University of Pennsylvania Working Papers in Linguistics

Volume 16
Issue 1 *Proceedings of the 33rd Annual Penn Linguistics Colloquium*

Article 22

2010

The Perception of Complex Onsets in English: Universal Markedness?

Mieko Sperbeck
City University of New York

Winifred Strange
City University of New York

The Perception of Complex Onsets in English: Universal Markedness?

Abstract

Second language (L2) learners of English whose native languages have relatively simple syllable structure have a strong tendency to modify complex onsets in production. Past studies have shown that such modification is often correlated with sonority-based markedness. According to this principle, the marked bi-consonantal sequences are such that the sonority distance between the first consonant and the subsequent consonant is relatively small. For instance, /pl/ is considered to be less marked than /bl/ since the former has larger sonority distance. A question of interest here is whether such "markedness" would be applicable to the perception of complex onsets by Japanese-speaking learners of English. The current study tested Japanese L2 learners and American English controls in a categorial ABX discrimination test of 8 contrasts between nonsense words with consonant cluster onsets CC(C)VCV vs. CVC(C)VCV sequences (e.g., /spani/ vs. /sepani/) and included /sp, sk, pl, bl, kl, gl, spl, skl/ clusters. Results showed that overall accuracy by Japanese listeners was significantly poorer than for the Americans (72 % and 98% correct, respectively). Certain clusters were harder for Japanese listeners (e.g., 76% correct for /pl/ but 64% for /bl/). However, in general, relative difficulty was not accurately predicted by sonority-based markedness. Alternative hypotheses for relative perceptual difficulties include the acoustic characteristics of the stimulus materials and effects of native phonological structures.

The Perception of Complex Onsets in English: Universal Markedness?

Mieko Sperbeck and Winifred Strange*

1 Introduction

Late second language (L2) learners have a strong tendency to transform L2 syllable structures into native ones in producing L2 speech. For example, Japanese-speaking learners of English often break up English consonantal clusters by inserting a vowel [u] (e.g., [s*u*p*u*rey] for "spray"). Since complex onsets are not allowed in Japanese, an English word such as "spray" cannot be mapped into a single syllable. A question of interest here is how Japanese listeners perceive phonotactically illegal consonantal sequences and whether they have different degrees of difficulty in perceiving them.

In generative phonology, the creation of a syllable is regulated by the Sonority Sequencing Principle, which states that sonority rises toward the nucleus of a syllable and lowers away from it (Clements, 1990). Each phoneme can be categorized in terms of a sonority scale where low vowels are considered as the most sonorous, followed by mid vowels, high vowels, flaps, laterals, nasals, voiced fricatives, voiceless fricatives, voiced stops, and voiceless stops (Hogg and McCully, 1987:33). Possible exceptions to the Sonority Sequencing Principle in English are /s/ plus voiceless stop sequences (e.g., /sp, st, sk/) since sonority between two phonemes falls toward the nucleus.

Markedness is related to frequency of a linguistic structure in languages of the world and is often applied to predict relative difficulty in acquisition (see Eckman, 2008 for reviews). The marked linguistic forms are less frequent and considered to be harder to acquire. For instance, triconsonantal (hereafter, CCC) sequences are regarded as more marked than bi-consonantal (hereafter CC) sequences. Research in first language acquisition often reports that young children tend to produce CC clusters more accurately than CCC clusters (Gierut and Champion, 2001; Smit, 1993). With respect to the acquisition of CC sequences in English, the markedness depends on the sonority distance between two phonemes of a complex onset. For instance, complex onsets such as /bl/ (e.g., "*bl*ow") are considered as more marked than /pl/ (e.g., "*pl*ay"). The sonority distance between /b/ and /l/ is smaller than that of /p/ and /l/ since voiced stops are more sonorous than voiceless stops and thus, closer to laterals in sonority.

Such sonority-based markedness has also been explored in the acquisition of L2 complex onsets (e.g., Broselow and Finer, 1991; Broselow, 1995; Broselow, Chen, and Wang, 1998; Carlisle 1997, 1998, 2006; Cardoso 2008; Eckman 1977, 1981, 1991, 1996, 2004; Hancin-Bhatt and Bhatt, 1997; Major, 1996). These studies have often tested whether universally more marked linguistic structures are harder to acquire than less marked ones regardless of learners' first language. Eckman states such a hypothesis in his Structural Conformity Hypothesis as follows: "the universal generalizations that hold for primary languages hold also for interlanguages" (Eckman, 1991: 24).

Eckman (1991), for instance, examined L2 learners' productions of three sets of related onsets: /spr/, /sp/, and /pr/; /str/, /st/, and, /tr/; and /skr/, /sk/, and /kr/. The selection of these stimuli was based on the Resolvability Principle, which is defined as "if a language has a consonantal sequence of length *m* in either initial or final position, it also has at least one continuous subsequence of length *m*-1 in this same position" (Eckman, 1991: 25). Participants were Japanese, Cantonese, and Korean, whose first languages do not allow initial consonant clusters. The tasks that participants performed included reading words from a list, eliciting a word from the picture, and having a natural conversation with native speakers of English. Results showed that the error rate was correlated with markedness. That is, participants had more difficulties in producing marked sequences (e.g., /spr/), and thus the author concluded that the Structural Conformity Hypothesis was confirmed.

*We would like to thank the audience at the 33rd Penn Linguistics Colloquium, colleagues at the Speech Acoustics and Perception Lab at the CUNY Graduate Center, and the Phonetics & Experimental Phonology Lab at New York University for their suggestions and comments on this study. This study is supported by the CUNY Graduate Center research grant (2008-2009).

With respect to the acquisition of CC sequences among L2 learners, Broselow and Finer (1991) also reported that more marked structures were harder to acquire than less marked ones. Production of six types of complex onsets were examined among Japanese and Korean-speaking learners of English: [pr], [br], [fr], [pj], [bj], [fj]. Based on the markedness principle of sonority, Broselow and Finer predicted relative difficulty to increase from least marked to most marked: [pj]—[pr]—[bj]—[br]—[fj]—[fr]. Participants were, first, asked to learn the pseudo-words that began with these six types of consonant clusters and then, to produce them. The production data was transcribed by native speakers of English. Results showed that the more marked clusters had higher error rates than the less marked ones for both language groups. The authors concluded that universal markedness of sonority applied to the production of L2 syllable onsets.

To summarize, past research on the production of consonantal sequences has shown that not all consonantal sequences are acquired equally. Rather, the acquisition seems to be constrained by sonority-based markedness. That is, the more marked sequences produce more errors than the less marked ones (however, see Davidson 2006 for the articulatory-based account of the production of nonnative complex onsets).

Perceptual studies in the past have revealed that L2 learners' inaccurate production of the target language is often a result of inaccurate perception of nonnative speech forms (see Strange, 1995; Strange and Shafer, 2008, for reviews). With respect to the perception of syllable structures, past studies have shown that listeners' native phonotactic properties seem to bias perception of nonnative phoneme sequences (e.g., Altenberg, 2005; Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi & Mehler, 2001; Fais, Kajikawa, Werker, & Amano, 2005; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak and Idsardi, 2007; Sperbeck and Strange, 2008). In other words, perception research has shown that the L2 input is not necessarily parsed by L2 learners in a native-like manner.

Dupoux, Kakehi, Hirose, Pallier, and Mehler (1999) tested Japanese listeners' perception of consonantal clusters that occurred word-medially (e.g., /abge/). Using a categorial ABX discrimination task, two types of contrasts were tested: an epenthesis contrast (e.g., "*abge-abuge*") and a vowel length control contrast (e.g., "*abuge-abuuge*"), the latter being phonologically distinctive in Japanese. Japanese participants heard three stimuli "ABX" and responded whether "X" was the same as "A" or "B". Results showed that Japanese participants had more difficulty with the epenthesis contrast than with the vowel length contrast (16% vs. 4% errors, respectively). Dupoux et al. interpreted these results as showing that Japanese listeners perceived "*abge*" as the perceptually-modified form, "*abuge*". That is, the authors claimed that the modification of L2 syllable structures occurs at the level of L2 speech perception. However, results of individual stimulus materials were not mentioned, and thus the relationship of the results to the markedness for L2 perception could not be examined.

Berent, Steriade, Lennertz, and Vaknin (2007) tested whether perception of onset clusters was affected by the universal markedness of sonority. Three types of contrasts based on the markedness were tested: sonority rises (e.g., "*bwif-bəwif*"), sonority plateaus (e.g., "*bdif-bədif*"), and sonority falls (e.g., "*lbif-ləbif*"). Participants were native speakers of American English, whose phonology did not allow the consonantal sequences that were used for the experiment, as well as native speakers of Russian, whose phonology allowed all the clusters used for the experiment. Berent and colleagues hypothesized that accuracy in identifying onset clusters would correlate with the markedness of onsets. That is, the rate of vowel epenthesis would be higher for the more marked onset clusters (e.g., "*lbif*") than the less marked ones (e.g., "*bnap*") since the sonority fell in the former type of clusters. The task was a forced choice identification task where listeners selected whether the auditory stimulus had one syllable or two syllables by pressing the button. Results showed that the overall percent correct for three types of clusters reflected universal markedness of sonority among native speakers of English. That is, as the markedness increased, the accuracy rate fell. The overall percent correct was about 60% for sonority rise, 30% for sonority plateau, and 15% for sonority fall, respectively (Berent et al., 2007:604; see Figure 1). Such effects of markedness were also obtained in the AX discrimination task where listeners had to respond whether two stimuli (e.g., "*lbif – lebif*") were the same or different. Berent et al. interpreted these results as showing that markedness of sonority operates in perceiving unattested onset clusters.

The current study investigates perception abilities of Japanese late learners of English on initial consonant clusters in English. If the epenthesized vowel in productions (e.g., the *u* and *o* in [s*u*po:t*o*] for "sport") is perceptually-derived as Dupoux et al. (1999) claim, it is predicted that Japanese participants will have difficulty in perceiving differences between structures such as CəCVCC (e.g., "support") vs. CCVCC (e.g., "sport"), which differ in word onset. Additionally, it is predicted that Japanese participants will show different degrees of difficulty in perceiving complex onset clusters if markedness operates in L2 perception as shown in Berent et al. (2007).

The choice of the stimulus materials was inspired by Eckman's 1991 study, which tested the production of sets of related onsets (e.g., /skr/ - /kr/ - /sk/). Since the production and perception of /r/ is known to be notoriously hard for native speakers of Japanese (e.g., Bradlow, 2008), /r/ was avoided as stimulus materials. The perception of two sets of related onsets was examined: /spl/, /sp/, and /pl/ and /skl/, /sk/, and /kl/. Additionally, stimulus materials included /bl/ and /gl/ in order to see the effects of markedness within CC sequences.

	More marked	Less marked
1	CCC sequences (e.g., /spl/)	CC sequences (e.g., /sp/)
2	voiced stop + /l/ (e.g., /bl/)	voiceless stop + /l/ (e.g., /pl/)
3	/s/ + voiceless stop (e.g., /sp/)	voiced & voiceless stop + /l/ (e.g., /pl/)

Table 1: Predicted difficulty based on the markedness

Based on the sonority-based markedness, it is predicted that the perception of CCC sequences will be harder than that of CC sequences since the former has more complex linguistic structures. In addition, the accuracy in perceiving voiceless stop plus /l/ clusters should be higher than voiced stop plus /l/ clusters since the former has larger sonority distance between two phonemes. With respect to the perception of the initial /s/ plus voiceless stop clusters, /s/-clusters are predicted to be more marked than other CC sequences since the sonority distance between two phonemes is smaller. Although the status of the initial /s/-clusters as a complex onset has been controversial since these clusters do not follow the Sonority Sequencing Principle (e.g., Kenstowicz, 1994), it is assumed here that there is no structural distinction between /s/- and non-/s/ clusters, following the analysis of Boyd (2006).

2 Method

2.1 Participants

Thirty native speakers of Japanese (19 female and 11 male; 23 to 52 yrs; mean = 32 yrs) served as the experimental group while five native speakers of American English (three female and two male; 23 to 34 yrs; mean = 28 yrs) formed the control group. Japanese participants' mean length of residence in the United States was 3.5 years (range one month to 20 years) and their mean age of arrival was 29 years old (range 19 to 39 years). All participants passed a pure-tone hearing screening prior to the experiment and were paid twenty dollars for their participation.

2.2 Stimulus Materials

A female speaker of American English who was a linguist produced stimulus materials for the experiment. Nonsense words were of the form /CC(C)áni/ and /CəC(C)áni/. An unstressed schwa served as the vowel between initial and subsequent consonants. There were six CC contexts (/sp, sk, pl, kl, bl, gl/) and two CCC contexts (/spl, skl/). These target words were produced in a short carrier sentence, "Say ____ now". Four tokens of each word were digitally recorded in a sound-attenuated room at 22.050 kHz with 16-bit resolution, using SOUND FORGE 4.5. The speaker was instructed to speak with normal conversational effort.

Acoustic measurements of the target words were made using Praat (Boersma & Weenink, 2005). The absence of an epenthetic vowel in the CC(C)ani materials (e.g., "splani") and its presence in the CVC(C)ani stimuli (e.g., "seplani") were verified. Durations of the schwa in the CəC(C)ani materials were measured (see Table 2 below). The average duration of schwa for all cluster types was 38 ms (SE = 2.3).

In addition, formant frequencies of these schwas in the CəC(C)ani materials were measured. Since the vowels were short and steady, measurements for the first two formants, F1 and F2, were taken by highlighting the vowel in selection from a Formant object in Praat (see Table 2, rows 2 and 3).

	/pəl/	/kəl/	/bəl/	/gəl/	/səp/	/sək/	/səpl/	/səkl/
Duration (in ms)	30	48	44	52	24	33	30	45
F1 (in Hz)	611	684	620	640	514	456	493	502
F2 (in Hz)	1263	1346	1322	1452	1860	1950	1758	2020

Table 2: Mean duration (in milliseconds) and formant frequencies (in Hertz) of schwa for each type of stimulus.

Moreover, the duration of other segments of the target words were measured. Mean durations of each segment are shown in milliseconds in Tables 3, 4, 5 below. /lá/ in the C(ə)láni-type materials (e.g., /pláni/) was measured as one unit, as it was sometimes difficult to segment between these two phonemes. With respect to the voiced and voiceless stop plus /l/ pairs shown in Table 3, note that the duration of Voice Onset Time (VOT) was distinctively longer in voiceless stop + /l/ stimuli (range = 37 - 96 ms) than in voiced stop + /l/ stimuli (range = 7 - 35 ms). Such durational difference was especially significant for /kláni/ (t [2] = 8.5, $p < 0.05$). The duration of /la/ was not significantly different among stops + /l/ clusters.

With respect to /s/ plus voiceless stop (sC) pairs shown in Table 4, note that the VOTs were significantly longer in /səC/ sequences (range = 7 – 20 ms) than in /sC/ sequences (range = 41 – 50 ms). Neither the duration of frication nor that of silence seemed to have signaled as distinctive acoustic cues, as there were statistically no significant differences among them in /sC/ and /səC/ sequences. These acoustic characteristics seemed to be also true for CCC contexts. That is, the VOTs were distinctively longer in /səCC/ sequences (range = 38 – 56 ms) than in /sCC/ sequences (range = 8 – 23 ms). Additionally, it showed that there was significant difference in the duration of /la/ in /splani - səplani/ pairs (t [2] = 6.2, $p < 0.05$), which might serve as an additional acoustic cue for listeners.

Stimuli	VOT	/la/		Stimuli	VOT	Vowel	/la/
/bl/	9	228		/bəl/	7	44	258
/gl/	23	229		/gəl/	33	52	255
/pl/	59	200		/pəl/	47	30	257
/kl/	83	191		/kəl/	49	48	225

Table 3: Mean duration of each distinct segment for stop + liquid pairs (in milliseconds).

Stimuli	Frication	Vowel	Silence	VOT
/sp/	139	0	79	7
/sk/	145	0	58	19
/səp/	126	24	85	44
/sək/	123	33	68	47

Table 4: Mean duration of each distinct segment for /s/-clusters (in milliseconds).

Stimuli	Frication	Vowel	Silence	VOT	/la/
/spl/	106	0	94	11	198
/skl/	126	0	55	22	164
/səpl/	120	30	86	39	163
/səkl/	111	45	59	52	172

Table 5: Mean duration of each distinct segment for tri-consonantal sequences (in milliseconds).

2.3 Procedure

The experiment was carried out individually in a sound-attenuated room at the Graduate Center, CUNY. The perception task was created with Paradigm beta-5d (written by Bruno Tagliaferri, 2008). A categorial ABX task was employed in which participants heard three short sentences in a row that contained the target words (e.g., “say spani now” – “say s*e*pani now” – “say spani now”) and answered whether the third target word was the same as the first or the second one by pressing a button (the first one is correct in this example). The tokens of each word in a triad were different productions so that there was no physical match between X and either A or B.

The perception experiment consisted of seven blocks. The first block was for familialization and thus, the score was not counted. Each block contained 32 trials (eight consonantal contexts (/sp, sk, pl, bl, kl, gl, spl, skl/) × 4 combinations (ABA, BAB, ABB, BAA). Thus, there were 224 trials (32 trials × 7 blocks) in total. The inter-trial interval was self-paced. The presentation of the trials was randomized for each participant, thus controlling for possible order effects.

3 Results

Table 6 shows mean percent correct for the perception task for the two groups. As predicted, overall accuracy in perception by the Japanese was poorer than for the Americans (72 % and 98% correct, respectively). There was greater variability within groups in the Japanese group (SD = 11.6) than in the American groups (SD = 1.5).

Participants	Mean	SD	SE
American (n = 5)	98	1.5	0.68
Japanese (n = 30)	72	11.6	2.12

Table 6: Percent correct by language groups: mean standard deviation (SD) and standard error of the mean (SE).

Figure 1 shows the mean percent correct for each stimulus type by each language group. Note that the American group showed ceiling effects for all types of consonantal sequences (98% to 100%). In contrast, the Japanese group's performance for all eight contexts was consistently lower (64% to 80%).

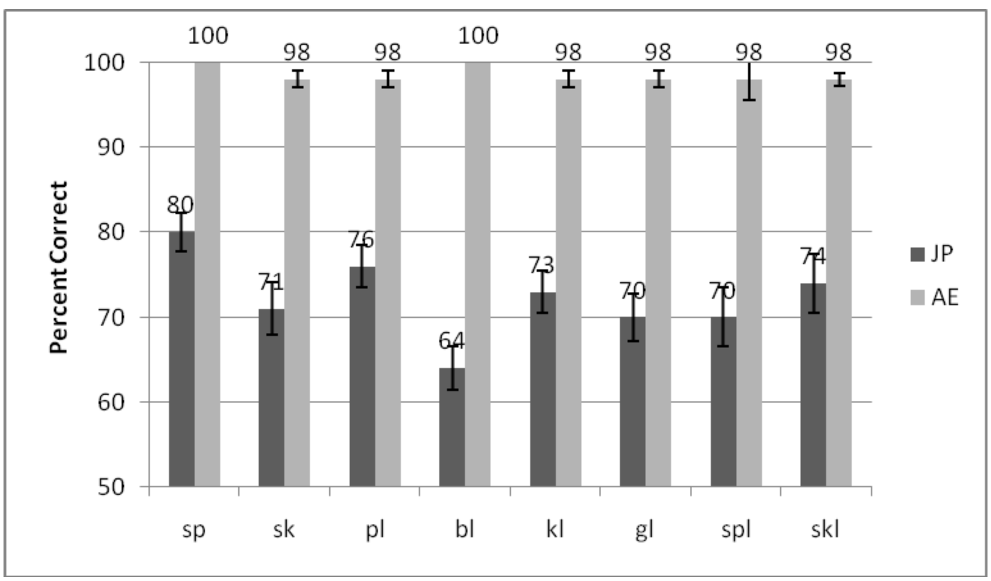

Figure 1: Mean percent correct for each stimulus type by language groups. JP = Japanese. AE = American English. Error bar represents standard error.

Since the American group's performance was at ceiling on all stimulus types, within-group comparisons were performed only for the Japanese data. The Friedman's test showed that Japanese listeners' performance differed significantly by stimulus types (χ^2 (7df) = 30.61, $p < 0.001$). The rank of average accuracy was inconsistent with the prediction made by markedness. For instance, despite the fact that /sp/ had the smallest sonority distance between the two consonants and thus was considered "the most marked" among CC sequences, the Japanese speakers performed best on this contrast (80% correct). As can be seen in Figure 1, overall performance seems to be divided into three groups; the high (/sp, pl/), the medium (/skl, kl, gl, sk, spl/), and the low (/bl/).

Stimuli	Mean	SD	SE
CC (e.g., /pl/)	73 (range = 57 - 97)	10.5	1.91
CCC (e.g., /spl/)	72 (range = 40 – 98)	18.2	3.33

Table 7: Overall percent correct by cluster types.

Table 7 shows the comparison of overall accuracy for CC vs. CCC conditions. Contrary to the sonority-based prediction, Wilcoxon signed-ranks revealed that there was no significant difference between CC and CCC contrasts ($T = 226$, $p > 0.05$). However, Japanese listeners' performance for CCC conditions was more variable (SD = 18.2) compared to those of CC conditions (SD = 10.5).

Stimuli	Mean	SD	SE
Voiceless stop + /l/	75 (range = 48 – 96)	11.0	2.01
Voiced stop + /l/	67 (range = 54 – 98)	12.6	2.31

Table 8: Overall percent correct by voicing types.

With respect to the voiced vs. voiceless stop + /l/ comparison (Table 8 above), significant differences were observed in Japanese subjects' performance ($T = 54$, $p < 0.01$). Consistent with the prediction, the voiced stop + /l/-clusters were harder than the voiceless stop + /l/-clusters (67% and 75% correct, respectively). However, a closer look at these results revealed that such significance was only true for the /bl – pl/ pairs. The average accuracy for the /gl – kl/ pairs did not differ significantly.

Stimuli	Mean	SD	SE
stop + liquid	71 (range = 53 – 96)	11.0	1.92
/s/ + voiceless stop	76 (range = 57 – 100)	12.6	2.6

Table 9: Overall percent correct of /s/-clusters and stop + liquid clusters.

Table 9 shows the comparison of the average accuracy of /s/-clusters and stop + liquid clusters. Although overall percent correct was higher for /s/-clusters, Wilcoxon signed-ranks revealed that statistically this difference was not significant ($T = 144.50$, $p > 0.05$).

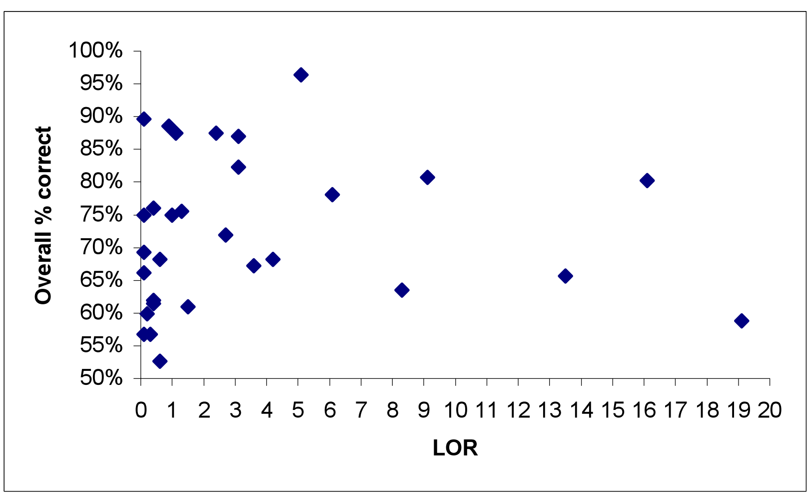

Figure 2: Japanese listeners' LOR and overall performance.

Lastly, further statistical analysis revealed that Japanese listeners' overall performance did not correlate with their length of residence (LOR) in an English speaking country (rho = +. 213), as Figure 2 shows. In the present study, Japanese participants' LOR was not controlled. The majority of participants had less than five years of LOR. Among them, there were 13 participants who had been staying in the U.S. less than one year at the time of the current experiment. There were only seven participants whose LOR were longer than five years. One possible explanation for the lack of correlation between Japanese group's performance and their LOR might be due to the small sample size for those with longer LOR. However, the great variability in performance for recent arrivals argues against this hypothesis. Another explanation might be due to difference in the amount of exposure to native English among those with shorter LOR. Four participants (JP 5, JP14, JP18, JP29) had 85% accuracy despite the fact that they had been staying in the U.S. less than one year at the time of the perception test. However, these four participants reported that they had had a good amount of exposure to native speakers' English while they lived in Japan. Thus, in the present study, LOR might not serve as a good index for accounting of English conversational experience.

To summarize, these results revealed that Japanese late L2 learners of English have difficulty in perceiving complex onsets. However, their overall performance did not support the sonority-based markedness hypothesis about relative difficulty. The predicted difficulty was only true for the /bl-pl/ pairs. Thus, markedness cannot, by itself, account for relative perception difficulty of English complex onsets.

4 Discussion

4.1 The more acoustic cues, the better perception?

The current study tested whether more marked structures were harder to perceive than less marked structures by L2 learners. Results showed that Japanese listeners' performance on different complex onsets did vary, but markedness alone did not accurately predict significant differences. What other factors contributed to perceptual difficulty? For instance, what caused the significant difference in performance for the /bl-pl/ pairs but not the /gl-kl/ pairs? In other words, why was the perception of /bl/ so difficult for Japanese subjects?

One possible explanation would be acoustic characteristics of the blani-belani stimuli. As mentioned in 2.2 above, the mean VOT duration for /bl/-clusters was 9 ms while that for /bəl/ was 7 ms. In other words, there was no difference between the two types of stimuli with respect to the duration of VOT. Thus, an unstressed schwa might be the only good acoustic cue available for listeners. Hence, distinguishing the stimulus such as "blani" and "belani" might be extremely difficult. Why, then, was the perception of the /glani-gelani/ stimuli better than that of the /blani-belani/ stimuli? It could be the case that the durational difference of VOT was slightly larger for the /glani-gelani/ stimuli (the durational difference = 10 ms) than the /blani-belani/ stimuli (the durational difference = 2 ms), and thus might have helped listeners discriminate CC from CəC.

4.2 Effects of devoiced vowel?

Contrary to the predicted order of difficulty, the performance accuracy for /sp/-clusters was the highest for Japanese participants, and /sk/-clusters were not significantly more difficult than the less marked stop + liquid clusters. One possible explanation is that the syllable structure of /s/-clusters is different from other CC clusters. As mentioned earlier, the status of /s/-clusters has been controversial among linguists. Some say that /s/-clusters have the same branching onset as other complex onsets (e.g., Boyd, 2006; Cardoso, 2008) while others claim that /s/ does not belong to onset but is either directly connected to the syllable node (e.g., Kenstowicz, 1994) or is left unsyllabified (e.g., Steriade, 1988). If the latter analysis is true, /s/-clusters might be analyzed as intermediate between regular complex onsets and simple onsets. Does this mean that the perception of /s/-clusters is less constrained than regular complex onsets since linguistic structure of /s/-clusters is closer to simple onsets?

Another possible account might be the effects of listeners' first language phonology. Vowel devoicing in Japanese phonology is a well-known phenomenon. In general, Japanese high vowels /i/ and /u/ become devoiced when surrounded by voiceless consonants regardless of speaking rate (Kondo, 2005; Nagano-Madesen, 1994; Varden, 1998). Acoustic characteristics of devoiced vowels are different from those of voiced vowels. Vowel devoicing is due to a lack of vocal fold vibration during the production of vowels that are surrounded by voiceless consonants (Kondo, 2005). Speech waveforms of devoiced vowels do not show any periodicity (Tsuchida, 1997). In addition, Kondo (2005) has demonstrated that devoiced vowels are not only shorter but also less intense than regular vowels.

If Japanese speakers repair phonotactically illegal consonant sequences by inserting a vowel in perception as Dupoux et al. (1999) claim, do they insert "a devoiced vowel" between voiceless stops? Suppose that Japanese hear an illusory devoiced [u] when presented with /sp/. Since /ə/ between these voiceless stops was not a devoiced vowel, it might be the case that these vowels somehow stood out for Japanese listeners. Thus, distinguishing between "s*u*pani" with an illusory devoiced vowel and "s*e*pani" with voiced schwa might have been easier. With respect to Japanese participants' overall performance for the /skani-sekani/ stimuli, the lower performance rate might have been due to large individual differences. In fact, the standard deviation was higher for the /sk/-conditions (SD = 16.9) than for the /sp/-conditions (SD = 12.5).

4.3 Limitations of the present study

The present study has demonstrated that there are significant differences in Japanese L2 learners' ability to distinguish complex onsets. However, such difficulty did not coincide with the sonority-based markedness hypothesis. Other factors such as the devoiced vowel phenomena and acoustic characteristics of the stimuli might have played a role.

One of the limitations of the present perception study is that we are unsure about what these Japanese participants hear when they are presented with consonantal sequences. Do they epenthesize a vowel? If so, do degree and/or quality of "an illusory vowel" differ depending on consonantal environments? For instance, do Japanese hear an illusory devoiced vowel when presented with voiceless consonantal sequences? These questions remain unanswered under the current paradigm. In an ongoing study, we are examining how consonantal sequences are produced after the presentation of such auditory stimuli in a delayed imitation production task.

References

Altenberg, Evelyn. 2005. The judgment, perception, and production of consonant clusters in a second language. *IRAL* 43:53–80.

Berent, Iris, Donca Steriade, Tracy Lennertz, and Vered Vaknin. 2007. What we know about what we have never heard: Evidence from perceptual illusions. *Cognition* 104:591–630.

Boersma, Paul and David Weenick. 2005. Praat: doing phonetics by computer (Version 4.3.14). Retrieved April 01, 2008 from: http://www.praat.org.

Boyd, Jeremy. 2006. On the representational status of /s/-clusters. *San Diego Linguistic Papers* 2:39–84.

Bradlow, Ann. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ - /l/ contrast. In *Phonology and second language acquisition*, ed. J. G. Hansen Edwards and M. L. Zampini, 287–308. Philadelphia: John Benjamins.

Broselow, Ellen. 1995. Skeletel positions and moras. In *The Handbook of Phonological Theory*, ed. J. Goldsmith, 175–205. Cambridge: Blackwell.

Broselow, Ellen, Su-I Chen, and Chilin Wang. 1998. The Emergence of the unmarked in Second Language Phonology. *Studies in Second Language Acquisition* 20:261–280.

Broselow, Ellen and Daniel Finer. 1991. Parameter setting in second language phonology and syntax. *Second Language Research* 7:35–59.

Broselow, Ellen and H.B. Park. 1995. Mora conservation in second language prosody. In *Phonological acquisition and phonological theory*, ed. J. Archibald, 151–168. Hillsdale, NJ: Erlbaum.

Cardoso, Walcir. 2008. The development of sC onset clusters in interlanguage: markedness vs. frequency effects. In *Proceedings of the 9th Generative Approaches to Second Language Acquisition Conference (GASLA 2007)*, ed. Roumyana Slabakova et al., 15–29. Somerville, MA: Cascadilla Proceedings Project.

Carlisle, Robert. 1997. The modification of onsets in a markedness relationship: Testing the Interlanguage Structural Conformity Hypothesis. *Language Learning* 47:327–361.

Carlisle, Robert. 1998. The acquisition of onsets in a markedness relationship. *Studies in Second Language Acquisition* 20:245–260.

Carlisle, Robert. 2006. The role of sonority cycle in core syllabification. In *English with a Latin Beat – Studies in Portuguese/Spanish English Interphonology*, ed. B. Baptista and M. Watkins, 105–138. Amsterdam: John Benjamins.

Clements, George. 1990. The role of the sonority cycle in core syllabification. In *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech*, ed. J. Kingston and M. Beckman, 283–333. Cambridge: Cambridge Press.

Davidson, Lisa. 2006. Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics* 34:104–137.

Davidson, Lisa. 2007. The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology* 24:261–286.

Dehaene-Lambertz, Ghislaine, Emmanuel Dupoux, and Ariel Gout. 2000. Electrophysiological correlates of phonological processing: a cross-linguistic study. *Journal of Cognitive Neuroscience* 12:635–647.

Dupoux, Emmanuel, Kazuhiko Kakehi, Yuki Hirose, Christophe Pallier, and Jacques Mehler. 1999. Epenthetic vowels in Japanese: a perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance* 26:1568–1578.

Dupoux, Emmanuel, Christophe Pallier, Kazuhiko Kakehi, and Jacques Mehler. 2001. New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes* 16:491–505.

Eckman, Fred. 1977. Markedness and the contrastive analysis hypothesis. *Language Learning* 27:315–330.

Eckman, Fred. 1981. On predicting phonological difficulty in second language acquisition. *Studies in Second Language Acquisition* 4:18–30.

Eckman, Fred. 1991. The structural conformity hypothesis and the acquisition of consonant clusters in the interlanguage of ESL learners. *Studies in Second Language Acquisition* 13:23–41.

Eckman, Fred and Gregory Iverson. 1993. Sonority and markedness among onset clusters in the interlanguage of ESL learners. *Second Language Research* 9:234–252.

Eckman, Fred. 1996. A functional-typological approach to second language acquisition theory. In *Handbook of second language acquisition*, ed. W. C. Ritchie and T. K. Bhatia, 195–211. N.Y.: Academic Press.

Eckman, Fred. 2004. From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second language Acquisition,* 26:513–549.

Eckman, Fred, Abudullah Elreyes and Gregory Iverson. 2003. Some principles of second language phonology. *Second Language Research* 19:169–208.

Eckman, Fred. 2008. Typological markedness and second language phonology. In *Phonology and second language acquisition*, ed. J. G. Hansen Edwards and M. L. Zampini, 95–115. Philadelphia: John Benjamins.

Fais, Lureal, Sachiyo Kajikawa, Janet Werker, and Shigeaki Amano. 2005. Japanese listeners' perception of phonotactic violations. *Language and Speech* 48:185–201.

Gierut, Judith and Annette Hust Champion. 2001. Syllable onsets II: Three-element clusters in phonological treatment. *Journal of Speech, Language, and Hearing Research* 44:886–904.

Hallé, Pierre, Juan Segui, Uli Frauenfelder, and Christine Meunier. 1998. Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance* 24:592–608.

Hancin-Bhatt, Barbara. 2000. Optimality in second language phonology: codas in Thai ESL. *Second Language Research* 16:201–232.

Hancin-Bhatt, Barbara and Rakesh Bhatt. 1997. Optimal L2 syllables: Interactions of transfer and developmental effects. *Studies in Second Language Acquisition* 19:331–378

Hogg, Richard and Cris McCully. 1987. *Metrical phonology: A Coursebook*. Cambridge: Cambridge University Press.

Ito, Junko and Armin Mester. 1995. Japanese phonology. In *The Handbook of Phonological Theory*, ed. J. Goldsmith, 817–838. Cambridge: Blackwell.

Kabak, Baris and William Idsardi. 2007. Perceptual distortions in the adaptation of English consonant clusters: syllable structure or consonantal contact constraints? *Language and Speech* 50:23-52.

Kenstowicz, Michael. 1994. *Phonology in Generative Grammar*. Cambridge, MA: Blackwell.

Kondo, Mariko. 2005. Syllable structure and its acoustic effects on vowels in devoicing environments. In *Voicing in Japanese,* ed. J. van de Weijer, K. Nanjo, and T. Nishihara, 229–245. Berlin: Mouton de Gruyter.

Major, Roy. 1996. Markedness in second language acquisition of consonant clusters. In *Variation and Second Language Acquisition*, ed. R. Bayley and D. Preston, 75–96. Amsterdam: John Benjamins.

Nagano-Madesen, Yasuko. 1994. Vowel devoicing in Japanese. *Lund University Working Papers* 42:119–129.

Smit, Ann. 1993. Phonological error distributions in the Iowa-Nebraska Articulation Norms Project: Word-initial consonant clusters. *Journal of Speech, Language, and Hearing Research* 36:931–947.

Sperbeck, Mieko and Winifred Strange. 2008. An epenthetic vowel between consonantal sequences in perception and production by Japanese. *The Journal of the Acoustical Society of America* 124(4):2595.

Steriade, Donca. 1988. Reduplication and syllable transfer in Sanskrit and elsewhere. *Phonology* 5:73–155.

Strange, Winifred. 1995. Cross-language studies of speech perception: A historical review. In *Speech perception and linguistic experience: Issues in cross-language research,* ed. W. Strange, 3–45. Baltimore: York Press.
Strange, Winifred and Valerie Shafer. 2008. Speech perception in second language learners: the re-education of selective perception. In *Phonology and second language acquisition*, ed. J. G. Hansen Edwards and M. L. Zampini, 153–191. Philadelphia: John Benjamins.
Tagliaferri, Bruno. 2008. Paradigm (Version Beta 5d). Retrieved April 01, 2008 from: http://perceptionresearchsystems.com
Tsuchida, Ayako. 2001. Japanese vowel devoicing: Cases of consecutive devoicing environments. *Journal of East Asian Linguistics* 10:225–245.
Vance, Timothy. 1987. *An introduction to Japanese Phonology*. New York: State University of New York Press.
Varden, John Kevin. 1999. On high vowel devoicing in standard Japanese: implications for current phonological theory. Doctoral Dissertation. University of Washington.

Department of Linguistics
The Graduate Center of the City University of New York
365 Fifth Avenue
New York, NY 10016
msperbeck@gc.cuny.edu
strangepin@aol.com