

The effect of F0 peak-delay on the L1 / L2 perception of English lexical stress

Shinichi Tokuma^{1,2}, Yi Xu²

¹ Department of English, Faculty of Commerce, Chuo University, Japan

² Research Department of Speech, Hearing and Phonetic Sciences, UCL, United Kingdom

tokuma@tamacc.chuo-u.ac.jp, yi.xu@ucl.ac.uk

Abstract

This study investigated the perceptual effect of F0 peak-delay on L1 / L2 perception of English lexical stress. A bisyllabic English non-word 'nini' /nɪni/ whose F0 was set to reach its peak in 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 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. A supplementary experiment conducted on Japanese speakers confirmed the location of the categorical boundary. These findings are supported by the data provided by previous studies on L1 acoustic analysis and on L1 / L2 perception of intonation.

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

1. Introduction

A discrepancy between phonological tonal association and phonetic F0 peak alignment is commonly observed in English and Japanese. In particular, the F0 peak can be aligned after the accented syllable / mora. This phenomenon is called F0 peak-delay, and is observed both in English and Japanese. In English, the F0 peak-delay is demonstrated, for example, by the data in [1] and [2]. In Japanese, see [3] and [4] among others.

Although F0 peak-delay is found in both languages, the role of F0 in the perception of lexical stress / accent is not symmetrical: in Japanese, F0 is the sole cue to the perception of lexical accent [5], while in English, F0 is just one of the cues, on a par with duration and intensity. (For reviews, see [6].)

This raises the following question relating to L2 perception: how is the F0 peak-delay in English lexical stress perceived by Japanese speakers? Do Japanese speakers perceive it differently from English speakers, since the roles assigned to F0 in lexical stress / accent are different, or are the perceptual patterns of both groups of speakers similar, since the F0 peak-delay is observed in both languages?

This topic has been little researched, and one of the few cross-linguistic papers is [3], but her experiment was about the perception of Japanese F0 peak-delay by American adults. Our previous study [7] investigated the perceptual effect of duration and F0 peak-delay on L1 / L2 perception of English lexical stress by using a bisyllabic English non-word. The results showed that in the perception of English lexical stress, F0 peaks that immediately followed the stimulus words perceptually affected the subjects in a different manner:

Japanese speakers perceived these F0 peaks as a cue to lexical stress in the preceding syllable, while English speakers were not as sensitive to delayed F0 peaks.

However, there are three potential issues that could be raised regarding the experimental method of [7]. First, in [7], the non-word was embedded in a sentence, 'Will you put ____ back on the table?' and the F0 peaks were aligned along the 'back on' continuum. This means that the F0 alignment was beyond the word / phrase boundary, which may have affected the perception of English listeners. Second, the non-word used in [7] was 'nur-nur' /nʊ:nʊ/, following [8] and [9], which studied the effect of intensity and duration, but in [7], the duration of the second syllable, manipulated from 180ms to 240ms by 20ms step, may be too long to cause F0 peak-delay. For example, [2] reports that in disyllabic words, "it is when the duration of the stressed syllable is shorter than 200 ms that the F0 peak occurs in the following syllable" (p.180). Third, the increment of F0 alignment shift in [7] was 50ms, but this may be too large, since the previous studies on F0 alignment perception used smaller values as one step: 25ms in [10] and 20ms in [11], [12].

In this study, we aim to re-investigate how F0 peak-delay influences the perception of English lexical stress by Japanese and English speakers. An English non-word, embedded in a frame sentence, is used as a stimulus, since the existing English word pairs with a lexical stress contrast, such as 'record'(noun) - 'record'(verb), have an asymmetrical syllable and segmental structure. Also, a shorter vowel /ɪ/, instead of /ə/, is used in the non-word. The F0 increment step is 20ms, and the F0 peaks are aligned with the temporal points in the second syllable of the word.

2. Experiment 1

2.1. Participants

Two groups of participants were tested in Experiment 1.

(A) Native speakers of English (henceforth EN): Twelve native speakers of South-East British English. Two of them were work-experience students from Latymer School, London, who voluntarily participated in the experiment. The other ten speakers were recruited through UCL Psychology Subject Pool System, and they were paid for their participation. All were aged between 17 and 35 years old, and none had a history of hearing or language impairment.

(B) Native speakers of Japanese (henceforth JP): Forty-three native Japanese first-year or second-year undergraduate students at Chuo University in Tokyo, aged between 18 and 20. No subject 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

An English non-word ‘nini’ /nɪnɪ/ was used for the experiment, and it was embedded in a frame sentence: ‘Lee may _____ my niece.’ to avoid the perceptual intervention of sentence-final lengthening. The vowel /ɪ/ was chosen since its duration is shorter than /ə:/ and, in vowel quality, the unstressed /ɪ/ is assumed to be very close to its stressed counterpart¹.

The choice of the frame sentence is based on the mean F0 contour pattern presented in [2]. In [2], the word ‘mimic’, which has a similar phonological structure to our non-word ‘nini’, was put in the sentence ‘Lee may mimic my niece,’ and the mean contour indicated a strong F0 peak-delay: although lexical stress was on the first syllable of ‘mimic’, its F0 peak was in the temporal middle point of the second syllable. This sentence has another advantage: its consonants are all sonorants except the final /s/, which helps to produce a continuous F0 contour.

Three productions of ‘Lee may nini my niece’ were made by a female RP speaker. They were recorded in a sound-attenuated recording room at the Department of Speech, Hearing and Phonetic Sciences, University College London, and the speaker was instructed to place the nucleus on ‘nini’ and the lexical stress on the first syllable, i.e. /nɪnɪ/. Her F0 contours were analysed by Praat software ver. 5 (downloadable from www.praat.org), and close inspection showed that the first production had the clearest F0 peak-delay: the peak was aligned with the temporal end point of /n/ in the second syllable. Hence, the first production was used for the frame sentence. Its F0 contour was stylised by Praat with a frequency resolution of 2 semitones, and the pitch tier was saved. Its stylised F0 starts at 230 Hz in ‘Lee’, linearly up to 260 Hz in ‘nini’, and after the word, it linearly declined to 150 Hz by the end of the sentence. In ‘nini’, 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. The rise rate, 30 semitone / second, and the fall rate, 62 semitone / second, do not exceed the limit reported by [13].

It must be pointed out that, as it was, the non-word ‘nini’ in the frame sentence could not be used for the stimulus, since the difference in syllable duration, intensity and vowel quality between its two syllables is a strong cue to stress perception. Therefore, using Speech Filing System ver. 4.7 (obtainable from www.phon.ucl.ac.uk/resource/sfs/), its first stressed syllable /nɪ/ was annotated in the above frame sentence. Next, two copies of the syllable were extracted and concatenated, with a temporal overlap of 20ms for smoothing. This process created /nɪnɪ/ with identical syllable duration and vowel quality. The duration of each syllable was also modified from 175ms to 160ms, and the duration of /n/ and /ɪ/ was set to 60ms and 100ms, respectively. This value of 160ms is less than 168ms, the average duration of the initial /mɪ/ in ‘mimic’ that showed F0 peak-delay in [2]. Finally, the modified non-word was embedded in the sentence, replacing the original.

Then the peak F0 contour was manipulated by editing the stylised pitch tier using Praat, and the detail of the F0 peak

pattern was as follows. The peak was shifted in 20ms increments by 7 steps, in the second syllable of the non-word ‘nini’. The shift increment was set to start after the onset of /n/. This set-up produced 7 delayed F0 peak patterns located at 0ms / 20ms / 40ms / 60 ms after the onset of /n/, and at 20ms / 40ms / 60 ms after the onset of /ɪ/, in the second syllable of ‘nini’. Figure 1 shows a schematic diagram of this.

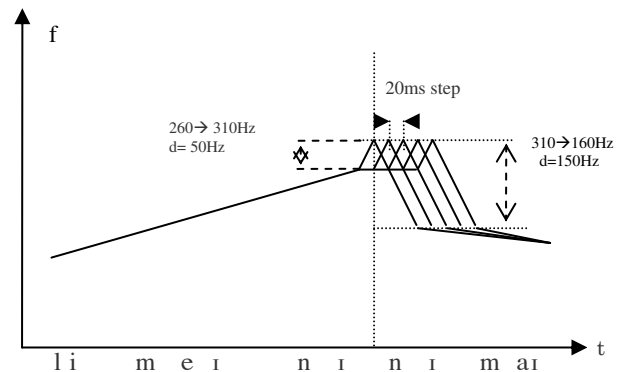


Figure 1: Schematic diagram of F0 manipulation

2.3. Procedure

Each stimulus sentence was presented in a random order to the participants 6 times, producing a total of 42 presentations (7 F0 peak patterns x 6 repetitions) per participant, and they were preceded by five 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 which of the syllables in the stimulus word /nɪnɪ/ was stressed. They were asked to circle or tick the syllable of the word (written as ‘nini’ in an answer sheet) which they thought was stressed. 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 the EN group, while all the JP participants listened to the stimuli through covered-ear headphones. None of them 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, it was found that three participants in the JP group and two in the EN group had chosen exclusively either the first or second syllables for nearly all the presentations. These five participants, therefore, were excluded from the analysis. This reduced the JP group to 40 members and the EN group to ten.

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 2 and 3 show the percentages of the first / second syllable choices for each F0 peak and for each group. Figure 2 is for the EN group and Figure 3 for the JP group. In these figures, F0 peak locations are plotted on the X axis.

¹ Sugito [3] claims that the Japanese F0 peak-delay tends to occur when the vowel in the following mora is non-high, but her observation was based on a very small set of words without quantitative analysis, and Ishihara [4] demonstrated with ample data that this is not the case. Hence, this study uses /ɪ/.

Figures 2 and 3 show that the participants of both groups perceive the delayed F0 peak as the cue to the lexical stress on the first syllable, unless the peak is delayed well into the vowel of the second syllable. In Figure 3, the perceptual pattern of the JP group shows a clear categorical curve, while the curve is less steep in Figure 2. In both figures, the 50% boundary is between the end of /n/ and 20ms after the onset of /i/ in the second syllable.

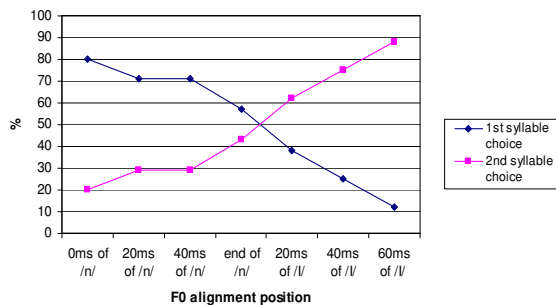


Figure 2: Results of EN group

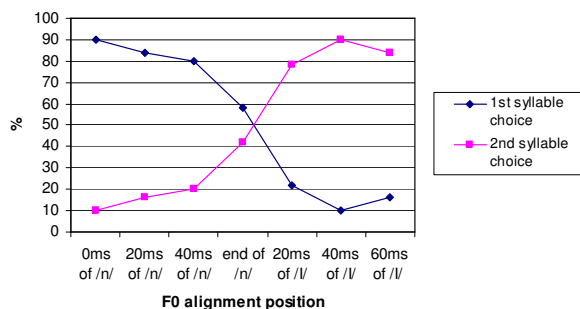


Figure 3: Results of JP group

To confirm the 50% boundary location, Probit Analysis was performed on the covariate, F0 peak location in ms (from the end of /n/), using SPSS for Windows version 15.0 software. The mean estimated 50% location, calculated by: $-\text{[intercept]} / \text{[slope]}$ across participants, was 7.1ms after the end of /n/ for the EN group, and 6.2ms after the end of /n/ for the JP group. These figures agree with the observations made about the Figures 2 and 3.

There is one potential problem to be addressed regarding the JP results: they may have perceived the stimuli in an analytic listening mode, and consequently, the categorical boundary shown in Figure 3 may not be a product of linguistic / auditory perception, but the result of simply selecting the middle stimulus point. Hence, another experiment was conducted on Japanese speakers, using an identical procedure, but with different F0 peak alignment positions.

3. Experiment 2

3.1. Participants

Thirty-seven Japanese undergraduate first-year students at Kanagawa University in Yokohama, Japan. No subject had lived in an English-speaking country or reported on hearing / language impairment. As in Experiment 1, they were not paid for participating.

3.2. Materials

The same English non-word 'nini' /nɪnɪ/, embedded in 'Lee may ___ my niece,' used in Experiment 1. As in Experiment 1, the F0 contour was manipulated by Praat, shifted in 20ms increments by 7 steps, in the second syllable of the non-word 'nini'. However, in this experiment, the shift increment was set to start 20ms after the onset of /n/, one step later than in Experiment 1. This set-up produced 7 delayed F0 peak patterns located at 20ms / 40ms / 60ms after the onset of /n/, and at 20ms / 40ms / 60ms / 80ms after the onset of /i/, in the second syllable of 'nini'.

3.3. Procedure

Same as in Experiment 1. 42 presentations (7 F0 peak patterns x 6 repetitions) per participant were made, and the task was to judge the lexical stress position of the non-word. The test was carried out in one group in a quiet Language Laboratory room at Kanagawa University, and the stimuli were played through loudspeakers for the participants. As in Experiment 1, none of them reported that their attention had been compromised by noise, or by the presence of other participants.

3.4. Results

Five participants were excluded either because they had chosen exclusively one of the syllables for nearly all the presentations, or because they had left some of the answers blank. This reduced the number of the group to 32 members. The percentages of the first / second syllable choices were obtained across all the participants and are shown in Figure 4.

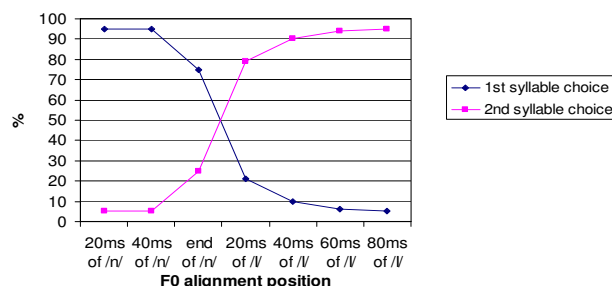


Figure 4: Results with shifted alignment stimuli

Figure 4 confirms the results of Experiment 1, namely that the Japanese speakers perceive the delayed F0 peak as the cue to the lexical stress on the first syllable in a categorical way. Furthermore, it shows that the 50% boundary still lies between the end of /n/ and 20ms after the onset of /i/ in the second syllable, in the non-word 'nini'. The results of Probit analysis, performed on F0 peak location in ms (from the end of /n/), justify this observation: the mean estimated 50% location was 9.5 ms after the end of /n/.

4. General Discussion

The results of Experiments 1 and 2 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. The categorical boundary is located immediately after the /n/ in the second syllable, for both English and Japanese speakers, and the boundary position for the Japanese is preserved even if the F0 peak alignment range is shifted, as shown in Experiment 2.

These observations are supported by the results of the previous L1 acoustic analysis. In English, [2] claims that, if the duration of the stressed syllable in disyllabic words is shorter than 200 ms, the F0 peak is aligned with the following syllable, and in this study, the end of /n/ is 220ms from the beginning of the first syllable. Similarly, in Japanese, [4] proposes that in an initially accented CVCV sequence, 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/.

Another proof from perceptual studies is found in the research review by [11], which suggests that the F0 peak shift induces the categorical change in intonation if the F0 rise goes into the following vowel more than 60ms. In our experiment, the end of /n/ in the second syllable is also 60ms. However, this must be treated with caution, since our data is about lexical, not prosodic, perception.

The categorical F0 perception by L2 listeners is also reported by [14], although this is about intonational difference: [14] discovered that non-German speakers are able to distinguish categorically two German intonation patterns evoked by the F0 peak location.

The results of this study also imply that this sensitivity to the delayed F0 peak found in L1 and L2 perception, as demonstrated by the fact that the categorical boundary position was identical, could be ascribed to lower-level, auditory perceptual processes, rather than to high-level language-specific processes. This is also suggested by [14], which concluded that the aforementioned categorical discrimination of F0 peaks by L2 listeners is due to a universal auditory principle, independent from semantically determined categorical identification. This hypothesis awaits further verification.

There is one issue to be mentioned here: in Figure 3, the perceptual curve shows a steep categorical curve for the JP group, while in Figure 2, it is less clear for the EN group, implying that Japanese speakers are more sensitive to F0 peak delay. This difference in sensitivity could be attributed to the role of F0 in lexical stress / accent perception in their L1, as discussed in the Introduction.

In English, F0 is just one of the perceptual cues in lexical stress perception and [7] proposed that duration can be a more salient cue than F0. By contrast, F0 is the sole perceptual cue in the perception of Japanese lexical accent. Hence, when listening to the stimuli, English speakers, due to the absence of 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 managed to tune into the F0 change of the L2 stimuli, as they normally do in their L1, and responded to it better. This assumption, however, would have to be confirmed by further research.

Finally, since the frame sentence of [2], ‘Lee may mimic my niece,’ that showed clear F0 peak-delay was utilised as the stimulus in this study, another potential criticism is that the obtained results may not reflect the perception of lexical stress in the non-nuclear position. To address this, we are currently conducting a series of L1 / L2 perceptual experiments on F0 peak-delay of a non-word in non-nuclear positions. We use as the stimulus the identical ‘Lee may nini my niece,’ whose nucleus is assigned either to ‘Lee’ to ‘niece’. The preliminary results indicate that Japanese speakers are able to perceive the delayed F0 peak with the height of 1.4 semitone in ‘nini’ as a stress cue in the post-nuclear position (i.e. when the nucleus is on ‘Lee’) but not in pre-nuclear position (i.e. when the nucleus is on ‘niece’). We are planning to publish a detailed report on these results, as well as the English data we are currently analysing, in the near future.

5. Conclusions

The results of Experiment 1 showed that in the perception of English lexical stress, delayed F0 peaks which were aligned with the second syllable of the stimulus word perceptually affected Japanese and English speakers in the same manner: both 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. The results of Experiment 2 confirmed that even if the F0 alignment range is shifted, the 50% boundary still lies in the same place, between the end of /n/ and 20ms after the onset of /i/ in the second syllable. These findings are supported by the data provided by previous studies on L1 acoustic analysis and on L1 / L2 perception of intonation.

6. Acknowledgements

The authors cordially appreciate the comments made by Stuart Rosen and Mark Huckvale of University College London, and by Takahito Shinya of Sophia University. This research was partially funded by Chuo University Overseas Research Programme.

7. References

- [1] Silverman, K. and Pierrehumbert, J., “The timing of prenuclear high accents in English,” in *Papers in Laboratory Phonology I*, J. Kingston and M.E. Beckman, Eds. Cambridge: Cambridge University Press, 1990, pp. 72-106.
- [2] Xu, Y. and Xu, C.H., “Phonetic realization of focus in English declarative intonation.” *J. of Phonetics*, vol. 33, pp. 159-197, 2005.
- [3] Sugito, M., *Nihongo Akusento no Kenkyu*. (in Japanese) Tokyo: Sanseido, 1982.
- [4] Ishihara, T., *Tonal Alignment in Tokyo Japanese*. PhD Diss. University of Edinburgh, 2006.
- [5] Beckman, M., *Stress and Non-Stress Accent*. Dordrecht: Foris Publications, 1986.
- [6] O’Shaughnessy, D., *Speech Communications: Human and Machine*. 2nd. ed, New York: IEEE Press, 2000.
- [7] Tokuma, S., “Perception of English lexical stress: effect of F0 peak location on English and Japanese speakers.” *Proc. of ICPHS*, Saarbruecken, 2007, pp. 1685-1688.
- [8] Sluijter, A.M.C., van Heuven, V.J. and Pacilly, J.J.A., “Spectral balance as a cue in the perception of linguistic stress.” *J. Acoust. Soc. Amer.*, vol. 101, pp. 503-513, 1997.
- [9] Tokuma, S., “Perception of English lexical stress by English and Japanese speakers: effect of duration and ‘realistic’ intensity change.” *Proc. of Eurospeech*, Geneva, 2003, pp. 2121-2124.
- [10] Redi, L., “Categorical effects in the production of pitch contours in English.” *Proc. of ICPHS*, Barcelona, 2003, pp. 2921-2924.
- [11] Niebuhr, O., “Perceptual study of timing variables in F0 peaks.” *Proc. of ICPHS*, Barcelona, 2003, pp. 1225-1228.
- [12] Niebuhr, O., “The signalling of German rising-falling intonation categories --- the interplay of synchronization, shape and height.” *Phonetica*, vol. 64, pp. 174-193, 2007.
- [13] Xu, Y. and Sun, X., “Maximum speed of pitch change and how it may relate to speech.” *J. Acoust. Soc. Amer.*, vol. 111, pp. 1399-1413, 2002.
- [14] Kohler, K., “Categorical speech perception revisited,” *From sound to sense: June 11-13, 2004*, pp. C157-C162. (Conference proceedings; downloaded from: www.rle.mit.edu/soundtosense, in January 2009.)