contrast in the same ways at the different levels of sound processing.

3. Method

3.1 Experimental paradigm

The experiment consisted of the contrastive stress (i.e. [MIpa] vs. [miPA][1]), consonantal phoneme (i.e. [TUki] vs. [TUpi]), and vocalic (i.e. [paFU] vs. [pə FU]) tasks. Each task was called, Stress, Phoneme, Schwa task respectively, and was constructed in three blocks. The first block contained a sequence of three words, the second one consisted of four words, and the third one had five words. All of the selected sequences are listed in Table 1.

Since this experiment was designed for the purpose of assessing phonological representations, the following aspects were considered based on Dupoux et al. (2001).

1. By using the sequences of three to five words, this experiment gradually increased the burden of memory for participants. As the burden is increased, participants have to encode the sound information at a deeper level.
2. By providing some phonetic variability in each word, manipulating a pitch, more abstract phonological representations, rather than acoustic, were assessed.
3. In order to prevent the participants from using echoic memory (which is defined as very brief sensory memory of some auditory stimuli and is typically stored for short periods of time) every sequence was followed by "OK" (Morton, Crowder, & Prussin, 1971; Morton, Marcus, & Ottley, 1981), and they could not begin typing their responses until they had heard this word.
4. The speed of presentation (the ISI) was kept very short, specifically 80 msec. in order to diminish the likelihood that the participants would use the strategy of mentally translating the words into the associated numbers while listening to the sequence.

3.2 Materials

The stimuli were created using an American male's and female's voices in Text to Speech of AT & T Labs (http://www.research.att.com/~ttsweb/tts/demo.php). In addition, the word "OK" was recorded by the female's voice. All the stimuli were recorded using the Praat software (Boersma & Weenink, 2007), and stored

Table 1 *Types of sequences*

Sequence	Sequence types
3-word sequence	111, 112, 121, 122, 211, 212, 221, 222
4-word sequence	1121, 1122, 1211, 1221, 2111, 2112, 2122, 2212
5-word sequence	11121, 12112, 12122, 12211, 21211, 21112, 21221, 22122

Note. Indicated number "1" is associated with [MIpa] and "2" with [miPA] on the computer keyboard.

on a computer disk.

Table 2 indicates the acoustic characteristics of the stimuli. The mean durations of each stimulus were manipulated to be as equal as possible (see Table 2-A). Let us look more closely at other phonetic characteristics of each contrastive stimulus. In the tokens of the contrastive stress, as shown in Table 2-B, the stressed vowels in [MIpa and miPA] were on average 48 msec. longer than the unstressed vowels. The maximum value of F0 in the stressed vowels was on average 68.5 Hz higher than that of the unstressed vowels. The intensity of the stressed vowels was on average 4.5 dB louder[2] than that of the unstressed vowels.

In terms of the tokens with the full vowel /a/ and schwa contrast, that is, [paFU vs. pəFU], as shown in Table 2-C, the full vowel and schwa in each token were almost the same in duration: 53 msec. for the full vowel, and 51 msec. for the schwa. This was intended to make the participants pay attention to the quality of the vowels, rather than to the durational cue. For the quality, the first and second formant frequencies (F1, F2) were evaluated based on the data obtained in the previous studies: for schwa, F1: 530-575 Hz and F2: 1700 Hz-1720 Hz (Peterson & Barney, 1952; Wallace, 1994) and for /a/, F1: 850 Hz and F2: 1200 Hz (Peterson and Barney, 1952). As a matter of fact, the F1 and F2 in the stimuli were higher than the criteria. This is probably attributable to the gender differences: it is generally said that the formant frequencies for females' voices are usually higher than those of males, and the studies used for the criteria had obtained the results from males, whereas the present study used a female's voice. In terms of the stressed vowels (i.e. second syllables) in the vocalic minimal pairs, they were on average 65.5 Hz higher in *F0* than the unstressed vowels (i.e. the first syllables). As for intensity, there was a difference of 22 dB between the full vowel and schwa. The absolute peak in the intensity for the schwa token was 56 dB, and this was higher than the one obtained from Wallace's (1994) study (30-40 dB). However, intensity generally varies,

Table 2 *Acoustical description of stimulus*

A: Duration of three contrastive stimulus (ms.)

(avg.)	stress stimulus	consonant phonemes	vowel phonemes
	MIpa: 307	TUki: 354	paFU: 310
	miPA: 311	tuPI: 353	pəFU: 304

B: Stress: [MIpa] vs. [miPA]

	duration: msec.	pitch: Hz	intensity: dB
1st/2nd syllable			
MIpa:	141/96	160/84	56/54
miPA:	89/140	100/157	56/63

C: Schwa: [paFU] vs. [pəFU]

	F1 & F2: Hz	duration: msec.	pitch: Hz	intensity: dB
1st/2nd syllable				
paFU:	923/1428	53/241	161/232	78/70
pəFU:	697/1955	51/247	175/236	56/66

D: Phoneme: [TUki] vs. [TUpi]

	duration: msec.	pitch: Hz	intensity: dB
1st/2nd syllable			
TUki:	193/171	215/145	58/44
TUpi:	191/174	209/155	59/42

depending on the recording conditions, so it could be assumed that the high intensity rarely affected the results.

With respect to the [TUki] vs. [TUpi] contrast, as shown in Table 2-D, the duration of *F0* and intensity were created to be as equal as possible.

Finally, to give each token more variations, its *F0* was changed by means of free software, WavePad, with 105, 101, 97, 93% including the original tokens, and in total the five variations were prepared for each item.

3.3 Experimental design

The experiment consisted of three tasks to explore stress, the consonantal and vocalic contrasts within disyllabic nonce words. Each was constructed with three blocks and contained eight sequences of the two contrastive nonce word sequences of three, four and five words. The first group contained the sequences of three words, the second one consisted of four-word sequences, and the third one had sequences of five words. All the selected sequences are listed in Table 2. The stress contrast, [1] was associated with [MIpa], while [2] was associated with [miPA]. For the consonantal contrast, [1] was associated with [TUki], while [2] was associated with [TUpi] and for the vocalic contrast, [1] was associated with [paFU] and [2] with [pəFU].

3.4 Procedure

1. The participants were told that they were going to learn two non-words. They could listen to the tokens of the two words as many times as they wanted. While the participants listened to [MIpa], [1] was shown on a computer screen. After that, they listened to its counterpart [miPA], and [2] was indicated on the screen at the same time.
2. Subsequently, the participants had to take a pre-test to verify that they had learned the distinction between the two words, as well as the correct association between the words and the number keys, that is, [MIpa] for key [1] and [miPA] for key [2]. They heard four trials consisting of a three-word sequence, two four-word sequences and a five-word sequence, and they had to reproduce each sequence by pressing the associated keys in the correct order. They received the message, "Correct!" on the screen if their response was correct. If not, the participants took the same trial until they answered correctly. After having finished the test, the participants moved to the main experiments.
3. During the test, the participants listened to 24 sequences constituted by the repetitions of the minimal pairs, divided into three blocks. For each participant, the order of the eight sequences in each block was randomized. The participants did not receive feedback as to whether or not their responses were correct. After a 1500-ms. pause, they moved to the next trial, but if they could type the answer key earlier than the set time, they were also allowed to move to the next one.

On average, the entire experiment lasted about 30 minutes. The experiments were conducted with Super Lab 4.0 (Cedrus), and responses and reaction times were recorded on a computer disk. The response time was measured from the onset to the offset of pressing the keys. The participants with more incorrect responses than correct responses were eliminated since they might have confused the number key associated with the first item with the one associated with the second item.

3.5 Participants

Seven NS aged between late 10s and 40s, nine JS aged between 20s and 50s with high English proficiency (adv. JS), and 13 JS aged between 18 and 50s with low English proficiency (beg. JS) participated in the experiment.

The criterion used to select the speakers for the two Japanese groups was based on the scores of English tests, such as TOEIC, TOEFL and "Eiken". The JS obtaining a score higher than 250 in TOEFL (the CTB version), or a score of 870 in TOEIC, or the first grade in "Eiken" were classified as the learners at the advanced level.

On the contrary, the JS who obtained scores between 200 and 500 in TOEIC or less than the pre-2nd grade in "Eiken" were categorized as the beginners. Five out of the nine JS with high English proficiency had been staying in English speaking countries for more than two years,

Table 3 *A trial of 3-word sequence*

stimuli	-80 msec.-	stimuli	-80 msec.-	stimuli	-80 msec.-	OK
(e.g. MIpa)		(e.g. miPA)		(e.g. miPA)		

whereas none of the JS who were at the beginner level had had such an experience.

All of the advanced English speakers – except one who had stayed in the US from the age of ten to twelve – began to learn English at junior high schools in Japan. At the time of data collection, the advanced speakers were either graduate students majoring in linguistics or English teachers at universities in Japan. All of the beginners of English – except one who was a math teacher at a university – were university students in Japan. As for the NS, one of them had been staying in Japan for half a year as an exchange student, and the others had been staying in Japan as English teachers for more than ten years. None of the participants had any problems with hearing and speaking.

3.6 Analysis

The participants' responses were recorded on a computer disk and classified as follows. If the input sequence was 100% correctly reproduced in the response, it was coded as correct; all other possible responses were coded as incorrect. A participant with 100% incorrect responses in one of the three tasks was rejected. The high percentage of incorrect responses suggests that they might have confused the number key associated with the first and second sound items, or they may not have concentrated on the experiment at all.

For the statistical analyses, an analysis with generalized linear models (GLM) was conducted. The dependent variable was Error rate and the independent variables included: *Memory load (3-, 4- and 5-word sequences), Group (beginner JS, advanced JS and NS) and Contrast (stress, schwa and phoneme)*.

4. Results

4.1 Results of descriptive analysis

To begin, Table 4 shows the description of error percentages for beginner JS, advanced JS, and NS participants for the contrasts as a function of sequence lengths. It was observed that all the groups performed well on the contrastive stress tasks: the error rates ranged from around 5 % to around 35 % in the three-word sequence and the error rates became higher as the memory load increased. However, the error rates in the five-word sequences were around 20 % (NS and Adv. JS) to 30% (Beg. JS).

In terms of the schwa task, overall the three groups performed more poorly than they did in the contrastive stress task, indicating above 35% in the three-word sequence. It should be noted that, in particular, the error rates of the two JS groups became higher as the memory load increased (69.7 % for Beg. JS; 65.3 % for Adv. JS; 49.9 % for NS in the five-word sequence). Beg. JS and Adv. JS made 1.4 times and 1.3 times as many errors as that of NS in the four- and five-word sequence, showing that JS had more difficulty in distinguishing between [paFU] and [pəFU].

Regarding phoneme contrast, the error rates fall between that of the stress and schwa tasks, indicating around 20% to 40% , and increased in parallel to the memory load increases.

4.2 Results of statistical analysis

First of all, a three-way ANOVA with the factors of *Group, Contrast*, and *Memory* revealed there was no interaction between *Group* vs. *Contrast* (p = .573), *Contrast* vs. *Memory load* (p = .906), and Group vs. Memory load (p = .978), showing that the three groups have no difference among the

Table 4 *Percentage of error with phoneme, schwa, and stress contrast as a function of sequence length for beg. JS, adv. JS, and NS*

Sequence length		3 words	4 words	5 words	Mean
Beg. JS	Stress	**7.7**	**14.0**	**36.5**	**33.8**
	Phoneme	24.8	25.5	37.5	29.2
	Schwa	**38.5**	**50.0**	**69.7**	**52.7**
Adv. JS	Stress	**15.3**	**12.5**	**33.0**	**20.2**
	Phoneme	18.9	19.4	43.1	23.2
	Schwa	**31.9**	**51.4**	**65.3**	**49.5**
NS	Stress	**5.4**	**19.0**	**35.7**	**20.0**
	Phoneme	17.9	32.1	41.1	30.3
	Schwa	**35.7**	**35.7**	**49.9**	**40.4**

performances of the three tasks and the degrees of memory load.

In terms of simple main effect, there was a main effect in *Contrast* ($p = .000 < .05$) and *Memory load* ($p = .000 < .05$), but it was not obtained for *Group* ($p = .941$). Post hoc comparisons within *Contrast* indicated there is a significant effect for *Stress* vs. *Schwa* ($p = .000 < .05$) and *Phoneme* vs. *Schwa* ($p = .000 < .05$), demonstrating that the participants have more difficulty in distinguishing the contrast between a schwa and [a] than contrastive lexical stress and consonantal phoneme contrasts. However, the significant difference was not yield for *Stress* vs. *Phoneme* ($p = .116$), showing that the participants performed in the two tasks in a similar way. In terms of the *Memory Load* factor, there was a significant effect of sequence for three-word vs. five-word ($p = .000 < .05$) and four-word and five-word ($p = .000 < .05$), but not for three-word vs. four-word ($p = .117$), revealing that overall the participants have difficulty in distinguishing the contrasts when more memory load was required.

5. Discussion

The purpose of this study is to examine whether or not Japanese learners perceive lexical stress and schwa/full vowel contrast in the same ways as NS at the different sound processing levels. The following sections discuss the results of the stress contrast task and the schwa/full vowel contrast task respectively.

5.1 Stress contrast task

In terms of stress contrast, JS regardless of their English fluency were able to perceive a stress contrast which was instantiated by three acoustic cues: *F*0, duration and intensity, at the deeper processing level in the experiment in the same manner as native speakers of English do.

This result is consistent with previous studies. Altmann (2006) showed that L2 speakers with a non-stress first language background (such as Chinese, Japanese and Korean) are good at perceiving stress based on the Stress Typology Model (STM). Chan (2005) showed that Cantonese-English bilinguals, who take advantage of using a pitch cue to perceive accent in their lexical words can perform in the same way as NS in the perception of stress contrast tasks, using the same experimental paradigm as the current experiment. Thus, it is reasonable to assume that language listeners who utilize pitch cue to recognize their L1 words like Japanese are also able to apply this strategy to other languages. In order to verify the explanation, however, a follow-up experiment is necessary, for example, an experiment in which the pitch cue is made less available by systematically manipulating it in lexical words, with the intensity and duration cues kept constant. In this way, it can be proved that the results in the previous experiments are attributed to the genuine representation of lexical stress in short-term memory.

5.2 Schwa/full vowel contrast task

As for the schwa/full vowel contrast task, the statistical analysis showed that no difference was found among the three groups, showing that there was no interaction between *Group* vs. *Contrast*. However, as aforementioned, JS made more errors numerically (1.3 to 1.4 times as many as those of NS) in the highest memory condition, implying that JS had more difficulty in distinguishing between [paFU] and [pəFU] than NS did. Also, the fact that even Adv. JS made almost the same percentage of errors as Beg. JS showed the difficulty of acquiring /ə/ after achieving a certain level in overall English skills. A possible reason for this result may be due to the lack of a central vowel in the Japanese vowel inventory. Also, the sound /ə/ might be marked and intrinsically difficult to acquire. L1 phonological acquisition research – although it is from the speech production perspective – has shown that children take time to acquire schwa in speech production (Kettemann & Wieden, 1993). Thus, it is expected that JS also take time to fully establish the representation of schwa.

6. Conclusion

The results obtained from this experiment lead to two general conclusions. First, speakers who does not have lexical stress, but use pitch cue for perceiving words in their L1, are able to perceive lexical contrastive stress at a deeper level than the phonetic surface level in the same manner as native speakers of English. In other words, they might be able to store and trigger the metrical structure of lexical stress at the phonological representation level.

Second, regarding the schwa task, it seems that distinguishing schwa from /a/ is difficult for both advanced JS and NS, probably due to the intrinsically short length of the stimuli. However, it was observed that both JS groups made more errors in the four- and five-word sequences in the

task compared to NS. That is, the JS groups could not deal with the sound in the same way as NS when they were required to process it at the deeper sound processing level.

In future studies – although the stimuli in this study were intentionally created with [CVCV] phonotactics to avoid giving the advantage to NS in the perception task – by conducting the experiment using real English words or nonce words with English phonotactics and adding words with a variety of different syllable patterns as stimulus, it might be possible to clearly reveal JS and NS's speech processing patterns.

[1] The capitalized letters stand for stressed syllables.

[2] In Dupoux et al. (2001), where the same experimental paradigm was used as this study, there were differences of 45.3 Hz in F0, 20 msec. in duration, and 1.6 dB in intensity between stressed and unstressed syllables in the contrastive stress stimuli.

References

Altmann, H.(2006). *The perception and production of second language stress: A crosslinguistic experimental Study.* Ph.D. dissertation, University of Delaware.

Altmann, H., & Vogel, I. (2002). *L2 Acquisition of stress: the role of L1.* Paper presented at the DGfS annual meeting "Multilingualism Today" in Mannheim, Germany, March 2002.

Beckman, M. E. (1986). *Stress and non-stress accent.* Dordrecht: Foris.

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

Chan, M. K. (2005). *The processing and representation of lexical stress in short-termmemory of Cantonese-English successive bilingual.* Unpublished MA thesis, University of Hong Kong, Pokfulam Road, Hong Kong.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45,* 207-228.

Cutler, A., & Clifton, C. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp.183-196). Hillsdale, NJ: Erlbaum.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language, 36,* 3, 406-421.

Dupoux, E., Peperkamp, S., & Sebastian, N. (2001). A robust method to study stress "deafness". *Journal of the Acoustical Society of America, 110,* 3, 1606-1618.

Flege, J. E. (1992). Speech learning in a second language. In C. A. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: models, research, and implications* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 229-273). Timonium, MD: York Press.

Haraguchi, S. (1977). *The tone pattern of Japanese: An autosegmental theory of tonology.* Tokyo: Kaitakusha.

Kettemann, B., & Wieden, W. (Eds.). (1993). *Current issues in European second language acquisition research.* Tübingen: Narr.

Matthews, J., & Brown, C. (1998). *Qualitative and quantitative differences in the discrimination of second language speech sounds,* Proceedings of the Boston University Conference on Language Development, 22, 499-510.

Morton, J., Crowder, R. G., & Prussin, H. A. (1971). Experiments with the stimulus suffix effect. *Journal of Experimental Psychology Monograph, 91,* 169-190.

Morton, J., Marcus, S.M., & Ottley, P. (1981). The acoustic correlates of "speechlike": A use of the suffix effect. *Journal of Experimental Psychology: General, 110,* 568-593.

Peterson, G. E., & Barney, H. l. (1952). Control methods used in the study of vowels. *Journal of the Acoustical Society of America, 24,* 75-184.

Sugito, M. (1969). Tokyo, Osaka ni okeru musei-boin ni tsuite [A study of voiceless vowels in Tokyo and Osaka]. *Onsei no Kenkyu [Study of Sounds], 14,* 249-264.

Wallace, K. (1994). *An acoustic study of American English Schwa in multiple speaking modes.* Unpublished PhD dissertation, New York University, New York. N.Y.

Werker, J. F., & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics, 37,* 35-44.