

PERCEPTION OF ENGLISH LEXICAL STRESS: EFFECT OF F0 PEAK LOCATION ON ENGLISH AND JAPANESE SPEAKERS

Shinichi Tokuma

Chuo University, Tokyo, Japan
tokuma@tamacc.chuo-u.ac.jp

ABSTRACT

This study investigated the perceptual effect of duration and F0 peak location on L1 / L2 perception of English lexical stress. A non-sense bisyllabic English word embedded in a frame sentence whose F0 was set to reach its peak after the word was used as the stimuli of the perceptual experiment. Native English and Japanese speakers were asked to determine lexical stress locations in the experiment. The results showed that in the perception of English lexical stress, F0 peaks that immediately followed the stimulus words perceptually affected the subjects in an opposite manner: Japanese speakers perceived these F0 peaks as a cue to lexical stress in the preceding syllable, while English speakers perceived them as an independent prominence peak and showed perceptual stress shift. The findings also confirmed the claim by previous studies that, while the perception of Japanese subjects is scarcely affected by duration, English subjects show great sensitivity to it.

1. INTRODUCTION

English and Japanese, as well as other languages, have peak-delay, where the F0 peak of a pitch accent is aligned outside the stressed or accented syllable [1] [2]. Particularly, in Japanese, which uses high / low pitch differences to distinguish words, a delayed F0 peak can determine a lexical accent position, in a way that an F0 peak that occurs *after* the mora is associated with the lexical accent location. It is a common phenomenon and called 'oso-sagari' [2].

F0 is a main acoustic correlate of lexical stress / accent in English and Japanese. Of four acoustic parameters contributing to the perception of English lexical stress (F0, segmental duration, intensity and vowel quality), F0 and duration are the most important cues (See, for example, the study review in [3]). In the lexical accent perception of Japanese

disyllabic words, Japanese speakers heavily rely on F0 [4].

This raises the following question of L2 perception: if F0 reaches its peak after a word in an English sentence, how will that affect the lexical accent perception of the word by native Japanese speakers? Will the Japanese be significantly influenced by their L1, where a delayed F0 peak affects the lexical accent perception of the preceding word?

Little has been found on the effect of the delayed F0 peak on L2 perception, although the perception of F0 peak-delay in L1 has been studied. For instance, Hasegawa & Hata [5] conducted an extensive study on the monolingual perception of F0 peak-delay in Japanese and English. With regard to English, they studied the perception of sentence accent, using the contrast of the sentential focus: 'This is MY net.' vs. 'This is my NET.' (The capitalised word means the nucleus assignment). One of the few cross-linguistic papers on the perception of F0 peak-delay is Sugito [2]. She conducted a cross-linguistic perceptual experiment on peak-delay, but her subjects were four native speakers of English, who listened to F0 peak-delay.

Another interesting issue is the effect of duration on the perception of English stress by the Japanese. It has been investigated: Mochizuki-Sudo & Kiritani [6], who examined the sensitivity to duration of stressed vowels and inter-stress intervals, claim that Japanese listeners with poor English command do not perceive inter-stress intervals. Or, Beckman [4] showed that in the perception of English lexical stress, native Japanese speakers are sensitive almost exclusively to F0, while the perception of English speakers are influenced by F0, duration and, to a slightly lesser extent, by amplitude. However these studies used existent English word pairs with a lexical stress contrast, such as 'object' (noun) - 'object' (verb). Hence their stimuli are not symmetrical in their syllable and segmental structure, and this could possibly affect

subjects' performance.

This paper investigates how the English lexical stress perception by Japanese speakers is affected by the shifted F0 peak. Perceptual data of native speakers of English was also collected for comparison. Furthermore this paper also attempts to confirm the insensitivity of Japanese speakers to duration, found in previous studies [4] [6], using nonsense stimulus words with symmetrical syllable structure.

2. EXPERIMENT

2.1. Subjects

Two groups of subjects participated in the perceptual experiment.

- (A) Native speakers of English (henceforth called EN): Five native speakers of British English, four of whom are university lecturers in Japan, the rest being a graduate of University of London. None of them speaks with a noticeable regional accent.
- (B) Native speakers of Japanese (henceforth called JP): Fifteen Japanese undergraduate first-year students of Chuo University in Tokyo. All of them had taken TOEIC tests, and the average score was 576 points. No subject had lived in an English-speaking country. They all took an English course by the author, who can safely assert that their English abilities are at the pre-intermediate level.

2.2. Materials

A nonsense word 'nurnur' /nə:nə:/ was used for the experiment. The choice of /n/ facilitated the manipulation of duration and the vowel /ə:/ is assumed to be the closest in vowel quality to an English weak vowel /ə/.

The duration of the syllable /nə:/ in the word was varied in 6 steps of 20ms from 160ms to 260ms, but the duration step of the first syllable ranged from 160ms to 240ms, while the duration step of the second syllable was varied from 180ms to 260ms. This asymmetry is based on previous studies [7] [8] which showed native English listeners' strong preference to stress on the first syllable, called 'stress bias'. The total duration of the word was kept constant. This process produced the stimuli of 5 durational types shown in Table 1 below. These stimulus words /nə:nə:/ were

embedded in a frame sentence: 'Will you put down in the yard?' to avoid the perceptual intervention of sentence-final lengthening. The phrase 'down in the yard' includes only vowels and voiced consonants to produce a continuous F0 contour.

Table 1: *Durational structure of stimuli*

	1st syllable duration	2nd syllable duration
Stimulus 1	160ms	260ms
Stimulus 2	180ms	240ms
Stimulus 3	200ms	220ms
Stimulus 4	220ms	200ms
Stimulus 5	240ms	180ms

In the synthesis, MBROLA synthesiser [9] was used and its British male voice database was implemented. As a reference, a recording of the sentence was separately made, using a male native speaker of South-East British English, and its segmental duration was used for the synthesis, while F0 synthesis involved further manipulation.

In the frame sentence, F0 was set to reach 130Hz at the 90% durational point of /ɪ/ in 'will', linearly up to 160 Hz in the middle part of 'put'. Then it linearly declined to 150 Hz by the end of /ə:/ in 'nurnur'. Finally, it was set to reach a high peak of 180 Hz, followed by a steep fall, in the continuum of 'down in'. This gives the frame sentences an auditory impression of having a high-fall nucleus after the target stimulus word 'nurnur'.

The detailed manipulation of the F0 peak pattern was as follows: Hasegawa & Hata [5] reported that in the perception of English sentential accent, the fall rate of 1.6 Hz/ms invoked 70% peak-shift responses (i.e. accent sensation shifted to the front, from 'my NET' to 'MY net') among English speakers. Hence in this study, the rate of 1.6 Hz/ms was adopted, and as in [5], the frequency range of the fall was kept constant: 180 Hz to 100 Hz. The F0 peak of the nucleus was shifted in 50ms increments by 5 steps, after the nonsense word /nə:nə:/, within the continuum of 'down in'. The shift increment was set to start 20ms after the release of /d/ to avoid a truncated peak in the F0 contour. This set-up produced 5 F0 peak patterns located at 20ms / 70ms / 120ms / 170ms / 220ms after the release of /d/ in 'down in'.

The synthesised sentences were checked by one native speaker of South-East British English, who ensured that they had an acceptable quality of synthesised speech.

2.3. Experimental procedure

Each stimulus sentence was presented to the subjects four times, producing a total of 100 presentations (5 durational patterns x 5 F0 peak patterns x 4 repetitions) per subject, and they were preceded by ten trial presentations designed to make the subjects familiar with the experimental setting and the nature of the stimuli. The interval between each presentation was 3 seconds, and a longer pause of 5 seconds with a beep was inserted after every 10 presentations.

The task of the subjects was to listen to the stimulus words embedded in a sentence and to judge which of the syllables in the stimulus word /nə:nə:/ was stressed. They were asked to circle or tick the syllable of the word (written as 'Nur-Nur' in an answer sheet) which they thought was stressed. The test was carried out in a quiet study room for EN and in a Language Laboratory room for JP subjects. All the subjects listened to the stimuli through covered-ear headphones and none of them reported that their attention had been diverted by noise, or by the presence of other subjects in the case of the JP group.

2.4. Results

After the experiment, it was found that one subject of the JP group had chosen the first syllables for nearly all the presentations, while three other subjects had left many answers blank. These four subjects, therefore, were excluded from the analysis in order to validate the reliability of the data. This reduced the JP group to eleven members.

In the analysis, the responses were accumulated and the numbers of the first or second syllable choices were counted for each F0 peak and duration type across all the subjects within the group, before the percentages of the first / second syllable choices were calculated. Figures 1 and 2 show the percentages of the second syllable choices for each F0 peak and durational pattern and for each subject group. Figure 1 is for the JP group and Figure 2 for the EN group. In these figures, F0 peak locations are plotted on the X axis. On the other hand, durational patterns of the second syllables are plotted as a separate category.

Figure 1 suggests that the JP group were sensitive to the F0 peak location, which is shown in the falling trend from left to right, regardless of the durational manipulation of the stimuli. This is confirmed by Figure 3, where the mean percentages

of the first / second syllable choices were obtained across all duration patterns for each F0 peak location. This means that, when determining the stress position of the stimulus word, Japanese subjects give preference to stress on the second syllable, if the following F0 peak is close enough.

Figure 1: Results of JP group; 'S' stands for Stimulus (See Table 1).

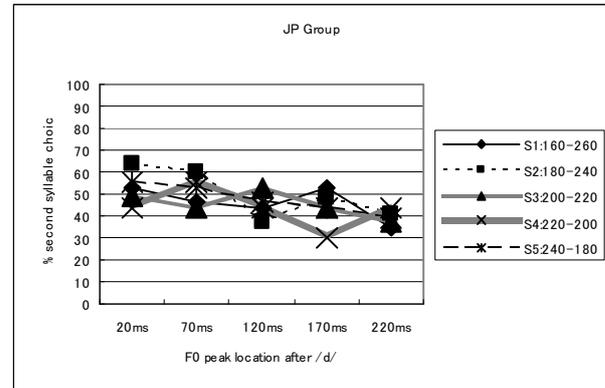


Figure 2: Results of EN group; 'S' stands for Stimulus (See Table 1).

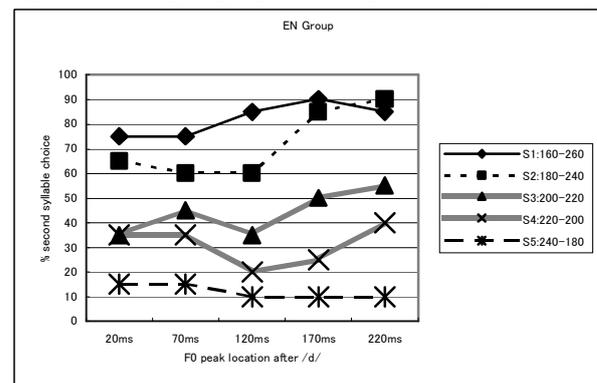


Figure 3: Mean responses of JP group across all duration patterns.

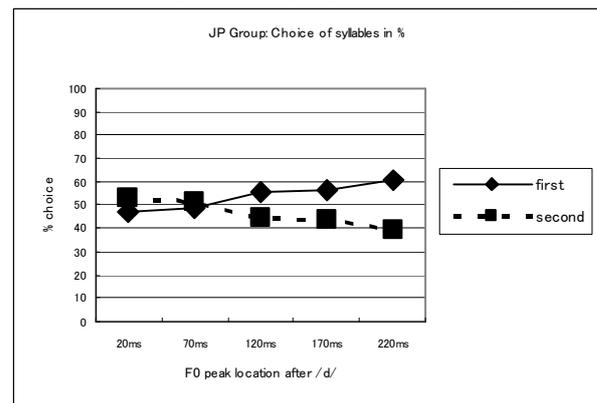


Figure 2 shows that for S4 and S5, the F0 peak location does not influence the lexical stress perception of the native speakers, since no positive or negative relation can be observed between the distance from the F0 peak location and the number of second-syllable choices; in contrast, the percentage of the second syllable choice is lower in

S1, S2 and S3, when the F0 peak is closer to the syllable (as shown in 20ms / 70ms / 120ms stimuli). This phenomenon is discussed in the next section.

Figures 1 and 2 also demonstrate the contrast in the sensitivity to the durational change: in Figure 1, for the JP group, the influence of the durational change on perception cannot be discerned but Figure 2 shows the increase of the second syllable choice from S5 (240ms, 180ms) to S1 (160ms, 260ms), which means that as the duration of a syllable increases, more subjects of the EN group perceive it as stressed.

3. DISCUSSION

The experiment undertaken in this study supports the claim by Beckman [4] and Mochizuki-Sudo & Kiritani [6] that Japanese subjects with poor English command did not perceive duration as a cue to English lexical stress. This is demonstrated in this study by the JP group's insensitivity to duration.

The results of the experiment also show that, in the perception of English lexical stress, Japanese listeners are influenced by the location of the delayed F0 peak. This is in line with the findings of Hasegawa & Hata [5] on monolingual F0 peak-delay perception.

There is one interesting contrast between the results of the EN and JP group. As noted in 2.4, Figure 2 (the results of the EN group) shows a lower percentage of the second syllable choice in S1 to S3 when the F0 peak is closer to the syllable (as shown in 20ms / 70ms stimuli), while, in Figure 3, Japanese subjects made a higher percentage of the second syllable choice for 20ms / 70ms / 120ms stimuli of S1 and S2.

This is perhaps due to whether the subjects of this study perceptually integrated the F0 peak that immediately followed the second syllable of the stimuli. Since F0 is the sole cue to lexical accent perception and peak-delay is a common phenomenon in Japanese, JP subjects perceptually integrated the F0 peak in the following syllable and perceived prominence in the second syllable. In contrast, the F0 peak that immediately follows the second syllable of the stimuli may have sounded like a separate prominence peak to EN subjects, and this resulted in stress clash sensation to their ears. Hence they shifted their perceptual stress location to the first syllable to keep a regular spacing of stress assignments in the sound structure,

as observed in the stress shift phenomenon.

This perceptual behaviour of EN subjects has some proof: Figure 2 shows that this perceptual stress shift does not occur in 170ms / 220ms stimuli of S1 and S2, where two prominence peaks are safely apart; and Figure 2 also shows that, in S5, where the durational difference (240ms-180ms) is distinctive enough to tell that the first syllable is stressed, the F0 peak location does not have any effect on the lexical stress perception.

This hypothesis, however, would have to be confirmed by future experiments.

4. CONCLUSION

Overall, the results of the experiment suggest that in the perception of English lexical stress, F0 peaks that immediately followed the stimulus words perceptually affected the subjects in an opposite manner: Japanese speakers perceived these F0 peaks as a cue to lexical stress in the preceding syllable, while English speakers perceived them as an independent prominence peak and showed perceptual stress shift. The findings also confirmed the claim by previous studies that, while the perception of Japanese subjects is scarcely affected by duration, English subjects show great sensitivity to it.

5. REFERENCES

- [1] Ishihara, K. (2003) 'A phonological effect on tonal alignment in Tokyo Japanese.' *Proc. ICPhS 2003, Barcelona*, 615-618.
- [2] Sugito, M. (1982) *Nihongo Akusento no Kenkyu*. (in Japanese) Tokyo: Sanseido.
- [3] O'Shaughnessy, D. (2000) *Speech Communications: Human and Machine*. New York: IEEE Press.
- [4] Beckman, M. (1986) *Stress and Non-Stress Accent*. Dordrecht: Foris Publications.
- [5] Hasegawa, Y and Hata, K. (1992) 'Fundamental frequency as an acoustic cue to accent perception.' *Language and Speech*. 35, 87-98.
- [6] Mochizuki-Sudo, M. and Kiritani, S. (1991) 'Production and perception of stress-related durational patterns in Japanese learners of English.' *J. of Phonetics*. 19, 231-248.
- [7] Tokuma, S. (2003). 'Perception of English lexical stress by English and Japanese speakers: effect of duration and "realistic" intensity change.' *Proc. Eurospeech 03, Geneva*, 2121-2124.
- [8] van Heuven, V.J. and Menert, L. (1996) 'Why stress position bias?' *J. Acoust. Soc. Amer.* 100, 2439-2451.
- [9] Dutoit, T., Pagel, V., Pierret, N., Bataille, F. and van der Vreken, O. (1996) 'The MBROLA Project: Towards a Set of High-Quality Speech Synthesizers Free of Use for Non-Commercial Purposes.' *Proc. ICSLP'96*. 3, 1393-1396.