Realisation of tonal alignment in the English of Japanese-English late bilinguals

Calbert Graham, Brechtje Post

Department of Theoretical and Applied Linguistics, University of Cambridge, UK

crg29@cam.ac.uk, bmbp2@cam.ac.uk

Abstract

Several factors have been attested to affect the temporal synchronisation of tonal targets such as syllable duration, segmental structure and proximity to word or intonational boundaries, e.g. [1], [2], [3]. Given the apparent language-specific nature of tonal alignment [4], it can be expected that late bilinguals who are acquiring a second language will need to learn the alignment implementation rules of that language, in addition to other aspects. This study compared the tonal alignment patterns of Japanese late bilingual English speakers and monolingual English speakers in order to investigate to what extent learners transfer their native implementation strategies to the interlanguage, and whether alignment changes with proficiency. The results show that, although initial-accented words were aligned later than final-accented words for all groups, as expected, the Japanese bilinguals aligned the former significantly later than the monolinguals. Further analyses revealed that their off-target realisations were generally limited to a specific type of syllable structure that we speculate may be linked to peak delay in their L1.These results are taken as evidence of prosodic transfer and suggest that late bilinguals will need to learn the L2 phonetic implementation rules of alignment independently of their acquisition of the phonology.

Index terms: alignment, English, Japanese, acquisition, bilingual

10.21437/Interspeech.2013-557

1. Introduction

The Autosegmental Metrical framework has become the dominant framework for analysing intonation. The central idea behind the theory is that intonational tunes are independent of the segmental constituents in the speech stream, and that intonation consists of a phonological as well as a phonetic component. The phonological component consists of binary tones (H and L), which can be associated with either a prominent syllable (i.e. H* or L* pitch accents) or phrase boundary tones (i.e. H% or L %.) The phonetic component applies language specific rules in the implementation of these underlying tones as tonal targets. Tonal alignment is therefore about the associative relationship that exists between the segmental structure with which these tones are associated as part of the intonational phonology of a language and the actual alignment patterns that result after the implementation of phonetic rules.

There is growing cross-linguistic evidence showing that prosodic context as well as segmental structure can affect the alignment and scaling of tonal targets. For instance, the upcoming (i.e. right-side) prosodic context can cause the f0 peak in prenuclear H* accents to occur earlier if the context includes a boundary or stress clash [3], and [5] reported on the effects of tonal crowding (i.e. a phenomenon in which two tones occur in close vicinity of each other resulting in a phonetic realignment of the tonal target) in Greek polar questions. Similarly, vowel duration has been reported to exert influence on accent peak alignment in Dutch [6], and f0 alignment in Japanese is more retracted in words beginning with two-mora syllable structures (i.e. CVV or CVN) than in words beginning with a single mora (i.e. CVCV) [7]. Differences in alignment patterns between nuclear and prenuclear accents have been found for several languages (for English [8], [9]; for Spanish [10]; for Dutch [11]; among others). Languages exhibit significant variability in the way they implement the tonal alignment of pitch accents and the scaling of phonological tonal targets, e.g. [4].

Given such cross-linguistic differences in alignment, it can be expected that late bilinguals who are acquiring a second language will have to learn the phonetic implementation rules for tonal alignment, even if their first language has a similar inventory of phonological categories or prosodic features. Speakers of languages that are structurally dissimilar in tonal specification in their intonational phonology – such as Japanese with only one pitch accent shape compared to American English (henceforth AE), for instance, with six pitch accents - will have to acquire both those basic phonological tones and the relationship they share in forming various contours, as well as the 'idiosyncratic' phonetic implementation rules that govern their function and distribution in speech. (This is in addition to acquiring semantic and pragmatic knowledge.) The relatively few existing studies of L2 prosody suggest that different aspects of intonation develop in different ways. That is, the acquisition of, for instance, lexical stress [12], accent distribution [13], and accentual lengthening [14] follow universal developmental paths, with learners producing 'default' or 'unmarked' values in their interlanguage irrespective of their L1, and progressing in the direction of the L2 as their proficiency increases. By contrast, aspects like pitch range, register, and direction (falling or rising) show transfer effects [15], [16], progressing from L1-like values towards L2 values as proficiency increases.

This study examines the acquisition of the tonal alignment of the peak of the H* accent of the AE contour (L) H*L by Japanese late bilingual speakers of AE. Specifically, this study addresses the following two primary research questions:

1. Are there differences in the phonetic implementation of tonal alignment patterns in the AE speech of late Japanese-English bilinguals and native speakers?
2. If yes, are differences attributable to patterns of alignment in the L1?

2. Method

2.1. Participants

Participants in the study were ten native speakers of Tokyo Japanese who learnt AE as a second language as young adults and five native age-matched speakers of AE (all from Northern Virginia). Based on the judgment of native AE raters, as well as the results of a standard Oxford proficiency test, the Japanese were placed in two groups: 'advanced' (CEFR C1)

and 'basic' (CEFR A2). It was expected that including at least two distinct levels could provide insights into how learners' knowledge of L2 alignment develops with proficiency.

2.2. Materials

The materials were adapted from the APriL project [17] for the purposes of this study. There were 32 test items which were varied for stress (two stress positions: initial-syllable accented words and final-syllable accented words) and segmental composition. (Test items actually varied three stress positions (initial, medial and final) for a different aspect of the study not reported here; only words with initial and final stress were analysed in the present study.) Syllable compositions were varied as follows: (1) syllables with both obstruent onset and coda (i.e. no sonorants); (2) syllables with a sonorant onset (and obstruent coda or no coda); (3) syllables with a sonorant coda (and obstruent onset or no onset) and (4) syllables with both onset and coda sonorants (i.e. no obstruents) – for ease of reference, henceforth No-son(orant), Son(orant)-onset; Son(orant)-coda, Son(orant)-onset:coda, respectively.) All target words were high frequency words with high familiarity ratings [18], for the benefit of participants with a lower reading proficiency. The target words were embedded in the phrase final position of sentences and were produced with identificational narrow focus in a declarative (e.g. 'She is looking for a *monkey*' and 'This is a *violin*', as situational responses to the questions 'What is Melanie looking for?' and 'What is this?', respectively.)

2.3. Recording procedure

The materials were recorded on a Nagra Aries II at 16 bits with a sampling rate of 44.1 kHz. Recordings were conducted in sound-attenuated recording booths at the University of Tokyo (Japanese participants) and the University of Colorado, Boulder (American participants). The target sentences were presented on a computer screen in PowerPoint.

2.4. Analysis

The data labelling, in Praat [19], was done according to ToBI annotation guidelines [20]. The three most common tunes observed were (1) L+H* L% with an onglide from the onset of the voiced segment of the syllable to the high target followed by a fall; (2) H* L% without the onglide of (1), and (3) L+H* H- L%, where the H* tone is sustained across the syllable and forms a plateau without any noticeable peak. In this study, we analysed the two contours that had clear f0 peaks (1 and 2) only. Following [21] we take the L+H* to be a minor variant of H*. There were 275 tokens in total. Each word was annotated with demarcation of accent position and type of syllable structure. The onset and offset of each accented syllable and the location of the H* accent were also marked. Duration measurements were extracted using a Praat script. Two measures of tonal alignment were used (distance in seconds from H* to the onset of the syllable, and from H* to the end of the syllable; however, as the results for both measures were comparable, and given limitations on space, only the results for the former are presented.

3. Results

In order to tease out the potential effects of stress and syllable structure, we conducted a MANOVA with alignment by syllable structure (four types: No-son, Son-onset, Son-coda & Son-onset:coda); stress (initial-accented and final accented) as within subjects factors; with proficiency as the independent variable. The results reveal a significant main effect of proficiency, $F(3, 14) = 8.143$, $p < 0.01$; a significant interaction between proficiency and stress, $F(3, 14) = 8.14$, $p < 0.01$, and between proficiency and syllable structure, $F(2, 14) = 4.35$, $p < 0.05$).

3.1. Effects of stress position on alignment

Post hoc tests, with a Bonferroni correction, further revealed that within the initial accented word condition, compared to the native speakers, the H* peak alignment was significantly later in both Japanese groups than in the native English group (L1 English v. Basic group, $p < 0.01$; and L1 English v. Advanced group, $p < 0.01$). There was no statistically significant difference between the two learner groups.

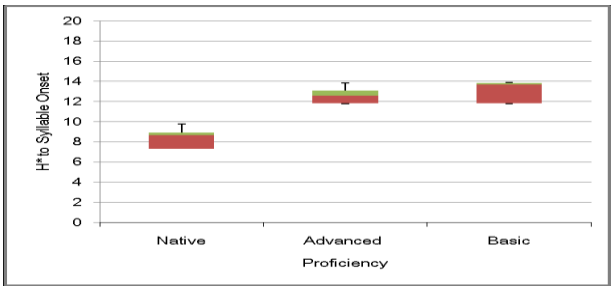

Figure 1: *Distance (in milliseconds) from H* to the syllable onset of mean initial-accented syllables for the three AE proficiency groups.*

To control for the effects of individual variability in syllable duration, which could potentially affect alignment patterns, a series of MANCOVAs was conducted with mean syllable duration of the words in each of the two-stress conditions as a covariate factor. The MANCOVAs confirm that even when syllable duration is controlled for, there was still a statistically significant main effect of proficiency on alignment patterns, $F(2, 10) = 8.18$, $p < .01$, partial eta-squared = .654. This confirms that both groups of bilinguals had significantly later H* peak alignment than the monolingual speakers in initial-accented words. Their alignment timing as a percentage of syllable duration in the two accented positions is shown in Figure 2 below:

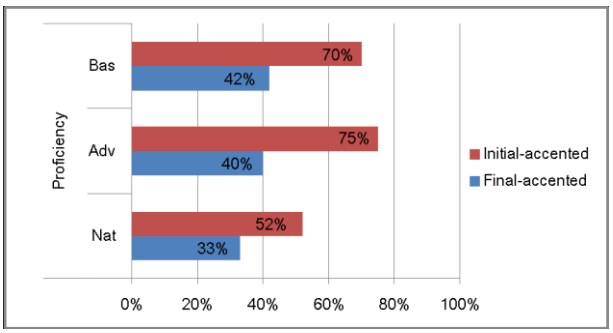

Figure 2: H* *alignment as a percentage of the total vowel duration by accent position across the three proficiency groups (Bas – basic; Adv – advanced & Nat – native).*

3.2. Effects of syllable structure

With regard to the significant interaction between proficiency and syllable structure, both groups of Japanese (Advanced, M=.159; SD=.025; Basic M = .159, SD = .018) were observed to have aligned their initial-accented f0 peaks significantly later than the Americans (M = .119, SD = .035) in the Son-onset:coda condition F (2, 14) = 4.35, $p < 0.05$. There were no overall significant differences in any of the other conditions among the three groups. In short, as can be seen in Figure 3 below, all the test groups were comparable in aligning their accent peaks significantly later in the Son-onset position than in the No-son position ; and were more retracted in the Son-onset:coda words than in the Son-onset words. However, crucially, the Americans appeared to have a more retracted mean f0 peak in Son-onset:coda syllables than in Son-onset syllables, whereas for the Japanese both these latter types were aligned later in the syllable. It is worth noting that the later alignment in syllables with a sonorant onset (son-onset) compared to syllables with obstruents in both onset and coda (no-son) could in part be due to segmentation and measurement criteria. Syllables with plosives in their onsets were segmented at the burst and were on average shorter than the three other segment types, which could make the raw alignment measures in this category appear shorter than they otherwise might have been. However, as the criteria were the same for all test groups, this has no effect on the native vs. non-native comparisons, which are the focus of this study.

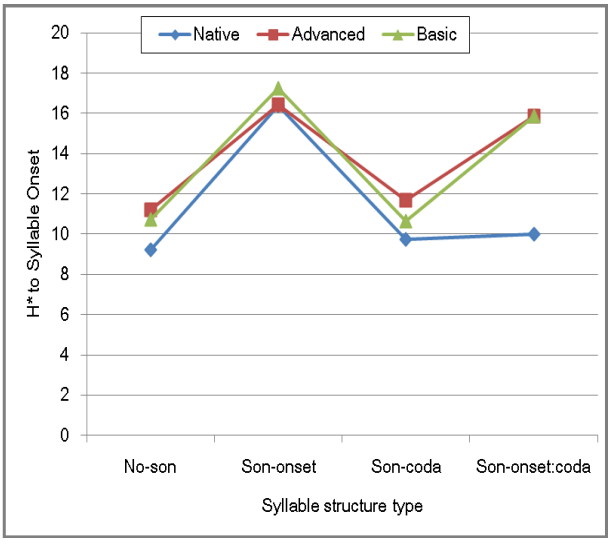

Figure 3: *Distance (in milliseconds) from H* to the syllable onset for the four syllable types in initial accented words by the three AE proficiency groups.*

With regard to the final-accented words, the analysis revealed no significant contrast between the native speakers and any of the two Japanese groups (see Figure 4). This confirms that all three groups had significantly more retracted alignment in this position than in the initial-accented words.

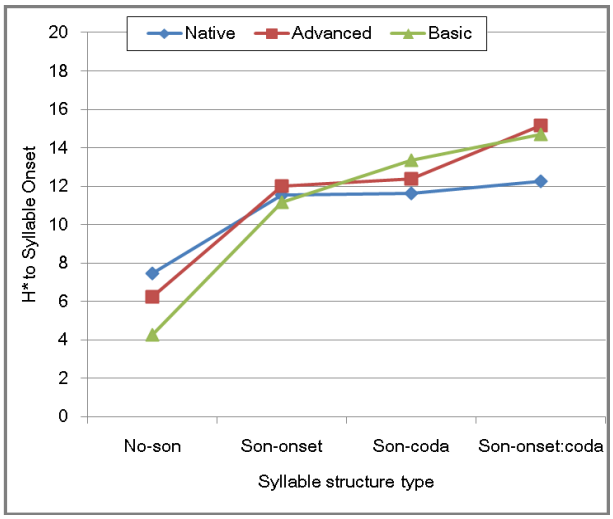

Figure 4: *Distance (in milliseconds) from H* to the syllable onset for the four syllable types in final-accented words by the three AE proficiency groups. There were no significant differences in this condition.*

The overall alignment distribution by proficiency and accent position for all 15 participants is summarised below in Figure 5. The graph shows that the three groups were generally comparable in alignment of final-accented words; however, for the initial accented words, the Japanese cluster further to the right on the x-axis, representing statistically significantly later alignment patterns than the native speakers:

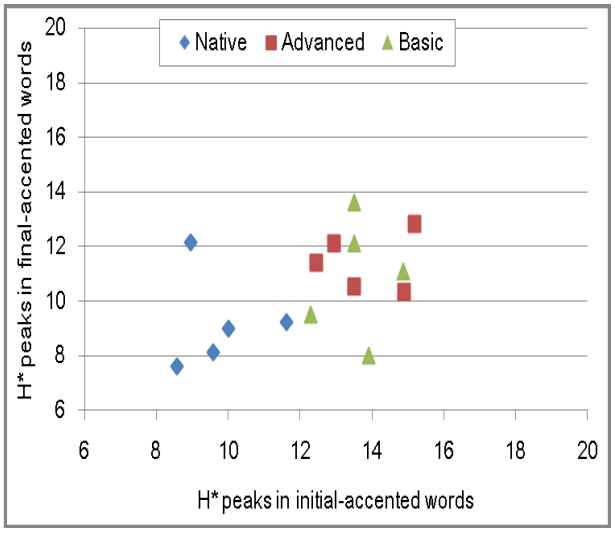

Figure 5: *Distance (in milliseconds) of H* peaks in both initial-accented and final-accented words.*

4. Discussion

We compared the tonal alignment of the (L) H*L pitch accent produced by native speakers of AE English with that of Japanese second language learners of English. Controlling for the possible effects of syllable duration, we see that both groups of Japanese learners consistently aligned the accent later than the monolinguals in the initial-accented words. Since

this is the position that has been shown to trigger *ososagari* or 'peak delay' in Japanese, e.g. [22] [7], and only peaks in this position showed the effect, this strongly suggests that peak timing transfers directly from the L1, and is not subject to a universal mechanism – at least in the L2 English of L1 Japanese learners. The finding that the Japanese appeared to have transferred their L1 alignment rules to the L2 is in line with findings in other studies that suggest that learners commonly transfer prosodic features from their L1 to the L2 (e.g. [15], [23]), including pitch accent distribution and realisation (e.g. [13], [23], [24]). Furthermore, the fact that there was no apparent proficiency-related progression toward native-like alignment patterns - given that both Japanese groups had comparable realisations - suggests that despite their general English speaking ability, the Japanese were slow to acquire or were unable to operationalise the rules for mapping phonology-to-phonetics in their L2 tonal realisation.

This finding contrasts sharply with the findings of an earlier study involving these same participants [25] which showed a marked difference between the two proficiency levels in the durational implementation of lexical stress. This implies that the phonetic implementation of prosodic features like stress in L2 speech may develop at different rates and in different ways for different phonetic correlates. It also underlines the necessity to control for individual differences in syllable duration in this study).

The final-accented words appeared in the vicinity of a prosodic boundary that has been attested cross-linguistically to trigger tonal readjustment as a mechanism to ease the effects of tonal crowding that would result from both the H* and the approaching intonation boundary tone (i.e. L%), being implemented on the same syllable. This is a fairly well-documented cross-linguistic phenomenon, so presumably Japanese functions similarly to English in this regard, which would explain why the Japanese generally had no apparent difficulty implementing the prosodic constraints of the two tones on this syllable.

The results confirming an effect of syllable structure on alignment, but limited to the initial-accented words category, suggest that whilst accent peaks in these initial-accented words were realised later than in the final-accented words for all three groups, the significantly later alignment in the late bilinguals may actually be due, at least in part, to the effects of syllable structure. Interestingly, the overall patterns produced by the Japanese were similar to the native realisations, with the only significant contrast being that the extent of the rightward pull in the Son-onset:coda condition was statistically significantly greater in the Japanese groups. It is not entirely clear what is responsible for this misalignment, but the fact that both groups of Japanese were comparable in their alignment patterns leads us to speculate that their productions are likely influenced by their native language rules of tonal alignment. It remains to be seen, however, how important a role other factors such as vowel length, distance from the word boundary, foot structure and other prosodic and segmental phenomena have on alignment patterns. More studies that examine these various factors cross-linguistically will be necessary to shed more light on the issue.

5. Conclusion

We have presented the results of an experiment examining the acquisition of the acquisition of the tonal alignment rules of English by Japanese-English bilinguals. The results revealed evidence of prosodic transfer in the actual implementation of the H* accent and suggest that late bilinguals – at least those in this study - may be selectively hindered by native language alignment rules, and are likely to be slow or possibly unable to acquire the more highly sensitive language specific rules of tonal alignment.

6. References

[1] Ladd, D.R., Faulkner, D., Faulkner, H. and Schepman, A., "Constant 'segmental' anchoring of f0 movements under changes in speech rate", Journal of the Acoustical Society of America, 106: 1543-1544, 1999.

[2] Ladd, D.R., "Phonological conditioning of f0 target alignment", in M.J. Sole´, D. Recasens, and Romero [Eds.], Proceedings of the XVth International Congress of Phonetic sciences, Vol.1:249-252, 2003.

[3] Silverman, K.E. & J. Pierrehumbert "The timing of prenuclear high accents in English", in J Kingston and M Beckman [Eds.], Papers in Laboratory Phonology I. Cambridge: Cambridge University Press, 1990.

[4] Atterer, M. & Ladd, D.R., "On the phonetics and phonology of "segmental anchoring" of F0: evidence from German". Journal of Phonetics 32: 177-197., 2004.

[5] Arvaniti, A., Ladd, D.R., Mennen, I. (2006). "Phonetic effects of focus and 'tonal crowding' in intonation: Evidence from Greek polar questions", Speech Communication, 48:667:696, 2006.

[6] Ladd, D.R., Mennen, I., & Schepman, A., "Phonological conditioning of peak alignment in rising pitch accents in Dutch", Journal of the Acoustical Society of America, 107:2685-2696, 2000.

[7] Ishihara, T., "A phonological effect on tonal alignment in Tokyo Japanese", in Proceedings of the 15th international congress of phonetic sciences, Barcelona: 615-619, 2003.

[8] Steele, S., "Nuclear accent f0 peak location: Effect of rate vowel, and number of syllables", Journal of the Acoustical Society of America, 80 (Suppl. 1): s51, 1986.

[9] Ladd, D. R., Schepman, A., White, L., Quarmby, L., Stackhouse, R., "Structural and dialectal effects on pitch peak alignment in two varieties of British English", Journal of Phonetics, doi:10.1016/j.wocn.2008.11.001, 2009.

[10] Prieto, P., van Santen, J., and Hirschberg, J., "Tonal alignment patterns in Spanish" Journal of Phonetics, 23: 473:500, 1995.

[11] Schepman, A., Lickley, R., Ladd, D.R., "Effects of vowel length and "right context" on the alignment of Dutch nuclear accents", Journal of Phonetics, 34:1-28, 2006.

[12] Archibald, J., "A formal model of learning L2 prosodic phonology". Second Language Research, *10*: 215-240, 1994.

[13] Rasier, L., & Hiligsmann, P., "Prosodic transfer from L1 to L2. Theoretical and methodological issues". Nouveaux cahiers de linguistique française, 28: 41-66, 2007.

[14] Barry, W. J., "Rhythm as an L2 problem: How prosodic is it?" in J. Trouvain & U. Gut (Eds.) Non-Native Prosody: Phonetic and Teaching Practise, 97-120, The Hague: De Gruyter, 2007.

[15] Backman, N. E., "Intonation errors in second language pronunciation of eight Spanish speaking adults learning English", Interlanguage Studies Bulletin 4 (2): 239–266, 1979.

[16] Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. Journal of Phonetics, 32, 543-563.

[17] Astruc, L., Payne, E., Post, B., del Mar Vanrell, M., Prieto, P., "Tonal targets in early child English, Spanish, and Catalan", Language and Speech, DOI: 10.1177/0023830912460494, 2012.

[18] Wilson, M. D., "The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2", Behavioural Research Methods, Instruments and Computers, 1988.

[19] Boersma, P., & Weenink, D., "Praat: doing phonetics by computer", Version 5.3.35, 2012, retrieved from http://www.praat.org/.

[20] Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J.,

"ToBI: A standard for labeling English prosody", in Proceedings of the 1992 International Conference on Spoken Language Processing (ICSLP):867-870, 1992.

[21] Pitrelli, J.F., Beckman, M., Hirschberg, J., "Evaluation of prosodic transcription labeling reliability in the ToBI framework", Proceedings of the 1994 International Conference on Spoken Language Processing (ICSLP-1994): 123-126, 1994.

[22] Sugito, M., "Nihongo akusento no kenkyuu", ["Studies on Japanese accent"], Tokyo: Sanseido. 1982.

[23] Lepetit, D., " Cross-linguistic influence in intonation: French/Japanese and French English", Language Learning 39:3: 397-413, 1989.

[24] Benet, A., Kireva, E., Pešková, A., & Gabriel, C., "Transferencia prosódica del italiano al español: el ritmo en el español de Buenos Aires y en el español como L2 hablantes nativos de italiano", in A. Plans, (Ed), Actas del V Congreso Internacional de Fonética Experimental. Caceres, Extremadura, 22-25 Octubre, 2011.

[25] Graham, C., & Post, B., "L2 English stress production by Japanese native speakers", BAAP, Leeds, 26-28 March, 2012.