IDENTIFICATION OF ENGLISH PRIMARY STRESS AND BIAS TOWARD STRONG WORD-INITIAL SYLLABLES: NATIVE VS. JAPANESE LISTENERS

Mariko Sugahara

Doshisha University, Japan
msugahar@mail.doshisha.ac.jp

ABSTRACT

Identification of English primary stress in trochaic and iambic words was compared between English native listeners and Japanese listeners in an environment where the two stress patterns were not distinguished by pitch while other acoustic contrasts were kept intact. Native listeners and Japanese listeners showed opposite response patterns when the pitch pattern of the stimuli was made into flat: native listeners preferred more word-initial primary stress while Japanese listeners preferred more word-final primary stress. English listeners' preference for word-initial stress can be accounted for by a language-specific factor, i.e., the most probable stress pattern in their native language vocabulary. On the other hand, Japanese listeners' result can be explained that they are so sensitive to F0 information that they even interpret the flat contour as carrying an informative cue for prominence difference between the two syllables by assuming the presence of "declination".

Keywords: primary stress, F0, English, Japanese

1. INTRODUCTION

English listeners are known to segment speech at the onset of strong syllables (Cutler & Norris [6]; Cutler & Butterfield [4]), and are biased toward word-initial stress patterns in forced-identification tasks (Leyden & Heuven [7]; Cooper, et al. [3]). According to Cutler & Carter, [5]; Leyden & Heuven, [7], it is because the majority (about 90%) of English lexical words begin with strong syllables, i.e., English native listeners' perception of stress and word boundaries is influenced by their knowledge that strong word-initial syllables are the most predominant in the English vocabulary. It has not yet been proved, however, whether listeners' preference for strong initial syllables is truly specific to languages with predominant word-initial stress patterns or more universal to listeners of languages with different prosodic patterns.

One indication that listeners' preference for strong word-initial syllables may be more universal comes from Sekiguchi's [9] observation. He reports that Japanese university students were faster in detecting SW (trochaic) words than WS (iambic) words in English. Our question, then, is whether Japanese listeners are also biased toward strong initial syllables when forced to identify the location of English primary stress.

Since Japanese listeners are known to heavily rely on F0 to identify the location of English stress (Beckman [2]), their perception of English primary stress should be tested in an environment where F0 patterns are ambiguous between stimuli of different stress patterns. Otherwise, their perception will be strongly influenced by F0, and their preference for strong initial syllables, if there is any, will not surface. Therefore, we examined the question above in an environment where F0 patterns were ambiguous between trochaic and iambic stimuli.

2. FORCED CHOICE IDENTIFICATION EXPERIMENT

2.1. Participants

Thirty-eight Japanese and twelve American undergraduate students were recruited as paid participants. The Japanese listeners were mostly English majors and all of them had received at least seven years of formal English education prior to the experiment. The American listeners were all native speakers of American English studying Japanese at one-year exchange program in a Japanese university. None of them had reported speech and hearing disorders.

2.2. Stimuli

Six noun-verb pairs in which primary stress and secondary stress alternate were adopted, e.g., *TRANSplant* (noun, trochaic) vs. *transPLANT*

(verb, iambic). Since both primary and secondary stress vowels are "stressed" with full vowel qualities, the trochaic nouns and their iambic verb counterparts share the same vowel qualities regardless of their stress pattern differences. The thirteen noun-verb pairs are listed in Table 1.

Table 1: The list of stimuli used in the experiment. Syllables carrying primary stress are capitalized.

Nouns	Verbs
IMpact	imPACT
IMport	imPORT
REmake	reMAKE
REtake	reTAKE
TRANSplant	transPLANT
TRANSport	transPORT

A female native speaker of American English read those words in a context where they were interpreted as new (presentational focus). Her speech was, then, directly recorded onto a hard disk (44.1 KHz, 16 bit), using an AKG C420 Cardioid Headset Condenser Microphone. The wave forms corresponding to the target words were segmented from the entire carrier sentences, and used as stimuli in the auditory identification task.

In Figs.1-4, we summarize F0, overall intensity, duration and formant values of the first and those of the second vowels/syllables.

Figure 1: The mean of F0 averaged over an initial syllable (v1) period and that averaged over a final syllable (v2) period.

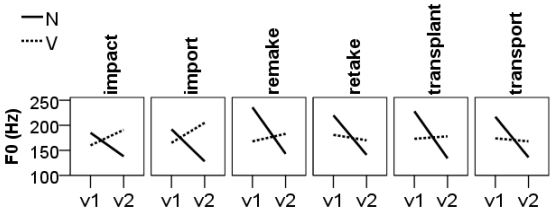
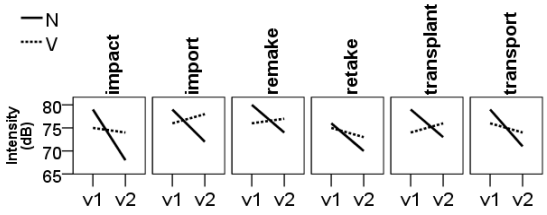


Figure 2: The mean of overall intensity averaged over a v1 period and that of a v2 period.



On the whole, pitch (Fig.1) and intensity (Fig.2) consistently distinguished trochaic N and iambic V. The duration of v2 was also consistently longer for V than for N across all word pairs (Fig.3a). The total syllable durations of s1 and s2 did not necessarily clearly differ between N and V (Fig.3b). F1 and F2 frequency differences between N and V was subtle even if there is any (Figs.4a, b).

Figure 3: The duration of **(a)** v1 and v2 and **(b)** the first syllable (s1) and the second syllable (s2).

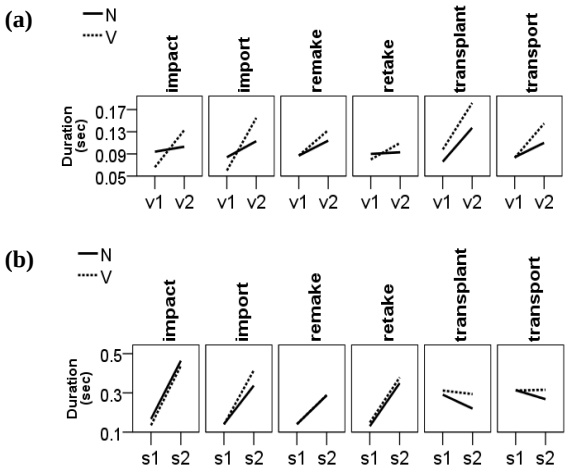
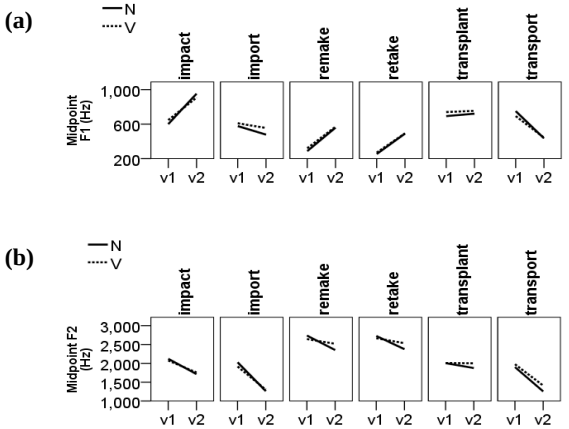


Figure 4: (a) The midpoint F1 values of v1 and v2. **(b)** The midpoint F2 values of v1 and v2.

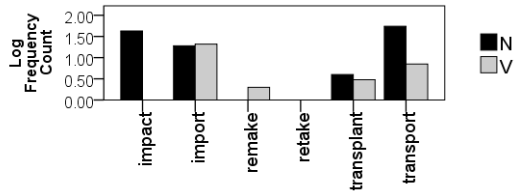


We further obtained the frequency information of the words from CELEX database (Baayen, et al. [1]), and their logarithmic values are shown in Fig. 5. The frequency of N is greater than that of V in the pairs of *impact* and *transport* while the rest of the pairs has similar frequency count for both N and V.

We created three different pitch patterns for each word: "Natural", "Flat" and "Declining". No pitch manipulation was done in the Natural condition. In the Flat condition, the F0 value of all syllables was set at 175 Hz, which is approximately the average of v1 and v2 of all words. In the Declining condition, F0 was manipulated so that it declined from the beginning of v1 to the end of v2. The F0 value at the onset point of v1 was set at 186 Hz (approximately the

average of all v1 over the two lexical categories) and that the end point of v2 was set at 161 Hz (approximately the average of v2 across categories). Other acoustic variables, such as duration, overall intensity and vowel quality were kept intact in all stimuli.

Figure 5: The frequency count (logarithmic scale) of the words used as stimuli.



2.3. Perception experiment procedure

We used E-Prime for the perception experiment. There were three blocks, in which only the stimuli of the same pitch pattern were grouped together. We did so in order to make listeners aware that pitch is not an acoustic factor to disambiguate N (trochaic) and V (iambic) when they listen to Flat and Declining stimuli. The three blocks were repeated three times. Each stimulus was presented to listeners twice in each block and their presentation was randomly ordered.

The stimuli were presented to listeners via headphones connected to a computer and they were asked to judge where the location of primary stress was. They pressed "1" on the key board of the computer if they judged primary stress to be at an initial syllable, and pressed "2" if they judged it to be at a final syllable.

2.4. Results

Fig. 6 summarizes correct response rates of N and those of V in each pitch condition and for each listener group, and Fig. 7 shows the results separately for each of the word pairs.

Figure 6: Correct stress response rates by categories, pitch conditions and listeners.

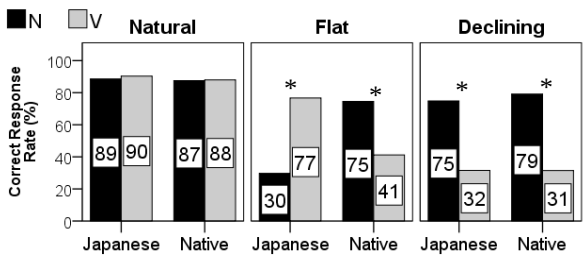


Fig. 6 shows that in the Natural condition, both N and V scored approximately 90% regardless of the listener group differences. In the Flat condition, the two groups showed opposite response patterns. Native listeners scored higher for N (75%) than for V (41%) while Japanese scored lower for N (30%) than for V (77%). The difference between N and V was significantly significant for both groups (Native:$\chi^2(1)=49.2$, $p<.001$; Japanese: $\chi^2(1)=303$, $p<.001$). In Declining, the two groups of listeners equally scored higher for N: the correct response rate of N (approx. 80%) was higher than that of V (approx. 30%), and the difference between N and V was significant (Native: $\chi^2(1)=99.4$, $p<.001$; Japanese: $\chi^2(1)=257.3$, $p<.001$).

Figure 7: Correct response rates plotted separately for each of the word pairs.

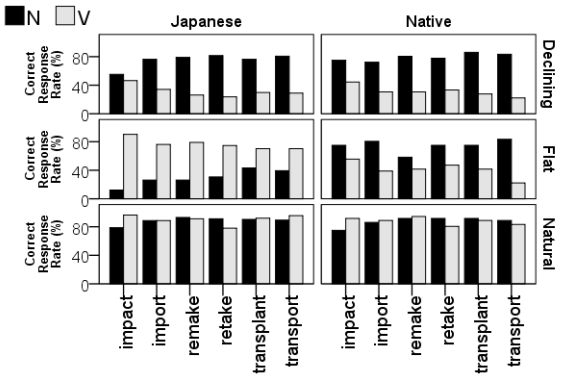


Fig. 7 shows that there were some differences among word pairs in their results. For both native and Japanese listeners, their correct response rates of *impact*(V) tended to be higher than those of V forms in other pairs regardless of the pitch pattern differences. In the same way, Japanese correct response rates of *impact*(N) tended to be lower than those of N forms in other pairs for all pitch patterns. That is, the stimuli of the *impact* pair were more likely to be perceived as having an iambic stress pattern than other pairs. Another finding is that the stimuli in the pairs of *transplant* and *transport* tended to be perceived as more trochaic than those in other pairs by the Japanese listeners in the Flat condition: their N forms had higher correct response rates than those in other pairs, and their V forms had lower correct rates than those in other pairs. A similar tendency was observed in the native speakers' response pattern of the *transport* pair in the Flat condition.

3. DISCUSSION

The main finding in the experiment is that Japanese listeners preferred final primary stress to initial stress when the trochaic N and the iambic V stimuli shared the same Flat pitch pattern. This contrasts with the result of English native listeners who chose more initial primary stress there. The outcome here supports the hypothesis that English speakers' preference for strong initial syllables is due to a language-specific factor, i.e., it is influenced by the most probable stress pattern in their native language vocabulary. That Japanese listeners preferred more word-final primary stress may be because they perceived the final syllable in the Flat stimuli "higher in pitch" or more "prominent" than the initial syllable in the same stimuli due to the effect of "declination": Pierrehumbert [8] among others reports that in order for two pitch peaks in speech to be perceived as equally prominent, the second one should be lower in pitch. As Japanese listeners are sensitive to F0 information, they would even interpret the flat contour as carrying an informative cue for prominence difference between the two syllables.

In the Declining pitch condition, not only English listeners but also Japanese listeners showed strong preference for initial primary stress. The best account for the outcome is that the listeners of both groups paid attention to the F0 difference between the initial and the final syllable in the same stimuli.

Another finding is that the stimuli of the *impact* pair were more likely to be interpreted as iambic by both native and Japanese listeners than the stimuli of other pairs. This may indicate that both native and Japanese listeners did not completely ignore acoustic information other than F0. As shown in Fig.3b, the s2 (second syllable) duration of the *impact* pair is the longest for both N and V among all pairs while their s1 (first syllable) duration is within the shortest range among all pairs. This makes the relative duration of s1 short and that of s2 long in both N and V stimuli, which might have triggered more iambic perception in this pair than in other pairs. What is interesting, however, is that the frequency count difference between *impact*(N) and *impact*(V) did not affect listeners' judgment. As shown in Fig.5, the trochaic N form of *impact* is used more frequently than the iambic V, which would be expected to affect the result in a reversal direction. That the *trans*- pairs triggered more trochaic judgment can be also accounted for by the duration of s1 and s2: s1's are intrinsically longer and s2's are relatively shorter in those pairs than in other pairs.

We started our research by asking whether Japanese listeners were also biased toward strong initial syllables when forced to identify the location of English primary stress. We, however, did not find the effect in the stress identification tasks of Flat pitch counter stimuli. What they rely on, instead, is likely an F0 cue even when an F0 contour is "flat".

4. REFERENCES

[1] Baayen, R.H., Piepenbrock, R., van Rijn, H. 1993. *The CELEX Lexical Data Base.* [CD-ROM.] Philadelphia: Linguistic Data Consortium, University of Pennsylvania.

[2] Beckman, M. E. 1986. *Stress and Non-Stress Accent*. Dordrech: Foris Publications.

[3] 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.

[4] Cutler, A., Butterfield, S. 1992. Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language* 31, 218-236.

[5] Cutler, A., Carter, D.M. 1987. The predominance of strong initial syllables in the English vocabulary. *Computer Speech & Language* 2, 133-142.

[6] Cutler, A., Norris, D. 1988. The role of strong syllables for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113-121.

[7] Leyden, K. van., Heuven, V. J. van. 1996. Lexical stress and spoken word recognition: Dutch versus English. In Cremers, C., den Dikken, M. (eds.), *Linguistics in the Netherlands*. Amsterdam: John Benjamins, 159-170.

[8] Pierrehumbert, J. 1979. The perception of fundamental frequency declination. *J. Acoust. Soc. Am.* 66, 363-369.

[9] Sekiguchi, T. 2008. Nihongowasha ni yoru eigo no bunsetsuka ni okeru sutoresutegakari no riyoo. (The use of stress-pattern cue in segmentation of English words by Japanese speakers.) *Tokyo Gakugei Daigaku Kiyoo: Soogoo-kyooikugaku-kei* (*The Bulletin of Tokyo Gakugei University: Comprehensive Educational Science*) 59, 163-169.