L1 / L2 perception of lexical stress with F0 peak-delay: effect of an extra syllable added

Shinichi Tokuma¹, Yi Xu ²

¹ Department of English, Faculty of Commerce, Chuo University, Japan
² Research Department of Speech, Hearing and Phonetic Sciences, UCL, United Kingdom

Abstract

This study addressed the problems of our previous study [5] and investigated the perceptual effect of F0 peak-delay on L1 / L2 perception of English lexical stress. A trisyllabic English nonsense word ‘ninimi’ whose F0 was set to reach its peak around the second syllable was embedded in a frame sentence and used as the stimulus 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, delayed F0 peaks which were aligned with the second syllable of the stimulus words perceptually affected both Japanese and English groups, although slightly in a different manner: the Japanese group perceived the delayed F0 peaks as a cue to lexical stress in the first syllable when the peaks were aligned with, or before, the end of /n/ in the second syllable, while the English group had the boundary shifted in an earlier temporal position. It was also discovered that the Japanese group had greater sensitivity to the delayed peak positions. The difference could be attributed to the dependency upon F0 as a perceptual cue and upon its alignment position in a constant / relative timing, as well as the speech rate difference between the two languages.

Index Terms: L2 speech perception, English lexical stress, F0 peak-delay, F0 alignment position

1. Introduction

In speech, F0 peaks can be aligned after the accented syllable / mora, and this phenomenon is called F0 peak-delay. This is observed both in English, (For example, [1] and [2]) and in Japanese (See [3] and [4] among others).

We investigated the perceptual effect of F0 peak-delay on L1 / L2 perception of English lexical stress [5]. A bisyllabic English nonsense word ‘nimi’ whose F0 was set to reach its peak around the second syllable was embedded in a frame sentence “Lee may my niece” and used as the stimulus 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, delayed F0 peaks which were aligned with the second syllable of the stimulus words perceptually affected Japanese and English groups in the same manner: both groups perceived the delayed F0 peaks as a cue to lexical stress in the first syllable when the peaks were aligned with, or before, the end of /n/ in the second syllable (See Figures 1 and 2, where the horizontal axis represents the temporal positions of F0 alignment and the vertical axis the percentage of the responses). A supplementary experiment conducted on Japanese speakers confirmed the location of the categorical boundary.

However, there are potential issues that have been raised regarding the experiment in [5]. First, the nonsense word used as the stimulus ‘nimii’ /nimii/ has a homophonic real word ‘ninii’ /ninii/, although the word is used as a noun¹. This might have provided a bias towards the preference of the first syllable. Second, English has a phonotactic constraint where stressed /i/ does not occur in an open syllable, and as the previous research, such as [6] and [7], demonstrated the influence of phonotactic constraints on L1 / L2 perception, this might prohibit the perception of second-syllable stress on the word of the structure given.

The nonsense word could have a closed syllable in the end, like /nimi/, but this would create an asymmetry in the duration of /ni/ in each syllable. Another suggested solution was, instead of /i/, to use long vowels such as /a/. However, the duration of the replaced vowel would be certainly longer than that of /i/, making the second syllable too long to cause F0 peak-delay: [2] reports that in disyllabic words, “it is when the duration of the stressed syllable is shorter than 200ms that the

¹ A few native speakers of English suggested they use it as a verb, meaning “to fool”, but the informal feedback obtained during and after the presentation of [5] showed that most native speakers of British English prefer not to use it as a verb.
F0 peak occurs in the following syllable” (p.180). This means that a short vowel is required to simulate the natural F0 peak-delay environment.

Our solution to this quandary is to utilise a trisyllabic nonsense word /ninini/, to align the F0 peak with the second syllable and to ask the participants of the experiment to identify whether the stress is on the first or the second syllable. Having an extra syllable in the test word is not uncommon: in [8], the word “millionaire” was used to test whether the first or the last syllable was stressed while F0 peaks were aligned with the second syllable.

In this study, we aim to re-examine how F0 peak-delay influences the perception of English lexical stress by Japanese and English speakers by using an English nonsense word /ninini/. The word is embedded in the same frame sentence used in [5] to compare the results. Also, as in [5], the F0 increment step is set to 20ms, and the F0 peaks are aligned with the second syllable of the word.

2. Experiment

2.1. Participants

Two groups of participants were tested in the experiment.

(A) Native speakers of English (henceforth EN): Eleven native speakers of South-East British English. They were recruited through UCL Psychology Subject Pool System, and they were paid for their participation. All were aged between 17 and 24 years old, and none had a history of hearing or language impairment.

(B) Native speakers of Japanese (henceforth JP): 24 native Japanese first-year or second-year undergraduate students at Chuo University in Tokyo, aged between 18 and 20. None of them had lived in an English-speaking country or reported on hearing / language impairment. They had studied English for at least six years at school, and their English abilities were judged to be at the pre-intermediate level by their instructor. They were not paid for their participation.

2.2. Materials

As stated in Introduction, an English nonsense word ‘ninini’ /ninini/ was used for the experiment, and it was embedded in the same frame sentence as in [5]: ‘Lee may _____ my niece.’ This was to avoid the perceptual intervention of sentence-final lengthening. The choice of the frame sentence is based on the mean F0 contour pattern presented in [2], which shows a clear F0 peak delay.

Five productions of ‘Lee may ninini my niece’ were made by the female RP speaker who had also provided the material for our previous work. They were recorded in a sound-attenuated recording room at the Department of Speech, Hearing and Phonetic Sciences, University College London. The speaker was instructed to place the nucleus on ‘ninini’ and the lexical stress on the second syllable, i.e. /ninini/, and one representative production was chosen as an input for stimuli resynthesis. The identical F0 contour as in [5] was utilised for resynthesis: it started at 230 Hz in ‘Lee’, linearly up to 260 Hz in ‘ninini’, and after the word, it linearly declined to 150 Hz by the end of the sentence. In ‘ninini’, there was a local peak with the steeper slopes, where F0 was set to rise from 260 Hz to a high peak of 310 Hz, followed by a fall to 160Hz (See Figure 3 below).

The procedure of [5] was followed to make the syllables identical in duration, intensity and vowel quality: using Speech Filing System ver. 4.7 (obtainable from www.phon.ucl.ac.uk/resource/sfs/), its second stressed syllable /ni/ was extracted and duplicated. The unstressed /ni/ at the end of ‘ninini’ was also extracted. These three syllables were concatenated, with a temporal overlap of 20ms for smoothing, in the order of ‘stressed-stressed-unstressed’. This process created /ninini/ where the first two syllables were identical in duration and vowel quality.

The duration of /n/ and /i/ was set to 50ms + 50ms, based on the average values of productions of the RP speaker. Finally, the modified nonsense word was embedded in the sentence, replacing the original.

Then the peak F0 contour was manipulated by editing the stylised pitch tier using Praat ver. 5.2 (www.fon.hum.uva.nl/praat/), and the same procedures as in [5] was followed again. The peak was shifted in 20ms increments by seven steps, around the second syllable of the nonsense word ‘ninini’. The shift increment was set to start after the onset of /n/. This set-up produced seven F0 peak patterns located at: 10ms before the onset of /n/ in the second syllable; at 10ms / 30ms / 50ms after the onset of /n/ in the second syllable; at 20ms / 40ms after the onset of /i/; and 10 ms after the onset of /n/ in the third syllable. Figure 3 shows a schematic diagram of this.

2.3. Procedure

Each stimulus sentence was presented in a random order to the participants seven times, producing a total of 49 presentations (7 F0 peak patterns x 7 repetitions) per participant, and they were preceded by ten trial presentations designed to make the participants 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 participants was to listen to the stimulus words embedded in a sentence and to judge whether the first or the second syllable in the stimulus word /ninini/ was stressed. They were asked to circle or tick the syllable of the word (written as ‘NI-NI-ni’ in an answer sheet) which they thought was stressed. Extra care was taken not to let the participants choose the third syllable as a choice, by giving instructions and feedback before and after the trial session.

The test was carried out individually in a sound-attenuated recording room at University College London for EN participants, and in one group in a quiet Language Laboratory room at Chuo University for JP participants. The stimuli were
played through loudspeakers for both EN and JP groups, none of whom reported that their attention had been diverted by noise, or by the presence of other participants in the case of the JP group.

2.4. Results

After the experiment, five JP participants were excluded because of one of the following reasons: (1) they had chosen exclusively one of the syllables for nearly all the presentations; (2) they had left some of the answers blank; or (3) they had chosen the third syllable. None of EN participants was excluded. This reduced the number of the JP group to 19 members.

In the analysis, the responses were accumulated and the numbers of the first or second syllable choices were counted for each F0 peak position across all the participants within the group, before the percentages of the first / second syllable choices were calculated. Figures 4 and 5 show the percentages of the first / second syllable choices for each F0 peak and for each group. Figure 4 is for the EN group and Figure 5 for the JP group. In these Figures, F0 peak locations are plotted on the X axis.

![Figure 4: Results of EN group](image)

![Figure 5: Results of JP group](image)

Figures 4 and 5 confirm the findings of [5] that both groups perceive the delayed F0 peak as the cue to the lexical stress on the first syllable. However, the 50% boundary positions in these Figures are different: in Figure 4, it is located immediately after the 30ms point of /n/ in the second syllable, while it is between the end of /n/ and 20ms point of /n/ in the second syllable in Figure 5, although the 50% boundary positions were the same in our previous study [5] (See Figures 1 and 2 in Introduction).

This means that for the JP group, the boundary position is consistently located just after the onset of /n/ in the second syllable, whether the test word is /ninimi/ or /ninini/, while EN speakers perceive delayed F0 peak differently, according to the number of the /ni/ sequences in the test word.

Figures 4 and 5 also confirm the observation made in [5] from Figures 1 and 2 that Japanese speakers are more sensitive to F0 peak delay. In Figure 5, the perceptual pattern of the JP group shows a clear categorical curve, while the curve is less steep in Figure 4. These findings are discussed in detail in the next section.

3. Discussion

The results of Experiment show that both Japanese and English speakers are sensitive to F0 peak-delay, perceiving the peak in the /n/ of the second syllable as the cue to the lexical stress on the first syllable, but the categorical boundary is located in different positions: for Japanese speakers, it is immediately after the end of /n/ in the second syllable, which is congruent with the previous results [5], while it is located immediately after 30ms point of /n/ in the second syllable for English speakers.

One hypothesis to explain this phenomenon is that Japanese speakers have strong sensitivity to F0 alignment positions, and the end of the first consonant in the second syllable is the boundary position, regardless of the duration of syllables / moras. This hypothesis has some empirical support from previous studies. For example, [4] proposes that in an initially accented CVCV sequence of Japanese, its F0 peak is aligned “just around the beginning of the vowel of the following syllable,” (p.30) which, in this study, corresponds to the end of the second /n/. Or, [9] demonstrates by his perceptual experiments that the F0 slope and the preceding context are basically irrelevant in the perception of Japanese lexical accent, only H* alignment determines the accent location and the difference between H* and L% determines the accent presence. This suggests that Japanese speakers will strongly rely on the peak alignment position when determining the lexical stress perception of English, which is the case in this study.

On the other hand, it is hypothesised that English speakers, when they refer to alignment positions in determining lexical stress, could rely on the relative temporal alignment position. In our previous study [5], the 50% boundary of ‘ninim’ sits 60ms-70ms after offset of the first syllable (see Figure 1), and since first syllable duration is 160ms, it is 37.5%-43.8% of the syllable duration after the offset. In the current study, the 50% boundary of ‘ninimi’ is 30-50ms after the offset (see Figure 4). Since the first syllable duration is 100ms in this study, it is 30%-50% of the syllable duration after the offset, which is very close to the first set of numbers. Figure 6 is a schematic diagram.

![Figure 6: Schematic figure of perceptual boundary positions for EN groups in [5] and the present study](image)
This could be taken as evidence that the proportional alignment position, in relation to the syllable / word duration, is perceived as a cue to lexical stress perception of English speakers. The relevant acoustic data can be found in the acoustic study by [10]. In [10], they investigated the influence of sentence position and phonological vowel length on the alignment of accent-related F0 peaks in Scottish Standard English and Southern British English, and state “alignment of nuclear peaks is proportionally invariant, relative to the duration of the word, the foot, or some other such phonological domain.” (p.155) However, this must be treated with caution, since our data is about lexical perception, not acoustic study of prosodic prominence. This explanation awaits further verification.

Another hypothesis for this the categorical boundary shift is the difference in speech rate between English and Japanese, in terms of the number of syllables per second. Although speech rate per segment does not show any difference between Japanese and English speakers [11], it is suggested that speech rate per syllable is faster in Japanese than in English [12]. Hence, when a pitch accent occurs on a short vowel, peak delay can be greater in Japanese than in English. If so, a perceptual bias of Japanese speakers may have developed in perception to allow for greater peak delay into the following syllable when the accent occurs on the current syllable. This hypothesis also requires further data to verify its validity.

The results of the experiment also confirm the findings in [5] that Japanese speakers are more sensitive to F0 peak delay, as shown by the steeper categorical curve for the JP group in Figures 2 and 5. In [5], this was attributed to the role of F0 in lexical stress / accent perception in their L1: due to the difference in the role of F0 in lexical stress perception, English speakers, without durational difference, had to rely solely on F0, and this resulted in more gradual slopes in the perceptual curves. On the other hand, Japanese speakers tuned into the F0 change of the L2 stimuli, as they normally do in their L1, and responded to it better. The results of our current experiment confirm this observation, and similar superiority of L2 subjects in F0 cue sensitivity is reported in [13]. She investigated the perception of English lexical stress by English and Chinese adults, and found that Chinese adults have significantly lower duration and intensity reliance scores but significantly higher F0 scores than English adults. Since the most important acoustic correlate for tone in Chinese is pitch, she attributed this superiority to the transfer of F0 reliance from their L1 to their L2, which is considered to be also the case in this study. This superiority of L2 subjects can be extended to other acoustic cues. For example, [14] reported that when perceiving /l/-/ɹ/ distinction in English, Japanese adults are more sensitive to variation along F2 dimension, although it is mostly irrelevant to the English /l/-/ɹ/ categorisation. Further research is required to establish these findings.

4. Conclusions

The results of the Experiment demonstrated that in the perception of English lexical stress, delayed F0 peaks which were aligned with the second syllable of the trisyllabic stimulus words perceptually affected Japanese and English groups, although slightly in a different manner: the Japanese group perceived the delayed F0 peaks as a cue to lexical stress in the first syllable when the peaks were aligned with, or before, the end of /a/ in the second syllable, while the English group had the boundary shifted in an earlier temporal position. It was also discovered that the Japanese group had greater sensitivity to the delayed peak positions. These findings are supported by the data provided by previous studies on L1 acoustic analysis and on L1 / L2 perception. The difference could be attributed to the dependency upon F0 as a perceptual cue and its alignment position in a constant / relative timing, as well as the speech rate difference between the two languages.

5. Acknowledgements

The authors cordially appreciate the comments made by Michael Ashby of University College London and Isao Hara of Newcastle University. This research was partially funded by Chuo University Personal Research Grant.

6. References