The Phonological Representation and the Use of Suprasegmental and Segmental Information in Spoken-Word Recognition by Japanese Speakers of English

Kaori Sugiura
(Shizuoka University of Art and Culture)

Although stress is an important aspect of L2 acquisition, few L2 research on the perception of the suprasegmental sounds has been conducted. Additionally, a reduced vowel, schwa, which is a necessary segment to build the metrical structures of English, has not been highlighted in L2 phonological acquisition. Using a sequence recall task, developed by Dupoux et al. (2001), the study examined whether or not Japanese speakers of English (JS) were as sensitive to lexical stress and schwa at the phonological level as native speakers of English (NS). A statistical analysis showed that there was no significant difference between JS and NS in terms of perceiving lexical stress. With respect to schwa, a statistically significant difference was not revealed between JS and NS's performances; however, the numerical result showed the tendency that JS had more difficulty in perceiving schwa in the sequences of 4 and 5 words as compared to the 3-word sequence. But this was not the case for NS.

Keywords: lexical stress, schwa, phonological representation, L2 learners of English

1. Introduction

English is one of the few languages which exploit both suprasegmental (i.e., stress) and segmental information (i.e., schwa) in perception of lexical stress (Cutler and Clifton, 1984; Fear et al., 1995; Cooper et al., 2002). Therefore, acquiring both phonological aspects is crucial for second language (L2) learners of English. A language employs different suprasegmental and segmental features to make lexical distinctions, L2 learners might perceive English lexical stress differently from NS or misperceive it.

Fewer L2 researches on the perception of suprasegmental sounds have been studied as compared to the research on segmental aspects. In addition, previous researches have not highlighted L2 perception of schwa despite the fact that it is a necessary segment to build the metrical structures of English.

The aim of the study is to investigate whether Japanese speakers of English (JS), whose native language is non-stress language, can encode contrastive stress and contrastive vowel (i.e., schwa /ə/ vs. /ɑ/) in their phonological representations, which play an important role in the language perception and production (Fowler, 1991).

2. Experiment

2.1 Research questions

1) Are JS as sensitive to the stress contrast

as NS at phonological level?

2) Are JS as sensitive to the vowel contrast (i.e., schwa vs. full vowel /ɑ/) at the phonological level as NS?

2.2 Experimental design

The experiment consisted of the contrastive stress (i.e., [MIpɑ] vs. [miPA][1]), consonantal (i.e., [TUki] vs. [TUpi]) and vocalic (i.e., [pɑFU] vs. [pə FU]) tasks. Each task was constructed with three blocks: the first block contained the sequences of three words, the second one consisted of those of four words, and the third one had those of five words. All the blocks had eight different sequences. All the selected sequences are listed in Table 1.

Since this experiment was designed for the purpose of assessing one's phonological representations, the following aspects were considered.

・By using the sequences of three to five words, and this experiment gradually increased the burden of memory for participants.

・By providing some phonetic variability in each word, manipulating a pitch, more abstract phonological representations rather than acoustic, phonetic representations were assessed.

・In order to prevent the participants from using echoic memory, every sequence was followed by the word "O.K." (Morton, Crowder, and Prussin, 1971; Morton, Marcus, and Ottley, 1981), and they could not begin typing their responses until they had heard this word.

・The speed of presentation was kept very short, specifically, 80 msec. in order to diminish the likelihood that the participants used the strategy in which they mentally translated the words into the associated numbers while listening to the sequence.

Table 1 Types of sequences

Sequence and 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

Notes: The indicted numbers were computer keys. "1" is associated with [MIpa] and "2" with [miPA] in the case of the contrastive stress task.

2.3 Materials

The stimuli were created by 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 on a computer disk. The mean durations of each stimulus were manipulated to be as equal as possible.

Table 2 Detail acoustic characteristics of the stimuli

[1] The capitalized letters stand for stressed syllables.

A: Duration of three contrastive stimulus (msec.)

	stress	consonant	vowel
MIpa:	307	tuki 354	pafu: 310
miPA:	311	tupi 353	pəfu: 304

B: Stress: Mipavs.miPA

	duration:msec.		pitch: Hz		intensity: dB	
First /Second syllable						
MIpa:	141	96	160	84	56	54
miPA:	89	140	100	157	56	63

C: Schwa: paFU vs. pəFU

	F1 & F*2:Hz	duration: msec.		pitch:Hz		Intensity:dB	
First /Second syllable							
pafu:	923 /1428	53	241	161	232	78	70
pəfu:	697/1955	51	247	175	236	56	66

*F=formant frequency

D: Consonant: TUki vs. TUpi

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

2.4 Procedures

1. The participants were told that they were going to learn two nonwords. They could listen to the tokens of the two words as many times as they wanted. While the participants listened to [MIpa], [1] was being shown on a computer screen. They, then, listened to its counterpart [miPA], and [2] was being 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 nonwords as well as the correct association between the words and the number keys, specifically, [MIpa] for key [1] and [miPA] for key [2]. They took the four trials in which they had to reproduce each sequence they heard by typing the associated keys in the correct order and received, then, the feedbacks.
3. During the test, the participants listened to twenty-four sequences and reproduced each one by typing the computer key. For each participant the order of the eight sequences in each block was randomized. The time interval between the trials was 1500 msec..(The same procedures were repeated for the consonantal contrast [TUki] vs. [TUpi] and the vocalic contrast [paFU] vs. [pə FU] tasks as well).

2.5 Participants

Seven NS aged between late 10s and 40s, JS aged between 20s and 50s with advanced (adv.) English proficiencies, and thirteen JS aged between18s and 50s with intermediate (inter.) English proficiency participated in the experiment. Note here is that adv. speakers started to learn English their target second languages from junior high school in Japan. All the participants did not have any problems with hearing and speaking.

2.6 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 responses were coded as incorrect. A participant with 100 % incorrect responses in one of the three tasks (stress contrast, schwa /ə/ vs. /ɑ/ and consonantal contrast) was rejected. The high percentage incorrectness suggest that they might have confused the number key associated with the first and second sound items or have not concentrated on the experiments at all.

3. Results

Error percentages for inter. JS, adv. JS and NS participants for the contrasts as a function of sequence lengths are shown in Table 3.

Table 3 Error rates (%) for the contrast as a function of sequence length

	Contrast/ Sequence	3 words	4 words	5 words
Inter.JS	Stress	7.7	14.0	36.5
	Consonant	24.8	25.5	37.5
	Schwa	38.5	50.0	69.7
Adv.JS	Stress	15.3	12.5	33.0
	Consonant	18.9	19.4	43.1
	Schwa	31.9	51.4	65.3
NS	Stress	5.4	19.0	35.7
	Consonant	17.9	32.1	41.1
	Schwa	35.7	35.7	49.9

A statistical analysis with generalized linear models (GLM) was conducted. The dependent variable was *Error rate* and the independent variables were *Memory load* (3-, 4- and 5-word sequences), *Group* (inter. JS, adv. JS and NS) and *Contrast* (stress, schwa and consonant). In addition, an analysis of an ANOVA with the factor: *Contrast* for each group was separately conducted.

In terms of the analysis with GLM, there was a main effect *of Contrast (p=.000<.05) and Memory load* (p=.000<.05), but there was not no main effect of *Group* (p=.941>.05). Post hoc comparisons in *Contrast* indicated a significant effect for Stress vs. Schwa (p=.000 <.05) and Consonant vs. Schwa (p=.000 <.05), but not for Stress and Consonant (p=.116>.05). There was a significant effect of sequence for 3-word vs. 5-word (p=.000<.05) and 4-word and 5-word (p=.000 <.05), but not for 3-word vs. 4-word (p=.117). However, there was no interaction between *Group* vs. *Contrast* (p=.573>.05), *Contrast* vs. *Memory load* (p=.906>.05), and *Group* vs. *Memory load* (p=.978>.05).

An analysis of an ANOVA with the factor: *Contrast* for each group was conducted separately. As indicated in Table 4, there was a main effect of *Contrast* for each group. Post hoc tests showed that both adv. and inter. JS had significant difference between Schwa (i.e., schwa vs. full vowel) and the Consonant contrast (i.e., TUki vs. TUpi) but this was not the case for NS.

Table 4 ANOVA results for the error data of the three groups (*p <.05)

Group	Inter. JS	Adv. JS	NS
Contrasts	F(1,14) =20.574 p=.00*	F(1.10) =13.84 p=.00*	F(1,8) =3.876 p=.027*
Schwa vs. Con.	p=.00 *	p=.00*	p=.374
Schwa vs. Str.	p=.00*	p=.00*	p =.020*
Stress vs. Con.	p=2.39	p= .701	p= .331

4. Discussion

4.1 R.Q 1: Are JS as sensitive to the stress contrast as NS at phonological level?

The results of the statistical analysis showed no interaction *Group×Contrast*, suggesting that JS were able to perceive a stress contrast at the phonological level in the same manner as native speakers of English. The good performance of the JS with regard to the perception of stress can be explained as follows.

First, in order to exploit the stress contrast, the JS might have utilized the pitch cue, which is primarily used to distinguish Japanese accent words. Chun (2005) has held a similar view to account for the good performances of Cantonese–English bilinguals in the perception of stress contrasts. Although the similarity between non-stress language speakers of Cantonese and Japanese with regard to the perception of stress is probably due to their use of a pitch in stress perception, to verify the explanation, a follow-up experiment will be 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 (Chun, 2005).

In addition, the results obtained in this study could be explained with Stress Typology Model (STM), proposed by Vogel (2000) and Altman and Vogel (2002). This model predicts that L2 speakers of non-stress first languages such as Chinese, Japanese and Korean are good at perceiving stress because they do not have any positive L1 parameter setting for the model stress that can possibly interfere with the L2 settings, whereas 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 impeded the acquisition of L2 stress. Recently, Heidi (2006) clearly has supported the STM by systematically examining the typologically different languages. Thus, JS's successful performance in the perception of lexical stress might be explained by these reasons.

4.2 R.Q 2: Are JS as sensitive to the vowel contrast (i.e., schwa vs. full vowel /ɑ/) as NS at the phonological level?

It was expected that JS groups were not as sensitive to the contrast, schwa vs. full vowel /ɑ/, as NS since the contrast does not exist in their native language. In addition, we predicted that even adv. JS were not able to perceive the contrast in the same way as NS. Since they are late learners of English, who started to learn English after a critical period, it could be assumed that it is difficult for them to establish new phonological representations (Lively, Logan, & Pisoni, 1993; Takagi & Mann, 1995; Pallier et et al., 1997; Sebastián-Gallé & Soto-Franco, 1999; Pallier et et al., 2001; Sebastián-Gallé et al., 2005; Dupoux et al.,2008).

Unexpectedly, however, the statistical analysis revealed that there was no interaction *Group×Contrast,* suggesting that there was no difference among three groups (i.e., two JS groups and NS group). But as indicated in Table 3, a numerical result showed the tendency that both JS groups have more difficulty in perceiving contrastive vowels (i.e., schwa vs. /ɑ/) as the word length becomes longer as compared to NS does (inter. JS: 38.5% for 3 words, 50% for 4 words, 69.7% for 5 words; adv. JS: 31.9 % for 3 words, 51.4 % for 4 words, 65.3% for 5 words; NS: 35.7% for 3 words, 35.7% for 4 words, 49.9% for 5 words).

One of the reasons that the statistical difference between NS and JS was not observed might be due to the fact that it was difficult even for NS to perceive the segmental items (i.e., schwa vs. full vowel /ɑ/) because of its short length in nature.

Another reason might lie in the phonotactic problem in the stimuli. Schwa rarely appears in the first syllable in English dissyllabic words, but the stimuli in this study were intentionally created as [pəFU] to avoid giving the advantage to NS in the task of the perception of schwa. However, the experiment using schwa in real English words or nonwords with English phonotactics also should be conducted in the near future.

5. Conclusion

This study examined how non-stress language speakers, Japanese, dealt with the phonological properties of the metrical structures in English. The result showed that JS could perceive stress in the lexical word at the phonological level in the same manner as the NS did. In terms of the segmental feature, schwa, there was also no statistical difference between JS and NS. Numerical results; however, showed that JS seemed to have more difficulty in distinguishing schwa from /ɑ/ as compared to NS.

Selected references

Chun, M. K. K. (2005). *The Processing and Representation of Lexical Stress in the Short-term Memory of Cantonese-English Successive Bilinguals.* Unpublished Master's thesis, University of Hong Kong.

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

Dupoux, E., Sebastian-Galles, N. Navarete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition*, **106(2)**, 682-706.

Heidi, A. (2006). *The perception and production of second language stress:a cross-linguistic experimental study.* Doctoral dissertation, University of Delaware.

Logan, J.S., Lively, S.E., & Pisoni, D.B. (1993). Training listeners to perceive novel phonetic categories: How do we know what is learned? *Journal of the Acoustical Society of America*, **94**, 1148-1151.

Vogel, I. (2000). *The Acquisition of Prosodic Phonology: Challenges for the L2 Learner.* Paper presented at "Structure, Acquisition, and Change of Grammars: Phonological and Syntactic Aspects" in Hamburg, Germany, Oct 2000.