

The perception of English F0 peak-delay by English and Japanese speakers: we hear what they do not hear

Shinichi Tokuma, Chuo University / UCL

Won Tokuma, Seikei University

tokuma@tamacc.chuo-u.ac.jp / s.tokuma@ucl.ac.uk

1 Introduction A discrepancy between phonological tonal association and phonetic F0 peak alignment is observed in many languages. In particular, the F0 peak can be aligned after the accented syllable / mora. This phenomenon is called F0 peak-delay, and is commonly observed both in English and Japanese. In English, the F0 peak-delay is demonstrated, for example, by the data in Silverman & Pierrehumbert (1990) and Xu & Xu (2005). In Japanese, see Ishihara (2006) 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 English, four acoustic parameters contribute to the perception of English lexical stress: F0, segmental duration, intensity and vowel quality. Of these four parameters, it is generally agreed that F0 and duration are the most important cues, while intensity, if not modified¹, is less significant and vowel quality contributes little to the perception. (See papers reviewed in O'Shaughnessy (2000)) By contrast, Japanese has a lexical accent system which uses high / low pitch difference to distinguish words, specified by the location of the nucleus H+L, and Japanese speakers heavily rely on F0 in the lexical accent perception of disyllabic words (Beckman (1986)).

This asymmetry 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?

2 Experiments We have been investigating the perceptual effect of F0 peak-delay on L1 / L2 perception of English lexical stress. In a series of our perceptual experiments, a bisyllabic English nonsense word 'nini' /nɪni/ was used as the stimuli, embedded in a frame sentence 'Lee may nini my niece'. Native English and Japanese speakers were asked to determine lexical stress location of the word in the experiments. The two syllables of 'nini' were acoustically identical, and their duration and intensity were not manipulated, while the alignment position of F0 peak was altered.

Of all the results obtained, now we present those with two different F0 patterns, *Focus* and *Post-Focus positions*:

---*Focus position*: the focus was on the target word 'nini' as in the data of Xu & Xu (2005), which showed great F0 peak delay. The high F0 peak was made of medium rise (3 semitones) and substantial fall (12 semitones). The F0 peak was aligned with the second syllable of 'nini', and its alignment position was shifted in 20ms increments by 7 steps. Ten English speakers and 40 Japanese speakers were tested. The detail of this experiment is reported in Tokuma & Xu (2009).

---*Post-Focus position*: the focus was on the first word 'Lee', making 'nini' in a post-focus position. The F0 peak in 'nini' was small and symmetrical, made of 1.4

¹ Sluijter et al. (1997), after studying Dutch stress perception, suggested that intensity can serve as a strong lexical stress cue in English if its spectral balance is altered above 500 Hz region.

semitone rise and fall. As in *Focus position*, the alignment position of the peak was shifted in 20ms increments by 7 steps. Due to shorter syllable duration of the word in a non-focus position, one stimulus had the F0 peak aligned 20ms before the onset of the second /n/ but the rest had the peak aligned within the second syllable of 'nini'. Ten English speakers and 13 Japanese speakers were tested.

3 Results Figures 1-4 show the results of the perceptual experiment: Figures 1 and 2 are for the results of *Focus position* while Figures 3 and 4 for *Post-Focus position*. In these Figures, the vertical axes represent the mean percentage of first / second syllable choices, and the horizontal axes stand for the temporal F0 peak alignment positions.

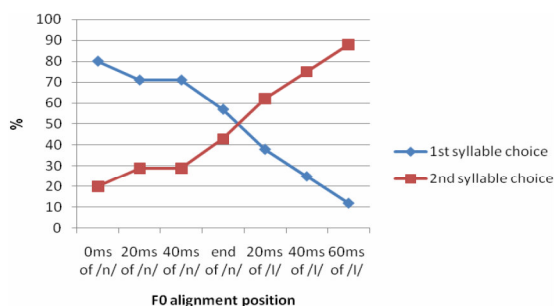


Figure 1. Focus position: English results.

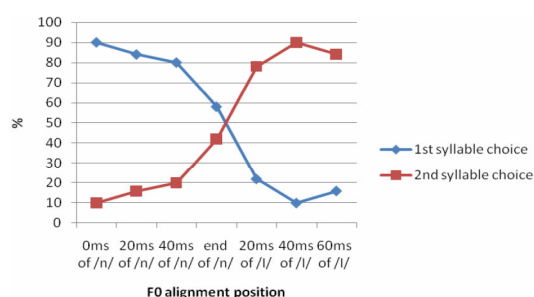


Figure 2. Focus position: Japanese results.

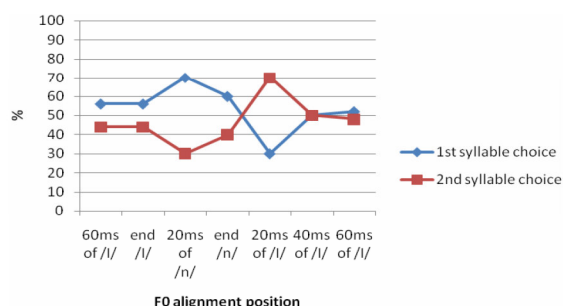


Figure 3. Post-Focus position: English results.

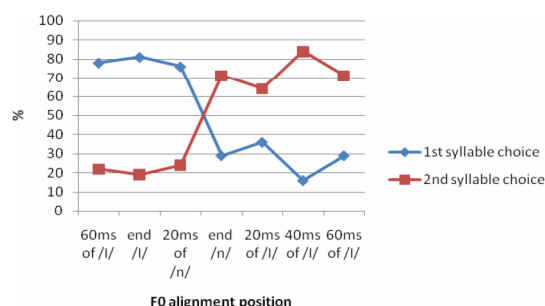


Figure 4. Post-Focus position: Japanese results.

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 participants in a mixed manner: both groups perceived the delayed F0 peaks in *Focus position* 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 but, in *Post-Focus position*, Japanese speakers were able to perceive the delayed F0 peak with the height of 1.4 semitones in 'nini' as a stress cue but English speakers were not, their answers being quasi-random. This suggests that Japanese speakers are more sensitive to F0 peak delay in English than English speakers. The closer inspection of *Focus position* supports this hypothesis: the perceptual curve of Japanese speakers in Figure 2 shows a steeper categorical curve than that of English speakers in Figure 1.

4 Discussion Why can Japanese speakers be more sensitive to F0 change? It is attributed to the role of F0 in lexical stress / accent perception in their L1, as discussed in Introduction. In English, F0 is just one of the perceptual cues in lexical stress perception, and Tokuma (2007) 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.

Similar superiority of L2 listeners in F0 cue sensitivity is reported by Wang (2008). 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 listeners can be extended to other acoustic cues. For example, Iverson et al (2003) 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.

There are two potential criticisms to be addressed regarding the choice of 'nini': (1) it has a homophonic real word 'ninny', which may have affected the English speakers' perception; and (2) English phonotactic constraints may prohibit second-syllable stress on the word. For the first criticism, the feedback from the English participants of the experiment, as well as the audience of a UCL seminar where the results of the paper were presented, show that the opinion is divided on whether the word 'ninny' is familiar to them or not. Hence, we presume that the impact of the existing word 'ninny' is not very significant. For the second criticism, a future experiment is required to establish whether the constraints truly affect the perception of lexical stress in English.

Do the findings of this study suggest anything to Japanese learners of English who wish to improve their English listening skills? Teachers in Japanese middle or high schools put not much emphasis on oral communication and teach English lexical stress as saying a vowel loudly, not as highlighting a syllable by loudness, pitch and length. Hence, in perception, Japanese learners associate the lexical stress with F0, since it is the sole acoustic correlate in Japanese lexical accent. Interestingly, Japanese learners are also able to pick intensity as a cue to English lexical stress perception because of the way it is taught, i.e., to associate stress with the loudness of vowel, as demonstrated by Tokuma (2003). This means that if somehow one can raise the sensitivity to the other crucial cue, duration, as well as to reduce the dependency on F0, in lexical stress perception by Japanese learners, it will lead to the improvement of their lexical stress perception.

Needless to say, improving Japanese learners' sensitivity to duration is not an easy task. In one of our pilot experiments, four intermediate Japanese learners of English participated in two perceptual experiments on English lexical stress perception, before and after a short phonetics lecture and intensive listening training. The lecture and training were about the importance of duration as a cue to English stress, and they lasted about 20 minutes. The task of the perceptual experiments was to identify the stress position of a nonsense word 'nur-nur' /nʊ:nʊ:/ in a frame sentence. The intensity and duration of the both syllables were manipulated, while F0 was in a linear declination from 160Hz to 150Hz within the word. The preliminary results are shown in Figures 5-8. In these Figures, the vertical axes represent the percentage of the first / second syllable choices. Figures 5 and 6 are for the results of intensity manipulation, and Figures 7 and 8 for those of duration manipulation. The horizontal axes of Figures 5 and 6 represent intensity manipulation steps (3dB) and those of Figures 7 and 8 the duration steps (20ms). These Figures demonstrate that after the training, the learners readily ignored intensity as a cue (although they are not supposed to) and that in the perception of durational difference, they showed slight improvement, but fell short of establishing a perceptual boundary. These learners might show more improvement if the training

Tokuma & Tokuma The perception of English F0 peak-delay by English and Japanese speakers: we hear what they do not hear: 4

session is repeated several times in a longer time span, such as a few months, perhaps supported by computer-assisted learning.

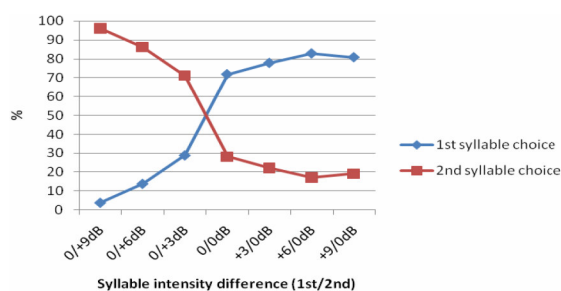


Figure 5. Sensitivity to intensity: before training

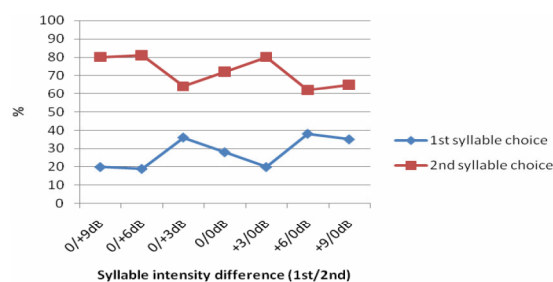


Figure 6. Sensitivity to intensity: after training

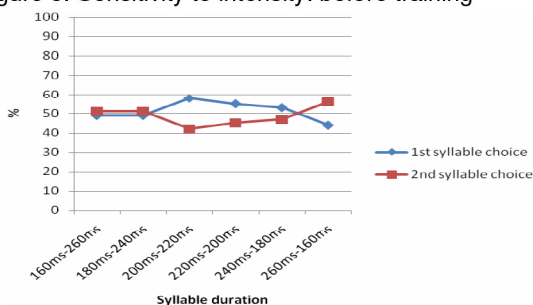


Figure 7. Sensitivity to duration: before training

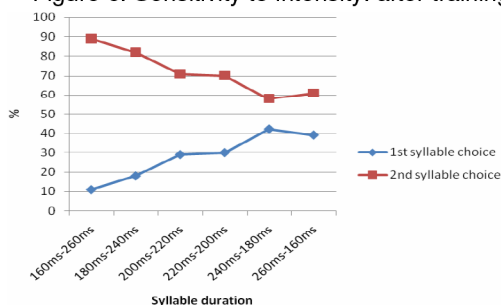


Figure 8. Sensitivity to duration: after training

5 Conclusion This study examined the perceptual effect of F0 peak-delay on L1 / L2 perception of English lexical stress, and the results demonstrated that in the perception of English lexical stress, delayed F0 peaks perceptually affected Japanese and English listeners in a mixed manner: (1) both perceived the delayed F0 peaks in a focus position 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, but (2) in a post-focus position, Japanese speakers were able to perceive the delayed F0 peak as a stress cue although English speakers could not. This suggests that Japanese speakers are more sensitive to F0 peak delay in English than English speakers. The importance to teach Japanese learners the role played by the acoustic cues to lexical stress perception is also suggested.

6 References

- Beckman, M. (1986) *Stress and Non-Stress Accent*. Dordrecht: Foris Publications.
- Ishihara, T. (2006) *Tonal Alignment in Tokyo Japanese*. PhD Diss. University of Edinburgh.
- Iverson, P., Kuhl, P.K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003) A Perceptual Interference Account of Acquisition Difficulties for Non-native Phonemes. *Cognition*, vol. 87, pp. B47-B57.
- O'Shaughnessy, D. (2000) *Speech Communications: Human and Machine*. 2nd. ed, New York: IEEE Press.
- Silverman, K. & Pierrehumbert, J. (1990) 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, pp. 72-106.
- Sluijter, A.M.C., van Heuven, V.J. & Pacilly, J.J.A. (1997) Spectral balance as a cue in the perception of linguistic stress. *J. Acoust. Soc. Amer.*, vol. 101, pp. 503-513.
- Tokuma, S. (2003) 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.
- Tokuma, S. (2007) Perception of English lexical stress: effect of F0 peak location on English and Japanese speakers. *Proc. of ICPHS, Saarbrücken, 2007*, pp. 1685-1688.
- Tokuma, S. & Xu, Y. (2009) The effect of F0 peak-delay on the L1 / L2 perception of English lexical stress. To appear in: *Proc. of Interspeech, Brighton, 2009*.
- Xu, Y. & Xu, C.H. (2005) Phonetic realization of focus in English declarative intonation. *J. of Phonetics*, vol. 33, pp. 159-197.
- Wang, Q. (2008) L2 Stress Perception: The reliance on different acoustic cues. *Proceedings of Speech Prosody, Brazil, 2008*, pp. 635-638.