

Title: Second language acquisition of intonation: Peak alignment in American English

Calbert Graham*, Brechtje Post

*Corresponding author: crg29@cam.ac.uk, Tel: +44 (0)1223760812

Phonetics Laboratory, DTAL, University of Cambridge, Cambridge CB3 9DA, UK

Abstract

The objective of the present study was to investigate (1) whether, and to what degree, late bilinguals of different L1 backgrounds are comparable to native speakers in the phonetic implementation of tonal targets in their L2, (2) whether they exhibit general patterns of acquisition irrespective of the typological closeness of their L1 to their L2, and (3) whether learners' choice of accent contours and the alignment of the high tone (H*) proceed in parallel with proficiency in the L2. More specifically, we examined the acquisition of the nuclear contour composition and the H* alignment of the American English (L)H*L- (i.e. pitch accent and boundary tone combination) in initial-stressed and final-stressed words by Japanese and Spanish late bilingual speakers at varying proficiency levels in American English. Results suggest that the L1 Spanish speakers were more comparable than the L1 Japanese speakers to the native English speakers in the phonological aspect of intonation (choice of pitch accent contour). In terms of peak alignment, we found that the late bilinguals generally tended to realise significantly later alignment than the native speakers, although the precise manifestation of this varied according to the L1 background of speakers and the stress pattern of words.

Keywords

alignment; pitch accent; Japanese; English; Spanish; acquisition

Second language acquisition of intonation: Peak alignment in American English

1.0 Introduction

It has long been established that the acquisition of a second language (L2) can entail some degree of interaction between the native (L1) and target language phonological systems of late bilinguals, often through the transfer of L1 properties to the L2. The vast majority of studies that have addressed this issue have examined the acquisition of segmental features, and only in the past two decades or so there has been a growing body of literature on the acquisition of prosodic phenomena by late bilinguals (see Gut 2009 for a review).

Prosody poses a particular challenge for second language acquisition (SLA) theories because of the difficulty in isolating the categorical or phonological sources of influences from gradient or phonetic sources. The relatively few existing studies of L2 prosody suggest that different aspects of prosody develop in different ways. That is, the acquisition of, for instance, lexical stress (Archibald 1994), accent distribution (Rasier & Hilgsmann 2007), accentual lengthening (Barry 2007) and variability in the timing of vowels (Li & Post 2014) follow universal developmental paths, with learners producing what are often referred to as '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, intonational aspects like pitch range, register, and direction (falling or rising) show transfer effects (Backman 1979), progressing from L1-like values towards L2 values as proficiency increases. Intonational features are by no means easy to define and characterise, not least because of the different assumptions researchers make and, consequently, the different frameworks they employ (e.g. see the debate on prosodic typology between Beckman & Venditti 2010, 2011 versus Hyman 2001a, 2006, 2011). It might be agreed, however, that the advent of the autosegmental-metrical framework has made it more feasible to compare the intonation of different language systems (e.g. Ladd 2001).

The autosegmental-metrical (AM) approach was originally developed as a descriptive analysis of English intonation by Pierrehumbert (1980) and others (on intuitions implicit in earlier works such as Liberman 1975, Goldsmith 1976, Bruce 1977). The AM approach has subsequently been extended to other languages (Japanese by Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988; Venditti 1995; Korean by Jun 1993; Dutch by Gussenhoven 1984, 2004; French by Post 2000 and Jun & Fougeron 1995; Greek by Arvaniti et al. 2005; among many others). The AM framework is the most widely used phonological framework for analysing intonation in various systems partly because of its simplicity, versatility and conceptual elegance.

The central idea behind the approach is that categorical phonological representations should be distinguished from gradiently varying phonetic realisation in the description. Phonologically, there are high (H) and low (L) tones that are associated either with metrically prominent syllables (indicated by an asterisk, e.g. H* or L* – often called 'starred' tones or pitch accents), or with phrase boundaries (H- and L- associate with the edges of intermediate phrases, while H% and L% mark intonation phrase boundaries). The inventory of pitch accents and boundary tones varies between languages, as do the phonotactics that constrain how they can be combined within utterances. Languages exhibit significant contrasts in their intonational phonology not only in terms of distribution and realisation of pitch accents but also in the number of tonal primitives they allow and the way in which they can be combined. The phonetic component controls the language-specific implementation of the underlying tones as tonal targets, determining their timing (alignment) and pitch height (scaling) relative to the segmental string. The present study will focus on pitch accent production and the alignment aspect of their phonetic implementation.

Previous studies (e.g. Pierrehumbert & Steele 1989; Rietveld & Gussenhoven 1995; D'Imperio & House 1997; Dilley & Brown 2007) report findings which suggest that listeners tend to process certain phonetic details of tonal targets categorically (see Ladd 2008 for a further discussion), although an alignment distinction cueing two different categories in one

language (e.g. L*+H vs. L+H* in American English) may not represent distinct categories in another language (e.g. Parisian French does not have this distinction in prenuclear position, although the timing of a rising movement does vary in French; see Post 2011, but also D'Imperio et al. 2007 for claims for the existence of contrastive alignment of non-final rises in French). Furthermore, there is evidence that this category (i.e. inventory of pitch accents) vs. gradience (i.e. the phonetic realisation of pitch accents) distinction may even vary across different dialects of the same language, for example, L+H* and H* in American English are manifested as the same category in some dialects (e.g. Minnesotan dialect), and as distinct categories in others (e.g. Southern California dialect) – (Arvaniti & Garding 2007). Moreover, research (e.g. Niebuhr et al. 2011) suggests that individual differences in alignment can also be found within the same language variety.

A key assumption of the original AM framework is that tonal association to segments is determined by phonetic alignment, though this view has been challenged by some researchers who argue that peak alignment may also be phonologised without any evident phonetic motivation. Note, for example, a study by Arvaniti and her colleagues (1998) that challenges the concept of starredness based on phonetic alignment by showing that a language like Greek has bitonal accents where none of the tones are aligned with the accented syllable. More generally, the available evidence from this line of research suggests that tonal alignment in pitch accents is not only phonetic but can also be phonologically specified (see Prieto et al. 2005 and D'imperio 2006 for insightful discussions of this point¹).

Second language learners may have difficulties not only in producing context-appropriate intonation patterns, but also in their implementation of the phonetic details of their production. We know very little about how the phonological and the phonetic component of intonation develop in L2 learners. Mennen (2014) introduced a new model – the L2 Intonation Learning Theory (LILt) – as a first step to standardise variables relevant to intonation learning. The model is based on the premise that cross-language differences in intonation can occur in a number of dimensions, including the inventory of pitch accents and their phonetic realisation. This opens the possibility for research on how these various dimensions are acquired by learners of different proficiency levels and typologically different L1s. This could shed light on a number of outstanding issues, including the following.

1. How learners acquire the phonological and the phonetic aspects of intonation. The findings in the area of L2 segmental acquisition (e.g. L2 acquisition of vowels) generally show that even when learners have acquired a specific category of sound (i.e. a phoneme) their manipulation of individual acoustic cues may still not be native-like.
2. How the various aspects of intonation develop with proficiency in a target L2. The experimental evidence for segmental processing, and indeed the limited research that has been done on L2 prosodic acquisition, suggest that experience with the target language typically playing a role in the degree to which L2 speech is perceived as natural or intelligible.
3. What the role of L1 typology is in L2 intonation. Studies in segmental acquisition and the limited research on L2 intonation suggest that L1 background plays a crucial role in L2 speech production.

The study reported in this paper examines the acquisition of pitch accent contour shape and alignment of the high tone in the (L+)H*² pitch accent in American English in the final falling (L+)H*L-L% contour by L1 Japanese and L1 Spanish speakers who are late bilingual speakers of American English. More specifically, the study compares the choice of pitch accent contour shape and tonal alignment (location of tones in relation to the segmental

¹ It may be worth indicating here that the notion of starredness is less controversial in the formalisation of American English intonation within the AM framework, and for this reason this point will not be discussed in further detail in the present study.

² Please note that (L+)H* is used in this paper to signify either one or the other of the following two pitch accents: H* and L+H*. As will be discussed later, L+H* can be treated as a minor variant of H*.

string) of the H* peak in the (L+)H*L-L% pitch contour in L1 and L2 American English within the autosegmental-metrical approach.

1.1. Intonation and pitch accent

English and Spanish have been described as stress accent languages in which stressed syllables act as the potential docking sites for the placement of accents in larger discourse contexts. Sentence-level prominence in English and Spanish is realised through the superposition of tonal patterns on prominent syllables (and phrasal boundaries, as described above). Bolinger (1985) showed that a syllable that is accented by means of pitch serves the dual function of signalling accent whilst simultaneously contributing to the melody of the utterance. This melody is primarily achieved through systematic fluctuation between high and low pitch. On the other hand, Japanese has been described as a pitch accent language in which the pitch accent is a lexical property of a word, and its presence or absence depends on what word is being uttered (Beckman & Pierrehumbert 1986; Pierrehumbert & Beckman 1988; Venditti 2005; *inter alia*). More precisely, Japanese makes lexical contrasts using pitch accents in two ways: (i) its presence vs. absence (thus maintaining a paradigmatic contrast between accented and unaccented words, according to Beckman & Pierrehumbert 1986), and (ii) if present, its location. The phonetic make-up of a pitch accent in a stress accent system is not equivalent to that of a pitch accent language. That is to say, in addition to pitch, which is the only cue used in marking accents in Japanese, stress accent languages like English and Spanish use other phonetic parameters, such as duration and intensity, to mark accent.

In AM descriptions as many as six distinct pitch accents have been proposed for American English: two monotonal accents (H*, L*), and four bitonal: H*+L³; H+L*; L*+H; L+H* (Pierrehumbert 1980; Pierrehumbert & Beckman 1988). According to Pierrehumbert's (1980) model of American English intonation, declarative statements or assertions are typically realised with a final falling (H*L-L%) pitch pattern. It has been shown that the American English L+H* bitonal pitch accent is typically associated with newness and is thus normally used to mark narrow focus in declarative utterances, although H* might also be used. Both these accents are typically followed by a fall within the accented syllable. Distinguishing between L+H* and H* can be problematic and there is some debate as to whether the two accents should be considered as phonologically distinct pitch accents or whether L+H* is a minor variant of H* (e.g. Pitrelli et al. 1994; Ladd 1996:84) – we will return to this issue later.

In Puerto Rican Spanish (PRS) – the native dialect of the Spanish speakers in this study – and indeed in other Caribbean varieties of Spanish, the !H+L* pitch accent is typical in the nuclear position of broad focus statements (Armstrong 2010). This pitch pattern is phonetically realised as a fall from a relatively high tone in a compressed pitch range to a low target. Neutral declarative statements in PRS typically end in a falling (low) pitch pattern (i.e. with a low boundary tone L%). For marking narrow focus declarative statements, PRS appears to employ similar pitch accents to English, though there are important differences in their phonetic realisation. In the nuclear position of narrow focus statements the typical pattern in PRS is the H* pitch accent realised as a high plateau with no preceding fundamental frequency (f₀) valley (Armstrong 2010). The bitonal L+H* is also attested and is realised as a rising pitch movement during the accented syllable with the f₀ peak located within the syllable, similar to American English. However, in Puerto Rican Spanish Armstrong observes that the peak may be reached early in the syllable in narrow focus statements.

³ With the development of the American English ToBI (AMToBI; Beckman et al. 2005) convention for prosodic transcription, however, a number of modifications have been incorporated including the deletion of the H*+L accent in American English, which Silverman and her colleagues (1992:868) describe as the 'down-stepped inducing version of H*'. This reanalysis means that instead of the original six only five pitch accents are currently postulated for American English according to the ToBI transcription standard.

The range of pitch accent types in Spanish and English contrasts with Tokyo Japanese, which has only one pitch accent type (represented as H*+L in J_ToBI, Venditti 2005) that is characterised by a sharp pitch fall from a high accented mora to the following mora (Pierrehumbert & Beckman 1988). The intonation pattern of Japanese declarative statements is similar to that of English in that both are canonically realised with a falling nucleus intonation (%L H*+LL% in Japanese and (L+)H*L-L% in English), when the word carries a pitch accent. However, if there are no accented words, there is usually no fall in pitch in Japanese. Note that whereas the fall of the Japanese H*+L pitch accent is crucial to the perception of which mora is accented, in an utterance-final nuclear syllable (i.e. when the last mora in the sentence bears a pitch accent) this contrast may be neutralised (cf. Vance 1995), as the phonation is about to stop and there is insufficient phonetic material remaining on which to realise the fall. (Note also that a high phrasal tone can cause the pitch to plateau until the accented syllable – e.g. Pierrehumbert & Beckman 1988.) On the one hand, it seems unlikely that suspension of a pitch fall would be detrimental to the basic meaning conveyed by the lexical pitch accent if the contrast is between a final-accented word and a word that is accented on a non-final mora. On the other hand, it might be expected that neutralisation may result in the loss of the phonological contrast between a final-accented word and an unaccented word appearing in an utterance-final position, or when they are produced in isolation. There is some evidence that this might indeed be problematic for Japanese native listeners (cf. Sugiyama 2012).

It might be expected that whilst there are important differences in pitch accent configurations, the prosodic systems of English and Spanish are generally more comparable than English and Japanese in terms of the function, distribution, and acoustic properties of pitch accents. As we will discuss below, there are also differences between these three languages in the alignment of target tones within a pitch accent. Investigating how speakers of different L1 backgrounds acquire the intonational and temporal aspects of accent production in their L2, and their relation to segmental structure, can potentially provide new insights into the role the L1 plays in the acquisition process. The present study is original in examining the late acquisition of pitch accent contour choice and tonal alignment (precisely, its relation to syllable structure and stress patterns of words of English) by speakers of Spanish (Spanish and English being prototypically stressed accent languages, as described earlier) and by speakers of Japanese (a pitch accent language) in the same experiment. Language-specific as well as typological differences in prosodic structure and their phonetic implementation are expected to play a significant role in the acquisition of the alignment of English phonological targets by late bilinguals (Spanish L1 & Japanese L1 speakers).

1.2. Cross-language research on tonal alignment

Tonal alignment refers to the temporal coordination of tonal targets with segmental structure that realises them. Various studies on alignment have found evidence of considerable consistency in the alignment of f₀ turning points with the segmental string (Caspers & Van Heuven 1993; Prieto et al. 1995; Arvaniti et al. 1998; Ladd et al. 2000). Ladd et al. (1999) refer to this phenomenon as the *segmental anchoring hypothesis*. In this hypothesis tones are seen as anchored or fixed to specific points in the segmental string, such as at the beginning of the stressed syllable or the following unstressed vowel. There is evidence for consistent tonal alignment in different languages (e.g. Arvaniti & Ladd 1995 for Greek; Atterer & Ladd 2004 for German, Xu 1998 for Chinese; Ladd et al. 2009 for English; Post 2011 for French; see D'Imperio 2011 for an excellent review).

Japanese is reported to manifest a peak delay phenomenon referred to as *ososagari* ('late fall'). Sugito (1982) claimed that in *ososagari* the beginning of the f₀ fall for a pitch accent occurs after the end of the associated mora. He also found that this was limited to initial-accented words whose second mora had a non-high vowel. Ishihara (2003) found evidence in support of *ososagari*, but suggests that segmental composition did not play a role, contra Sugito's claim. Ishihara (2003) also found that in non-initial-accented words there was earlier alignment of the peak of the pitch accent than those of initial-accented words (see

Ishihara 2003, pp. 65-70, for a discussion of possible explanations for this finding in Japanese).

Although segmental anchoring has been shown to be a considerably robust cross-linguistic phenomenon, however, there are several factors that have been reported to affect the temporal alignment relation between tonal targets and segmental string at a language-specific level. For instance, an upcoming prosodic boundary (or another accent in the case of prenuclear accents) can cause the f₀ peak in H* accents to align earlier with the segmental string (Silverman & Pierrehumbert 1990) in order to avoid tonal crowding, that is, a phenomenon in which two tones appear in close proximity to each other. This has been observed in Greek polar questions (Arvaniti et al. 2006). Phonological vowel length has also been reported to exert influence on accent peak alignment in Dutch (Ladd et al. 2000). (See Ladd 2003 and Ladd et al. 1999 for more on phonological factors in f₀ alignment.) Similarly, there is plenty of cross-linguistic evidence that the peak of nuclear accents is reached earlier with an increase in the number of syllables to the utterance end (a type of tonal repulsion). Arvaniti and her colleagues (1998) found that the H in L*H Greek nuclear accents shifted earlier with fewer syllables after the stressed syllables. Similar results have been found for Dutch (Schepman et al., 2006) and Irish (Dalton & Ní Chasaide, 2005), among others.

Differences in alignment patterns between nuclear and prenuclear accents have been found for several languages: Steele 1986 (for American English), Ladd et al. 2009 (for two varieties of British English), Nibert 2000 (for Spanish), Arvaniti & Baltazani 2005 (for Greek), Schepman et al. 2006 (for Dutch). In English, the H* peak appears to be aligned very close to the beginning of the stressed vowel with earlier alignment in initial-accented vs. in final-accented positions (Ladd et al. 2009). Some authors report that the H tone of rising pitch accents in Spanish is aligned with the posttonic syllable in prenuclear accents and within the tonic syllable in nuclear accents (see, inter alia, Face 2002, Face & Prieto 2007, Hualde & Prieto 2015, for detailed discussions). Note though that these language-specific patterns are likely to vary significantly from one dialect of a language to the next, as mentioned before.

Finally, the number and type of segments in a syllable and their intrinsic duration, as well as speech rate, may influence the alignment of tones. There is evidence that the peak location of nuclear accents in English is influenced by segment type (whether sonorant or obstruent) in onset and coda classes, as well as their corresponding durations (van Santen & Hirschberg 1994). Jilka and Möbius (2006) also confirm that the peak in German is aligned earlier in the syllable when there is a sonorant in the onset than when there is an obstruent. In codas, the effects were not as clear.

1.3. Previous research

A number of previous studies have examined the acquisition of intonational phonology and its phonetic realisation by bilinguals. Focussing on early bilinguals, studies such as Elordieta 2002 and Elordieta & Calleja 2005 have investigated the acquisition of L2 Spanish intonation by L1 Basque speakers. The local variety of Northern Bizkaian Basque of these speakers only has an H*+L pitch accent (similarly to Tokyo Japanese). The researchers found that these speakers do not transfer this pitch accent to their L2 Spanish; rather, they produce rising pitch accents in prenuclear and nuclear positions, like in standard Peninsular Spanish. They showed that the peaks of their L+H pitch accents are consistently realised within the stressed syllable, that is, realised as L+H*, even in prenuclear position. On the basis of these findings, the researchers proposed that it is the alignment of the H tone with the tonic syllable in Northern Bizkaian Basque that is transferred, that is, the H* of the native H*+L. This finding appears to illustrate target-like acquisition of pitch accent type (L+H) but non-target-like acquisition of the phonetic alignment of tones. The study of peak alignment in Spanish in contact with other language varieties generally shows early peaks in Spanish (e.g. Elordieta 2003 for Basque; O'Rourke 2004 for Quechua; Colantoni & Gurlekian 2004 for Italian; Hualde & Schwegler 2008 for Palenquero). This general finding suggests that early peak may be a general simplification pattern arising, although there are notable exceptions (e.g. early peaks were not found in the case of Miami Spanish in contact with English; cf. Alvord 2010a). It should be noted that there is some evidence that a preference for early peaks in Spanish in

contact cases may, at least in part, be explained as direct transfer of intonational patterns from the native language to Spanish (Colantoni & Gurlekian 2004 and Colantoni 2011). It may also be of some significance that early peaks in bilinguals in contact situations have only been consistently observed in the context of Spanish as a second language, and usually only in broad focus declarative utterances. This raises the question of the generalisability of these results to other target languages, speech materials, and whether these patterns can be observed in the speech production of late bilingual learners.

Mennen (2004) examined the tonal alignment of a phonologically identical f₀ pattern (LH) by late bilingual speakers of Dutch and Greek. Dutch and Greek differ in the phonetic realisation of this pattern such that Dutch canonically has earlier timing than Greek and is influenced by the phonological length of the vowel of the accented syllable. Mennen observed that even advanced L2 learners transfer the tonal alignment patterns of their first language to the second language, suggesting that L2 learners may be selectively hindered by their L1 in acquiring the alignment implementation regularities of a second language. Interestingly, the study also showed that in the Greek of Dutch native speakers, only the phonetic implementation was transferred to the L2, since the target pitch accent in both languages share a similar pattern. There was also evidence of a bi-directional influence of Greek on Native Dutch alignment measures, suggesting that the alignment of target tones in bilinguals may be not be entirely comparable to that of monolingual speakers in either of their languages.

There is evidence that some L2 learners may use narrower or less variable pitch range than native speakers of the target language (Backman 1979 for Spanish L2 learners of English; Willems 1982 for Dutch L2 learners of English; Busà & Urbani 2011 for Italian L2 learners of English; Ullakonoja 2007 for Finnish L2 learners of Russian). However, with increasing proficiency in the L2, learners appear to develop more target-like phonetic implementation (Ullakonoja 2007; Aoyama & Guion 2007). These studies suggest that structural differences in the prosodic organisation of the L1 and L2 may generally affect transfer patterns. D'Imperio (2002) argues that the alignment of tones and their recoverability by humans may be guided by universal characteristics as well as language-specific constraints. It can be expected then that language-specific differences in tonal alignment could pose a difficulty for L2 learners, particularly at the early stages of acquisition. It is not yet known how differences in the alignment realisation of intonational pitch accents affect an L2 learner's ability to produce target-like alignment patterns in their L2.

1.4. Goals of study

The objective of the present study was to investigate (1) whether, and to what degree, late bilinguals of different L1 backgrounds are comparable to native speakers in the phonetic implementation of tonal targets in their L2, (2) whether they exhibit general patterns of acquisition irrespective of the typological closeness of their L1 to their L2, and (3) how learners' choice of accent contours and the alignment of the high tone (H*) develop with proficiency in the L2.

1.5. Hypotheses

1. (a). Given the prevalence of H* plateaux in PRS in the nuclear position of narrow focussed utterances, a high proportion (compared to the native English speakers) of H* plateaux in the English of PRS would indicate a possible L1-related effect. Similarly for the L1 Japanese speakers, a high proportion of plateaux in final-accented words would indicate a possible L1-related effect in the suspension of the pitch fall.
(b). More target-like productions by the Spanish than Japanese would suggest a typologically motivated difference.
2. (a). Whilst we cannot determine precise transfer effects in alignment in this study, possible L1-related influence in alignment would be evidence of the potential effect of *ososagari* (late fall) in initial-accented words as opposed to relatively earlier

alignment in the final-accented words in the production of L1 Japanese speakers. We would expect PRS speakers to be similar to the English native speakers in realising H* peaks (of non-plateau-type accents) within the accented syllable. Further, we would generally expect all the language groups to realise earlier alignment in final-accented words, compared to later alignment in initial-accented words, due to the effects of tonal repulsion triggered by the presence of a boundary tone.

(b). It is expected that speakers will realise H* peaks earlier when there is a sonorant, as opposed to an obstruent, in the onset (cf. Jilka and Möbius 2006). Given cross-linguistic differences in segment articulation patterns, it might be expected that there will be significant differences based on L1 background in the alignment of H* within syllables composed of different segment types.

3. Significant differences in the late bilinguals' alignment of a similar tonal target, compared to native English patterns, would confirm that these two aspects are not acquired in parallel with each other. If, regardless of the L1 of the learners or their degree of proficiency, they use a similar modification strategy in their realisation of tonal alignment, this may suggest a general pattern in L2 acquisition.

2.0 Method

A total of 800 (32 words x 5 speakers per group x 5 groups) (L+)H*[L-L%] pitch contours were elicited in a naming task involving words that have been controlled for stress position, metrical structure and segmental composition – phenomena that may affect timing in cross-linguistically variable ways.

2.1. Participants

Participants in the study were 10 native speakers of the Tokyo dialect of Japanese, 10 native speakers of Puerto Rican Spanish and 5 native speakers of American English (specifically, all these speakers are from Northern Virginia/Washington and speak a mainstream variety with post-vocalic rhotics). We varied the L1 of bilinguals in order to assess how large a role the L1 background of late bilinguals plays in their L2 acquisition, independently of the role of proficiency. All the bilinguals reported the American dialect of English to be the variety to which they had been predominantly exposed and to be their L2 target. None of the participants reported any hearing or reading deficiencies.

The non-native speakers were divided into three proficiency groups (per L1), based on the independent judgements of three native speakers of English who were trained English assessors. Raters listened to a wide range of speech samples recorded in different contexts from a pool of 30 Japanese L1 speakers and 18 Spanish L1 speakers. Speech samples for the speakers were presented randomly using SuperLab software (Cedrus, Phoenix, Arizona), at the end of which raters were instructed to rate them in terms of how accented they perceived them to be on a scale of 1 to 3, where 1 was 'advanced', 2 'basic' and 3 'other'. Only the data for learners who were placed in the 'advanced' (i.e. CEFR C1) group or the 'basic' (i.e. CEFR A2) group by at least two of the three raters were included in the study. The phonetic experiment lasted approximately 30 minutes and participants received a token fee. There were 3 females and 2 males in each group as shown below (the age range of native speakers was 20-24; see Appendix A for further biographical details of the bilingual speakers).

Table 1: showing the language groups and their proficiency in this study

L1	English Proficiency	Female	Male	Total
English	Native	3	2	5
Japanese	Advanced (C1)	3	2	5
	Basic (A2)	3	2	5
Spanish	Advanced (C1)	3	2	5
	Basic (A2)	3	2	5

2.2. Materials

The test items (some of which we adapted from the APriL project, Astruc. et al. 2013) were varied for stress (two positions: initial-syllable stressed words and final-syllable stressed words) and segmental composition: (1) syllables with an onset sonorant (S_S/O) and syllables with an onset obstruent (O_S/O). For the sole purpose of ensuring consistency across these categories in determining the likely influence of consonants following the accented vowel, we assume intervocalic consonants to be ambisyllabic, and accordingly classified them as belonging to the coda of the preceding syllable. The target words were mostly high frequency words with high familiarity ratings (Wilson 1988), for the benefit of subjects with a lower level of proficiency. The target words, which were mostly disyllabic, were embedded in the phrase-final position of sentences and were produced in a declarative frame. The dialogue was structured to elicit narrow focus intonation (i.e. main accent assigned to the focussed constituent). A sample of the dialogue is as follows:

Experimenter: What is Melanie looking for?

Participant: She's looking for a monkey.

In this example the target word 'monkey' is stressed on the initial syllable /mon/ which contains onset (and coda) sonorants and accordingly satisfies the S_S/O condition. Either an animal or an object is given. See appendix B for the complete set of items and dialogues.

2.3. Recording procedure

The dialogue was presented in PowerPoint on a computer monitor with participants giving their responses to the experimenter's questions at their normal speaking rate. Participants performed this task twice (in the analysis, the repetition was only used if there was a problem with the first response, e.g. if a reliable pitch trace could not be automatically obtained). Participants were also instructed to repeat their answer if they made a mistake. The recordings were made on a portable Nagra Aries II, with 16 bit resolution and a sampling rate of 44.1 kHz in sound-attenuated recording booths at the University of Tokyo (for the Japanese participants), and the University of Colorado, Boulder (for the American English L1, all of whom were visiting from Northern Virginia/Washington, and for the Puerto Rican Spanish L1 participants).

2.4. Analysis

2.4.1. Intonation contours

The data were labelled by hand according to the AMToBI annotation guidelines for American English intonation (Beckman & Hirschberg 1992; Beckman & Ayers-Elam 1997). The three most common tunes observed were (type 1) L+H*L-L%, with an onglide from the onset on the sonorant part of the syllable to the high target followed by a fall; (type 2) H*L-L% without the onglide and (type 3) L+H*H-L% or H*H-L% (hereafter (L+)H*H-L% to represent the optional onglide), where the H* tone is sustained across the syllable and forms a plateau without any noticeable peak (see Figure 1 for stylised contours and Figure 2 for data

samples of the measurement points that illustrate the annotation scheme for the three accent types). The proportion of occurrence of the three observed accent types will be presented for cross-group comparisons, but only the two contours that had clear f0 peaks (i.e. types 1 and 2) were used in the acoustic analyses reported in this study (thus the H in type 3 accents was not marked). Since an exploratory analysis of the data indicated that the alignment of the peak had the same overall pattern for type 1 and type 2 accents within groups, we combined the two in the acoustic analyses reported here, which is also in accordance with Pitrelli et al. (1994). In other words, an assumption of either H* or L+H* did not alter the results of tonal alignment patterns reported in this study, and therefore does not appear to have any obvious repercussions for the conclusions we draw.

Insert Figure 1 about here

Insert Figure 2 about here

2.4.2. Coding and measurements

The target words of each sentence were segmented into syllables, consonants and vowels and both time and tonal measurements taken. As described earlier, in all cases the vowel nuclei were syllabified with the preceding and the following consonant, where present. The syllables were segmented into consonants and vowels according to standard criteria (Kiesling et al. 2006). More specifically, voiceless stop closures, where present, were segmented from the burst in all positions to make them comparable with cases in which they were word-initial. Sonorant consonants were identified at the onset of the voicing and a steady f0 transition. Change in the amplitude displayed in the waveform was used in cases where the voicing onset was not easily identifiable. Two segmental and three tonal targets were marked by hand with the following labels:

Segmental labels

1. Onset (O) of the accented syllable
2. End (E) of the accented syllable

Syllable segment type labels

The two syllable composition types (S_S/O, O_S/O) were annotated, and the target words were labelled for stress position (initial vs. final).

Tonal labels

H* peaks of target accents

The onset and offset of each accented syllable were annotated as well as the location of the peak for the H* target. Microprosodic effects (such as the typical dip produced by nasals) were disregarded. Measurements of syllabic durations, tonal targets (f0 and timing at H* peaks, and f0 at L%) were extracted using Praat scripts (Boersma & Weenink 2014). We used the autocorrelation method for f0 tracking, and checked it against narrow-band spectrograms. We manually corrected for octave errors and other artefacts. In these and all other transcriptions we combined the auditory impression of the accent with the visual representation of the f0 trace. Further, an independent annotator who is trained in intonation labelling conducted a complete labelling of all the test items. Inter-annotator agreement according to Cohen's kappa was 0.78 for agreement on the presence or absence of pitch accents in the utterance (but no disagreement with regard to nuclear pitch accents, which are the focus of this study) and 0.84 for agreement on the identification of pitch accent types. Where the annotators did not agree on the type of pitch accent, the authors presented these problematic cases at a research seminar for discussion with other trained annotators. All the cases were then resolved and one of the three pitch accent patterns assigned in each case.

2.5 Measures of alignment and normalisation procedure

Measures of tonal alignment used in this study: distance (in milliseconds) from (1) the onset of the syllable to the H* target (2) the H* target to the end of the syllable. Both measures have been commonly used in previous studies (e.g. Prieto et al. 1995; Chen & Mennen 2008; Astruc et al. 2013, but see Xu & Liu 2006 who argue that syllabic onsets provide a more robust anchoring point than the ends of syllables). Given that tonal alignment is a temporal phenomenon in being timed relative to segmental landmarks with the syllable it is likely sensitive to duration variation within the syllable. Thus we subjected the two alignment measures, H*-SO and H*-SE, to prior standardisation using the z-score transformation. Further, to normalise the data, a third measure (SylNorm) was used that measures alignment proportionally, relative to the duration of the accented syllable.

3.0 Results

3.1 Occurrence of plateau-type accents

The first analysis examines the proportion of plateau-type H* accent within each of the five language groups (i.e. L1 Japanese-Basic English (hereafter JBE), L1 Japanese-Advanced English (hereafter JAE), L1 Spanish-Basic English (hereafter SBE), L1 Spanish-Advanced English (hereafter SAE), and Native English (hereafter NE). Of the 800 tokens, 685 tokens (roughly 86%) were produced with an (L+)H*[L-L%] falling accent, with or without the onglide and the remaining 115 tokens (roughly 14%) produced with an extended plateau (Type 3). The Japanese produced more (type 3) plateaux (JBE: 28% and JAE: 19%) than the Spanish (SBE: 6%, SAE: 8%) and monolingual English speakers (NE: 11%). Interestingly, the plateau patterns for the monolingual speakers of American English are similar to findings for British English in previous studies (e.g. 8% in Astruc et al. 2013). The proportion of plateaux produced by the five groups in each of the two stress positions (initial vs. final) is as follows: JBE (initial: 10%, final: 90%), JAE (initial: 8%, final: 92%), SBE (initial: 44%, final: 56%), SAE (initial: 52%, final: 48%) and NE: (initial: 49%, final: 51%). The arcsine transform was applied to the data (formula: $Y = \arcsine \sqrt{p} = \sin^{-1} \sqrt{p}$; where p is the original proportion in percentage and Y is the result of the transformation) for the Type 3 (plateau-type) accent. A one-way ANOVA was then conducted to compare, by language group, the distribution of this accent in the productions of the late bilinguals to the native productions. The results reveal a main effect for language group ($F(4,24)=8.93$, $p<.01$). Post-hoc tests (Bonferroni corrected) revealed that only the JBE group produced the plateau-type accent significantly more frequently than the native speakers. The proportion of each H* variant produced by the five groups is shown in Figure 3 (though, as discussed earlier, type 1 and type 2 are treated as variants of the same accent):

Insert Figure 3 about here

3.2. Alignment

The statistical significance of the alignment results was tested using a series of mixed effects models in the R-environment (R Development Core Team, 2016). The model was run separately for each of the three dependent variables (H*-SO, H*-SE and SylNorm) with proficiency, L1 background, stress position and syllable structure as fixed effects. Speaker and items were included as random effects with all possible random intercepts and random slopes. Additionally, for each dependent variable, the model tested for interactions with proficiency and stress position. Linear mixed effects analyses were performed for each fixed factor through a series of iterative tests using the multilevel approach with the baseline model specifying the random variables (speaker and items). We used the *anova* function in R to compare fitted models.

The statistical analysis revealed no significant main effect of proficiency on any of the two delay measures of alignment. However, there was a main effect of stress for each measure (H*-SO: $\chi^2 = 6.00$, $p<.05$; H*-SE: $\chi^2 = 18.39$, $p<.001$). There was a significant

interaction between proficiency and stress for the H*-SO measure: $\chi^2 = 15.56$, $p < .01$), but no such interaction for the H*-SE measure. There was no statistically significant effect of syllable structure, and no other significant interactions for any of the two delay measures. Finally, there was a main effect of language (L1 background) for each of the two delay measures: H*-SO: $\chi^2 = 14.81$, $p < .001$ H*-SE: $\chi^2 = 5.15$, $p < .05$.

The lack of a significant main effect of proficiency, but its significant interaction with stress position, may be explainable as effects of the L1 background of speakers and/or the stress position of words. We therefore computed confidence intervals using parametric bootstrapping (via the *confint* function in R) to explore the interaction between proficiency and stress and the effect of L1 background. Native speaker group was the reference level of the categorical variable language group (the other groups being determined according to L1 background and proficiency level: JBE, JAE, SBE, SAE). The results for the H*-SO measure of alignment are as follows (as reported earlier H*-SE did not yield statistically significant results in the relevant analysis): in the initial-stress position all the non-native groups were significantly different than the native speakers. In the final-stress condition both L1 Spanish groups realised significantly later peak alignment than the native speakers, whilst the L1 Japanese groups were not significantly different. The results for each measure of alignment are depicted in Figure 4 and Figure 5.

Insert Figures 4-5 about here

The differences between the results obtained for the two delay measures could suggest a possible effect of syllable duration. Thus, as described before, the third measure, SyllNorm, represents alignment as a proportion of syllable duration. For this measure comparisons of the fitted models revealed that there was a main effect of proficiency ($\chi^2 = 16.88$, $p < .01$), a main effect of stress position ($\chi^2 = 43.46$, $p < .001$) and a significant interaction between proficiency and stress position ($\chi^2 = 10.88$, $p < .05$). The results of the confidence intervals were consistent with those obtained for the H*-SO measure: in the final-stress position only the L1 Spanish speakers were significantly different than the native speakers, while in the initial-stress condition all the L1 Japanese and the L1 Spanish groups were different than the native speakers. The mean alignment of each measure as a proportion of syllable duration by language group in the two stress positions is shown Figures 6.

Insert Figure 6 about here

4.0 Discussion

This study examined the effects of L1 background (Japanese and Spanish) and L2 proficiency (basic and advanced levels) in the proportions of contour choice of the high peak target on accented syllables and its alignment in the pitch accent contour (L+)H*[L-L%] in L2 English. The experiment varied test items according to stress position and syllable composition. With regard to the proportion of plateau-type accents, we found that the JAE learners produced far fewer plateaux in their English than their counterparts at the basic level of proficiency, while the Spanish were generally comparable to the native English speakers. Overall, these results suggest that (1) the choice of tonal contour shapes in the L2 may progress toward target-like patterns with increase in proficiency, and (2) the apparent faster acquisition in the Spanish late bilinguals compared to the Japanese may possibly be due to the typological similarity of the source (i.e. Spanish) and target (i.e. English) languages, in support of Hypothesis 1b. Given previous findings of the prevalence of plateau-type accents in the nuclear position of narrow focussed utterances in PRS, we expected that these speakers would produce significantly more of this pitch contour shape compared to the native English speakers (Hypothesis 1a). However, this was not borne out. Perhaps the difference here between the Spanish and Japanese is related to the fact that, unlike the Japanese, the PRS speakers had at their disposal a wider range of pitch contour shapes associated with narrow focus marking in their L1, with the falling pattern (L+H*L-L%) overlapping with the typical English pattern. This was their preferred pattern in our data. It is possible that they were able to exploit this optionality in producing the pitch accent form that most closely matched the canonical pattern of the target language. The prevalence of plateaux in the English of the Japanese speakers, compared to other speakers in our study, may in part be explained as a possible transfer from their native language (Hypothesis 1a). As described before, in final-accented words in Japanese, the fall of the H*+L pitch accent may be suspended, giving way to a high plateau.

With regard to alignment, the results were more variable for the H*-SE alignment measure than for the other two measures, H*SO and SylNorm, which were largely consistent with each other. This difference is clearly related to the individual variation in syllable duration, which appears to affect the H*SE measure than it does the H*SO measure of alignment. Thus, at least for this dataset, of the two delays measures the H*-SO was the more robust measure of tonal alignment, as it appears less sensitive and susceptible to variability in syllable duration than the H*-SE measure. This corroborates previous studies that have reported similar findings (e.g. Xu and Liu 2006, Ladd et al. 2009; and Astruc et al. 2013). The results showed significant differences between all four late bilingual groups and the American English monolingual speakers within the initial-stressed position, and between the Spanish L1 speakers and the native speakers in the final-stressed position.

In the initial-stressed position, the results showed that the off-target patterns were comparable for the L1 Spanish and the L1 Japanese speakers (i.e. all four learner groups produced statistically significantly later alignment than the native speakers). This might be interpreted as an indication that, at least for these speakers, the acquisition of peak timing by late bilinguals may perhaps be subject to a general simplification strategy. This finding of a general simplification strategy would corroborate other studies (e.g. Rasier & Hiligsmann 2007, Li & Post 2014) that have found general (possibly universal) developmental paths in the L2 acquisition of prosody. It must be noted, however, that the general finding of late peak alignment may not hold for early bilinguals, in light of previous studies in contact situations (Elordieta 2003; O'Rourke 2004; Colantoni & Gurlekian 2004; Hualde & Schwegler 2008; inter alia) that have found early peak alignment to be the preferred simplification strategy for these speakers, as described earlier.

An alternative explanation to a general mechanism in the acquisition of L2 tonal alignment could be that the realisation of English tonal alignment by the speakers in this study may be highly influenced by the tonal alignment patterns of their native language/dialect. For the (Tokyo) Japanese speakers, the occurrence of *ososagari* in initial-accented words, as described earlier in section 1.2, may be a contributing factor in the significantly later alignment (relative to the native English speakers) produced by these L1

Japanese speakers. For the L1 (Puerto Rican) Spanish speakers, although they produced the canonical L+H* pitch accent used in marking narrow focus in American English – similarly to the native speakers – their significantly later alignment than the native speakers may equally reflect systematic differences in alignment patterns between their L1 and L2. These two alternative explanations can only be teased apart adequately in a future study that includes various other L1-L2 pairings and different types of speech materials (both narrow and broad focus sentences) in both the source language and the target language of speakers.

Note also that all five speaker groups in this study consistently had earlier alignment in the vicinity of a prosodic boundary (i.e. in final-stressed words) than in initial-stressed words. This is evidently due to tonal repulsion by the upcoming L boundary. However, the finding that the L1 PRS speakers appeared to have aligned their H* peaks significantly later than the native speakers in the final-stressed words is somewhat confounding. On the one hand, it does not appear to be related to the effects of the L1 when taken with the observation of Armstrong (2010) that PRS speakers tend to realise the peak of L+H* ‘early’ in the syllable in narrow focus statements. (It must be noted here though that the author did not provide exact duration values for us to establish how early the peak is aligned relative to a given segmental landmark.) On the other hand, as the Japanese were target-like in their realisation of alignment in final-stressed words, these results would also appear to be less compatible (than the results for initial-stressed words) with our earlier speculation that there may be a general tendency of late bilinguals to realise later alignment in their L2. However, it is arguable that this apparent difference between the L1 Japanese and the L1 PRS speakers might simply be an ‘artefact’ of the study. More precisely, the fact that the Japanese realised significantly more plateaux than the other speaker groups meant that these patterns were excluded from the alignment analyses. Perhaps the prevalence of plateaux in the L1 Japanese groups is indicative of a truncated or neutralised fall being used as an alternative to a late fall, which, on the whole, would make them more comparable to the L1 Spanish speakers than the alignment results suggest. Further, the fact that L1 Japanese speakers were closer to target than L1 Spanish speakers may also be interpreted as an effect of their native H*+L pitch accent, compared to the PRS L+H* typically deployed in marking narrow focus. That is to say, a falling H*+L accent will likely always lead to an earlier alignment than a rising L+H* accent.

On the face of it, the results suggest more clearly that L2 intonational development progresses in a non-uniform way: learners who were producing target-like pitch contour shapes were still off-target in how tones are aligned with the underlying syllable. This finding appears to lend support to our third hypothesis that knowledge of phonetic implementation may not proceed in parallel with phonological knowledge.

There were no differences in H* alignment according to syllable composition. Tonal alignment is very closely linked to syllable structure and segmental composition, and owing to the fact that these aspects vary cross-linguistically, and given language-specific differences in segment articulation patterns, it is surprising that alignment patterns in these environments would yield no evidence of possible L1 effects as we had expected (Hypothesis 2b). It is notable though that, notwithstanding evidence of significant cross-language differences in segment articulation patterns, most of the segmental effects on tonal alignment patterns are related to the mechanics of articulation and show some consistency across languages.

Taken together, our assumption that the rate of progression in proficiency varies according to the typological closeness of the L1 and L2 intonation systems was only partly borne out. In terms of the phonological acquisition of American English pitch accent pattern, the results indicated that late bilinguals of a typologically similar L1 background (Spanish) were more on-target compared to speakers of a less similar language (Japanese) of the same proficiency, thus supporting other studies that have shown that typological closeness between languages can play a role in the rate of acquisition (Lepetit 1989; Altmann 2006; Kijak 2009; *inter alia*). Nonetheless, the observation that the L1 Japanese speakers appeared to progress toward the target as their proficiency increased confirms that, despite typological differences between an L1 and a target L2, learners can over time become more target-like in acquiring aspects of the phonological regularities of their L2 prosodic system.

It would appear that even at a high level of proficiency late bilinguals may have difficulty with implementing some aspects of the phonetic details of their target native language settings, in particular the rules for mapping phonology to phonetics in their L2 intonation. The results also suggest that the phonetic details of the intonational structure of a target language that is typologically similar to the source language of learners may not necessarily be any easier for them to acquire than for learners whose source and target languages are not structurally similar. Perhaps the similarity of pitch contour shapes between source and target languages is likely to have a greater influence on learners' phonetic realisation than typological similarity in the general structure of the intonation systems. It is also worth noting that there is some evidence that the phonetic details of the intonational structure of the dialect of a language can be successfully acquired or 'imitated' by speakers of another dialect of that language (cf. D'Imperio et al. 2014 who investigated tonal alignment in different (Southern) regional varieties of Italian). D'Imperio and her colleagues suggest that speakers may have a higher sensitivity to the phonetic details within dialects of the same language than across languages. It remains to be seen though whether this sensitivity will hold for dialects that are not close in geographical proximity to each other.

Overall, the picture that seems to emerge from this study is that whilst the phonology and the phonetics are closely linked, late bilinguals need to learn the strategies required to map phonological representation to phonetic targets separately.

5.0 Conclusion

This study reports an experiment examining the acquisition of pitch accent contour and the alignment of tonal targets in narrow focus marking by non-native speakers of English. With regard to the choice of pitch accent contour, the L1 Spanish speakers produced more target-like patterns than the Japanese compared to the native speakers, although the Japanese appeared to have developed more toward the target as proficiency increased. Compared to the native speakers the late bilinguals were generally off-target in their alignment realisation regardless of their L1 background or proficiency level in English, with differences according to L1 and stress position. We speculate that this may be due to the fact that alignment requires a delicate simultaneous control of pitch and duration with articulation. We further speculate that this finding may represent a general cross-linguistic trend in the acquisition of intonation, which would suggest that L2 learners are likely to be slow or unable to acquire the more highly sensitive language-specific regularities of peak alignment. More precisely, tonal alignment is sensitive to underlying phonological structure, which bilinguals will need to acquire in order to produce target-like implementation patterns.

Acknowledgements

This research was supported by a PhD scholarship awarded to the first author from the Cambridge Commonwealth Trust and St John's College, University of Cambridge, and by a PhD research support grant from the Japan Foundation. This support is gratefully acknowledged. We are also very grateful to Mariapaola D'Imperio, Martine Grice, Ineke Mennen, Francis Nolan, Mits Ota, and the anonymous reviewers for their constructive feedback.

References

- Altmann, H. (2006). The perception and production of second language stress: A cross-linguistic experimental study. PhD dissertation, University of Delaware, Newark.
- Alvord, S. (2010a). Miami Cuban Spanish Declarative Intonation. *Studies in Hispanic and Lusophone Linguistics*, 3:1, 3-39.
- Aoyama, K., & Guion, S. (2007). Acoustic analyses of duration and F0 range. In O. Bohn & M. Munro (eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege*, pp. 281-297. Amsterdam: John Benjamins.
- Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second Language Research*, 10, 215-240.
- Armstrong, M. (2010). Puerto Rican Spanish Intonation. In P. Prieto & P. Roseano (eds.), *Transcription of intonation of the Spanish language*, pp. 155-189. Munich: LINCOM Europa.

- Arvaniti, A., & Ladd, D. (1995). Tonal alignment and the representation of accentual targets. In *Proceedings of the 13th ICPHS*. Stockholm, 4, 220-223.
- Arvaniti, A., Ladd, D., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, 26(1), 3-25.
- Arvaniti, A. & Baltazani, M. (2005). Intonational analysis and prosodic annotation of Greek spoken corpora. In S-A. Jun (eds.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 84-117). Oxford: Oxford University Press.
- Arvaniti, A., Ladd, D., & Mennen, I. (2006). Phonetic effects of focus and ‘tonal crowding’ in intonation: Evidence from Greek polar questions. *Speech Communication*, 48, 667-696.
- Arvaniti, A., & Garding, G. (2007). Dialectal variation in the rising accents of American English. In J. Cole & J. Hualde (Eds.), *Papers in Laboratory Phonology*, 9, (pp.547-576). Berlin, New York: Mouton de Gruyter.
- Astruc, L., Payne, E., Post, B., Vanrell, M., & Prieto, P. (2013). Tonal targets in early child Catalan, Spanish, and English. *Language and Speech*, 56(2), 229-253.
- Atterer, M., & Ladd, D. (2004). On the phonetics and phonology of “segmental anchoring” of F0: evidence from German. *Journal of Phonetics*, 32, 177-197.
- Backman, N. (1979). Intonation errors in second language pronunciation of eight Spanish-speaking adults learning English. *Interlanguage Studies Bulletin*, 4 (2), 239–266.
- Barry, W. (2007). 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.
- Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, III, 255-309.
- Beckman, M., & Hirschberg, J. (1992). The ToBI annotation conventions. Ms. The Ohio State University.
- Beckman, M., & Ayers-Elam, G. (1997). *Guidelines for ToBI labeling* (version 3.0, March 1997). MS. The Ohio State University Research Foundation.
- Beckman, M., Campos, M., McGory, J., & Morgan, T. (2002). Intonation across Spanish, in the Tones and Break Indices framework. *Probus*, 14, 9-36.
- Beckman, M., Hirschberg, J., & Shattuck-Hufnagel, S. (2005). The original ToBI system and the evolution of the ToBI framework. In S.-A. Jun (Ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 9-54). Oxford: Oxford University Press.
- Beckman, M., & Venditti, J. (2010). Tone and Intonation. In W. Hardcastle & J. Laver (Eds.), *The Handbook of Phonetic Sciences* (pp. 603-652). Oxford: Blackwell.
- Beckman, M., & Venditti, J. (2011). Intonation. In J. Goldsmith, J. Riggle, & A. Yu (Eds.), *The Handbook of Phonological Theory*, (pp. 485-532). Oxford: Blackwell.
- Benet, A., Kireva, E., Pešková, A., & Gabriel, C., (2001). 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* (pp. 22-25). Cáceres, Extremadura.
- Boersma, P., & Weenink, D. (2014). *Praat: doing phonetics by computer*, version 5.3.35, retrieved from <http://www.praat.org/>.
- Bolinger, D (1985). Two Views of Accent. *Journal of Linguistics*, 21, 79-123.
- Bruce, G. (1977). Swedish word accents in sentence perspective. Lund: Gleerup.
- Busà, M., & Urbani, M. (2011). A cross-linguistic analysis of pitch range in English L1 and L2. In *Proceedings of the 17th ICPHS* (pp. 380-383). Hong Kong.
- Caspers, J., & van Heuven, V. (1993). Effects of time pressure on the phonetic realization of the Dutch Accent-lending pitch rise and fall. *Phonetica*, 50, 161–171.
- Chen, A., & Mennen, I. (2008). Encoding interrogativity intonationally in a second language. In Barbosa, P. Madureira, S. & Reis, C. (Eds.), In *Proceedings of the 4th International Conferences on Speech Prosody* (pp. 513-516). Campinas.
- Colantoni, L., & Gurlekian, J. (2004). Convergence and intonation: historical evidence from Buenos Aires Spanish. *Bilingualism: Language and Cognition*, 7, 107-119.
- Colantoni, L. (2011). Broad-focus declaratives in Argentine Spanish contact and non-contact varieties. In C. Gabriel & C. Lleó (eds.). *Intonational phrasing at the interfaces: cross-linguistic and bilingual studies in Romance and Germanic*. Amsterdam: John Benjamins. 183-212.
- Dilley, L., & Brown, M. (2007). Effects of pitch range variation on F0 extrema in an imitation task. *Journal of Phonetics*, 35, 523-551.
- D'Imperio, M. (2002). Italian Intonation: An Overview and some Questions. *Probus* (special volume on intonation in Romance languages), 14 (1), 37-69.

- D'Imperio, M. & House, D. (1997). Perception of questions and statements in Neapolitan Italian. In G. Kokkinakis, N. Fakotakis, & E. Dermatas (Eds.), *In Proceedings of the 5th Eurospeech*, Vol. 1 (pp. 251-254). Rhodes.
- D'Imperio, M. (2006). "Preface to Current Issues in Tonal Alignment". *Italian Journal of Linguistics/Rivista di Linguistica*, vol. 18, no. 1. 2006, p. 3-18.
- D'Imperio, M., Bertrand, R., Di Cristo, A. & Portes, C. (2007a). Investigating phrasing levels in French: Is there a difference between nuclear and prenuclear accents? In J. Camacho, V. Deprez, N. Flores, L. Sanchez (Eds.), *Selected Papers from the 36th Linguistic Symposium on Romance Languages (LSRL)*, 97-110. New Brunswick: John Benjamins Publishing Company.
- D'Imperio, M., Petrone, C., & Nguyen, N. (2007b). Effects of tonal alignment on lexical identification in Italian. In C. Gussenhoven & T. Riad (Eds.), *Tones and Tunes*, 2, 79-106. The Hague: Mouton de Gruyter.
- D'Imperio, M. (2011). Prosodic representations. In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Handbook of Laboratory Phonology*. Oxford: Oxford University Press (section on tonal alignment).
- D'Imperio, M., Cavone, R. and Petrone, C. (2014). Phonetic and phonological imitation of intonation in two varieties of Italian. *Frontiers in Psychology* 5: 1226. DOI: <https://doi.org/10.3389/fpsyg.2014.01226>
- Elordieta, G. (2003). "The Spanish intonation of speakers of a Basque pitch-accent dialect". *Catalan Journal of Linguistics* 2, 67-95.
- Elordieta, G. (2006). "Spanish pitch accent alignment by Northern Bizkaian Basque speakers". In B. Fernandez & I. Laka (eds.), *Andolin Gogoan. Essays in Honour of Profesor Eguzkitza*, 269-290. Bilbao: University of the Basque Country.
- Elordieta, G.; Calleja, N. (2005). "Microvariation in accentual tonal alignment in Basque Spanish". *Language and Speech* 48, 397-439.
- Face, T. (2002). Intonational marking of contrastive focus in Madrid Spanish. Munich: LINCOM Europa.
- Face, T. & Prieto, P. (2007). Rising accents in Castilian Spanish: a revision of Sp_ToBI. *Journal of Portuguese Linguistics*, 6.1, 117-146.
- Goldsmith, J. (1976). Autosegmental phonology. PhD dissertation. MIT. [Published 1979, New York: Garland Publishing.]
- Graham, C. (2014). The phonetics & phonology of late bilingual prosodic acquisition: a cross-linguistic investigation. PhD dissertation, University of Cambridge.
- Gussenhoven, C. (1984). *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.
- Gussenhoven, C. (2004). *The phonology of Tone & Intonation*. Cambridge: Cambridge University Press.
- Gut, U. (2009). Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German. Oxford: Peter Lang.
- Hualde, J., & Schwegler, A. (2008). Intonation in Palenquero. *Journal of Pidgin and Creole Languages*, 23: 1-31. DOI: 10.1075/jpcl.23.1.02hua
- Hyman, L. (2001a). Privative tone in Bantu. In S. Kaji (Ed.), *Cross-linguistic studies of tonal phenomena* (pp.237-257). Tokyo: Institute for the Study of Languages and Cultures.
- Hyman, L. (2006). Word prosodic typology. *Phonology*, 23, 225-257.
- Hyman, L. (2011). *In Defense of Prosodic Typology: A Response to Beckman & Venditti*. Berkeley Phonology Lab Annual Report, 220-235.
- Ishihara, T. (2003). A phonological effect on tonal alignment in Tokyo Japanese. In *Proceedings of the 15th ICPHS* (pp. 615-619). Barcelona.
- Jenner, B. (1976). Interlanguage and foreign accent. *Interlanguage Studies Bulletin*, 1, 166-195.
- Jilka, M., & Möbius, B. (2006). Towards a comprehensive investigation of factors relevant to peak alignment using a unit selection corpus. *Proceedings of Interspeech*, (pp. 2054-2057). Pittsburgh, Pennsylvania.
- Jun, S-A. (1993). The phonetics and phonology of Korean prosody. PhD dissertation, Ohio State University. [Published in 1996 by Garland, New York].
- Jun, S-A., & Fougeron, C. (1995). The Accentual Phrase and the prosodic structure of French. In K. Elenius & P. Branderud (Eds.), *Proceedings of the 13th ICPHS* (pp. 722-725). Stockholm.
- Kiesling, S., Dille, L., & Raymond, W. (2006). The Variation in Conversation (ViC) Project: Creating of the Buckeye Corpus of conversational speech. Ms. Columbus, OH: Department of Psychology, Ohio State University, available at www.buckeyecorpus.osu.edu.

- Kijak, A. (2009). How stressful is L2 stress? *A cross-linguistic study of L2 perception and production of metrical systems*. Utrecht: LOT (Netherlands Graduate School of Linguistics).
- Ladd, D. (1996, 2nd ed. 2008). *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. (2001). Intonation. In M. Haspelmath, E. König, W. Oesterreicher & W. Raible (Eds.), *Language Typology and Language Universals* (pp. 1380-90). Berlin: de Gruyter.
- Ladd, D. (2003). Phonological conditioning of f0 target alignment. In *Proceedings of the 15th ICPHS* (pp. 249-252). Barcelona.
- Ladd, D., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant 'segmental' anchoring of f0 movements under changes in speech rate. *Journal of the Acoustical Society of America*, 106, 1543-1544.
- Ladd, D., Mennen, I., & Schepman, A. (2000). Phonological conditioning of peak alignment of rising pitch accents in Dutch. *Journal of the Acoustical Society of America*, 107, 2685-2696.
- Ladd, D., Schepman, A., White, L., Quarmby, L., & Stackhouse, R. (2009). Structural and dialectal effects on pitch peak alignment in two varieties of British English. *Journal of Phonetics*, 37, 145-161.
- Lepetit, D. (1989). Cross-linguistic influence in intonation: French/Japanese and French/English. *Language Learning*, 39(3), 397-413.
- Li, A., & Post, B. (2014). L2 Acquisition of Prosodic Properties of Speech Rhythm: Evidence from L1 Mandarin and German Learners of English. *Studies in Second Language Acquisition* 36(2), 223 – 255.
- Liberman, M. (1975). The intonational system of English. PhD dissertation, MIT. [Published 1979, New York, NY: Garland Publishing.]
- Maekawa, K. (1994b). Is there 'dephrasing' of the accentual phrase in Japanese? In J. Venditti, *Ohio State University Working Papers in Linguistics*, 44, 146-165.
- Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics*, 32, 543-563.
- Mennen, I. (2015). Beyond segments: Towards a L2 intonation learning theory (LILt). In E. Delais-Roussarie, M. Avanzi & S. Herment (Eds.), *Prosody and language in contact: L2 acquisition, attrition and languages in multilingual situations*, (pp. 171-188). Berlin: Springer.
- Nibert, H. (2000). Phonetic and phonological evidence for intermediate phrasing in Spanish. Ph.D. Thesis, University of Illinois, Urbana- Champaign.
- Niebuhr, O. (2007). The signalling of German rising-falling intonation categories: The interplay of synchronization, shape, and height. *Phonetica*, 64, 174–193.
- Niebuhr, O., D'Imperio, M., Gili Fivela, B., & Cangemi, F. (2011). "Are there "shapers" and "aligners"? Individual differences in signalling pitch accent category". *Proceedings of 17th International Congress of Phonetic Sciences*, August 2011, Hong Kong. 120-123.
- O'Rourke, Erin. 2004. Peak placement in Peruvian Spanish. In Auger, J., Clements, C., & Vance, B. (eds.). *Contemporary Approaches to Romance Linguistics: Selected papers from the 33rd Linguistic Symposium on Romance Languages (LSRL)*, Bloomington, Indiana, April 2003. John Benjamins: Amsterdam. 321-341.
- Payne, E., Post, B., Prieto, P., Vanrell, M., & Astruc, L. (2012). Measuring child rhythm. *Language and Speech*, 55, 202–228.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD dissertation, MIT. Distributed 1988, Indiana University Linguistics Club.
- Pierrehumbert, J., & Beckman, M. (1988). *Japanese Tone Structure*. Linguistic Inquiry monographs (No. 15). Cambridge: MIT Press.
- Pierrehumbert, J., & Steele, S. (1989). Categories of tonal alignment in English. *Phonetica*, 46, 181-196.
- Pitrelli, J., Beckman, M., & Hirschberg, J. (1994). Evaluation of prosodic transcription labeling reliability in the ToBI framework. In *Proceedings of the 3rd International Conference on Spoken Language Processing* (pp. 123-126). Yokohama.
- Post, B. (2000). *Tonal and phrasal structures in French intonation*. The Hague: Holland Academic Graphics.
- Post, B. (2011). The multi-faceted relation between phrasing and intonation in French. In Lleo, C. & Gabriel, C. (Eds.), *Hamburger Studies in Multilingualism 10*, pp. 44-74. Amsterdam: John Benjamins.
- Prieto, P., van Santen, J., and Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, 23, 473:500.

- Prieto, P., D'Imperio, M. & Gili Fivela, B. (2005). "Pitch accent alignment in Romance: primary and secondary associations with metrical structure". *Language and Speech*, 48, 4, p. 359-396. (Special issue on Variation in Intonation, ed. by Paul Warren).
- Prieto, P. (2011). Tonal alignment. In M. Van Oostendorp, C. Ewen, B. Hume. & K. Rice (Eds.), *Companion to Phonology* (pp.1185-1203). Chichester: Wiley-Blackwell.
- Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française*, 28, 41-66.
- Rietveld, T., & Gussenhoven, C. (1995). Aligning pitch target in speech synthesis: effect of syllable structure. *Journal of Phonetics*, 23, 375-385.
- Savino, M., & Grice, M. (2007). The role of pitch range in realising pragmatic contrasts - The case of two question types in Italian. In *Proceeding of the 16th ICPhS*. (pp. 1037-1040). Saarbrücken.
- Schepman, A., Lickley, R., & Ladd, D. R. (2006). Effects of vowel length and "right context" on the alignment of Dutch nuclear accents. *Journal of Phonetics*, 34, 1-28.
- Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. In J. Kingston & M. Beckman (Eds.), *Papers in Laboratory Phonology I* (pp. 72-106). Cambridge: Cambridge University Press.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. In *Proceedings of the 2nd International Conference on Spoken Language Processing* (pp. 867-870). Alberta,
- Steele, S. (1986). Nuclear accent f0 peak location: Effect of rate vowel, and number of syllables. *Journal of the Acoustical Society of America*, 80, 1, 51.
- Sugito, M. (1982). *Nihongo Akusento no Kenkyuu* [Studies on Japanese Accent]. Tokyo: Sansendo.
- Sugiyama, Y. (2012). *The Production and Perception of Japanese Pitch Accent*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Ullakonoja, R. (2007). Comparison of Pitch Range in Finnish (L1) Fluency and Russian (L2). In *Proceeding of the 16th ICPhS*. (pp. 1701-1704). Saarbrücken.
- University of Michigan Library (n.d.). Online Exhibits: UM Library celebrates language. Retrieved from <http://www.lib.umich.edu/online-exhibits/exhibits/show/language/english>.
- Vance, T. J. (1995). Final accent vs. no accent: Utterance-final neutralization in Tokyo Japanese. *Journal of Phonetics* 23(4): 487-499.
- van Santen, J., & Hirschberg, J. (1994). Segmental effects on timing and height of pitch contours. In *Proceedings of the International Conference on Spoken Language Processing*, (vol. 1). (pp. 719 - 722). Yokohama, Japan.
- Vanrell, M. (2006). *The phonological role of tonal scaling in Majorcan Catalan interrogatives*, MA thesis. Universitat Autònoma de Barcelona, Barcelona.
- Venditti, J. (1995). Japanese ToBI Labelling Guidelines. *Ohio State University Working Papers in Linguistics* 50, 127-62.
- Venditti, J. (2005). The J ToBI model of Japanese intonation. In Jun (Ed.), *Prosodic Typology: The Phonology of Intonation and Phrasing* (pp. 172-200). Oxford: Oxford University Press.
- Welby, P. (2003). The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation. PhD dissertation. Ohio State University.
- Willems, N. (1982). *English intonation from a Dutch point of view*. Dordrecht: Foris Publications.
- Wilson, M. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2, Behavioural Research Methods, Instruments and Computers.
- Xu, Y. (1998). Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica*, 55, 179-203.
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics*, 27, 55-105.
- Xu, Y., & Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, 18, 125-159.

Appendix A Biographical information for late bilinguals in the study

Participants			Age at test	Age at onset of acquisition	Proficiency (assessor judgement)		Other languages spoken (proficiency) Self-reported
L1 background	Code	Gender			CEFR level	Oral proficiency	
Japanese	S1	F	23	12	C1	Advanced	None
	S2	F	26	8	C1	Advanced	None
	S3	F	23	11	C1	Advanced	Korean (basic)
	S4	M	19	5	C1	Advanced	None
	S5	M	24	9	C1	Advanced	French (basic)
	S6	F	23	10	A2	Basic	None
	S7	F	22	9	A2	Basic	None
	S8	F	19	8	A2	Basic	Italian (basic)
	S9	M	21	7	A2	Basic	None
	S10	M	20	9	A2	Basic	None
Spanish	S11	F	22	6	A2	Basic	None
	S12	F	21	13	A2	Basic	None
	S13	F	24	9	A2	Basic	None
	S14	M	23	5	A2	Basic	None
	S15	M	25	8	A2	Basic	None
	S16	F	22	6	C1	Advanced	None
	S17	F	21	7	C1	Advanced	None
	S18	F	24	8	C1	Advanced	French (basic)
	S19	M	22	7	C1	Advanced	None
	S20	M	25	9	C1	Advanced	None

Appendix B Test items

Sample dialogues		
A: What animal is Melanie looking for? [Picture of a monkey is shown]		
B: She's looking for a monkey .		
A: Who's/What's this? [Picture shown]		
B: This is mommy . This is a rat .		
A: Do you know this word?		
B: Yes. It's the word permit .		
Other test items (n – noun; v – verb).		
Stress position	Obstruent-onset	Sonorant-onset
Initial	<u>b</u> aby, <u>d</u> addy, <u>b</u> utterfly, <u>h</u> ippo, <u>h</u> elicopter, <u>p</u> ermit (n), <u>p</u> otty, <u>a</u> ngel	<u>m</u> usic, <u>r</u> ecord (n), <u>l</u> ion, <u>m</u> oney, <u>m</u> ommy, <u>M</u> elanie, <u>m</u> onkey, <u>m</u> ermaid
Final	<u>b</u> egin, <u>o</u> bject (v), <u>b</u> ee, <u>b</u> oot, <u>g</u> uitar, <u>s</u> un, <u>b</u> ear, <u>h</u> otel	<u>g</u> iraffe, <u>p</u> ermit (v), <u>k</u> angaroo, <u>J</u> apanese, <u>b</u> alloon, <u>m</u> oon, <u>v</u> iolin, <u>t</u> oday

Figures

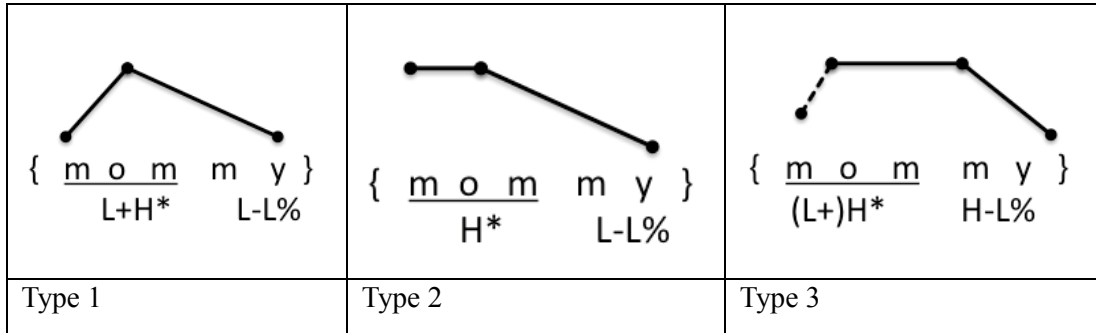
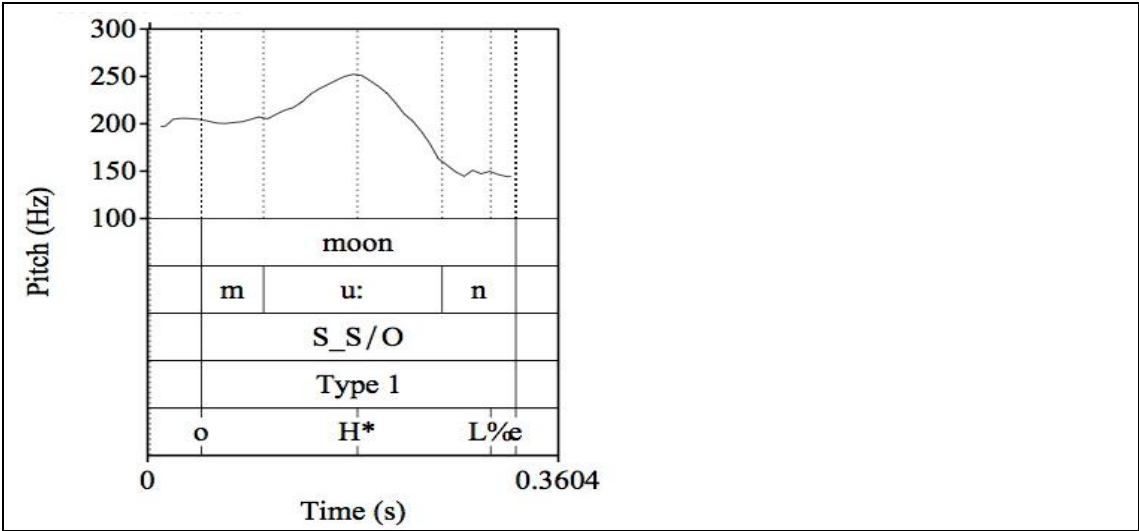
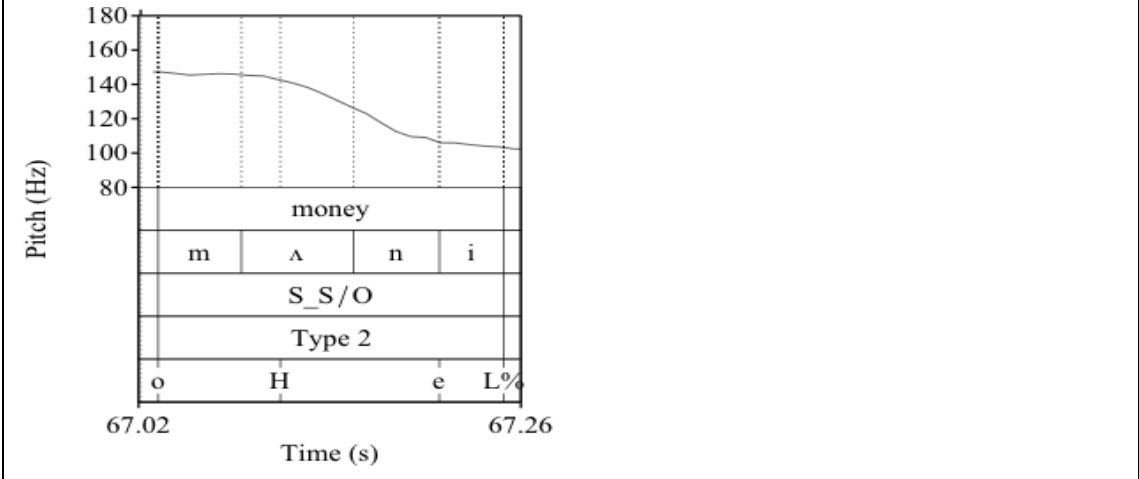


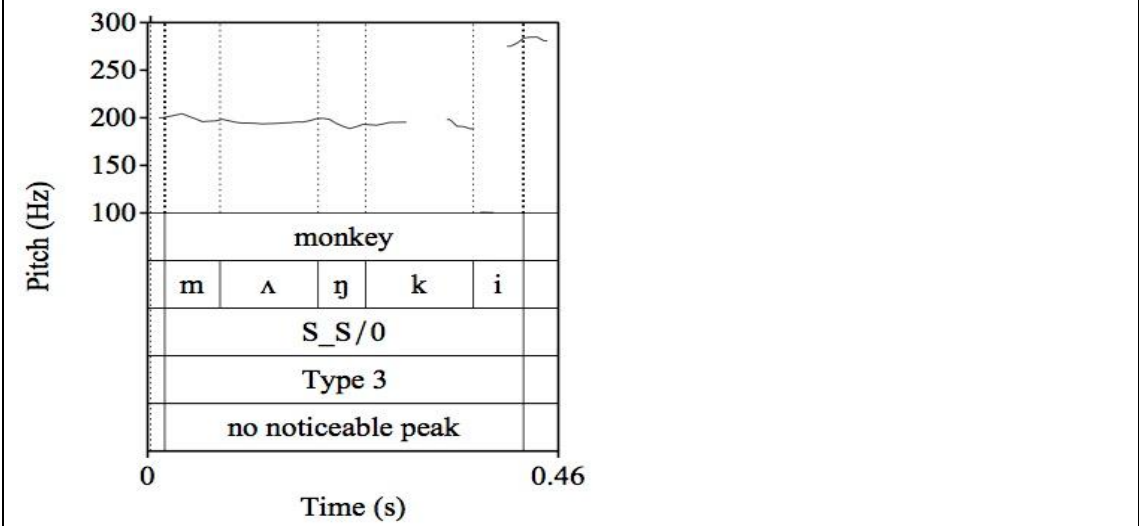
Figure 1 Stylised contours of the three accent types. Only types 1 and 2 were analysed in this study.



Type 1 accent produced by a female native speaker.



Type 2 accent produced by a male Japanese speaker (advanced English)



Type 3 accent produced by a female Japanese speaker (basic English)

Figure 2 Pitch traces of the three pitch accent shapes. 'O' marks the onset of the syllable, 'E' the end of the syllable, and 'H' the H* peak.

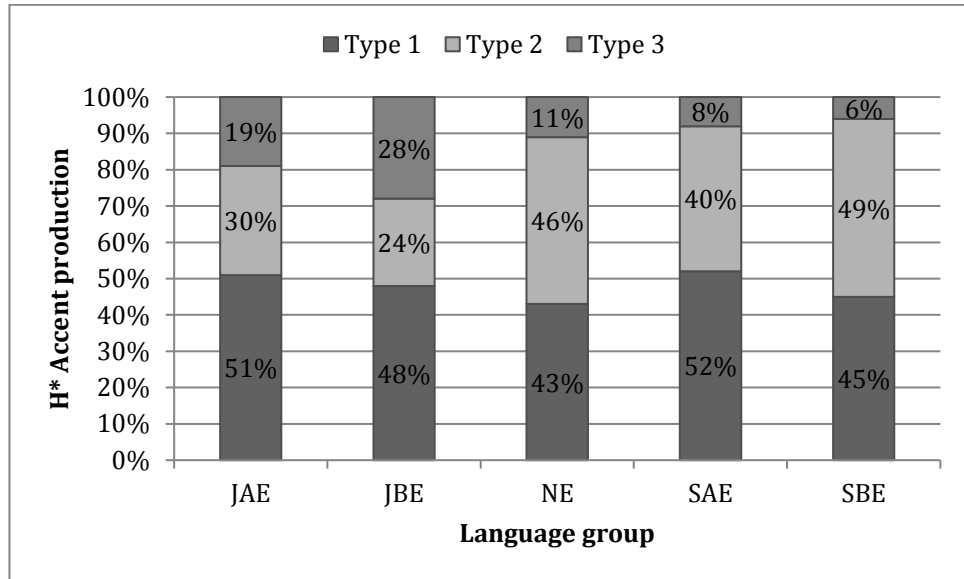


Figure 3 H* pitch accent shapes produced by the five language groups.

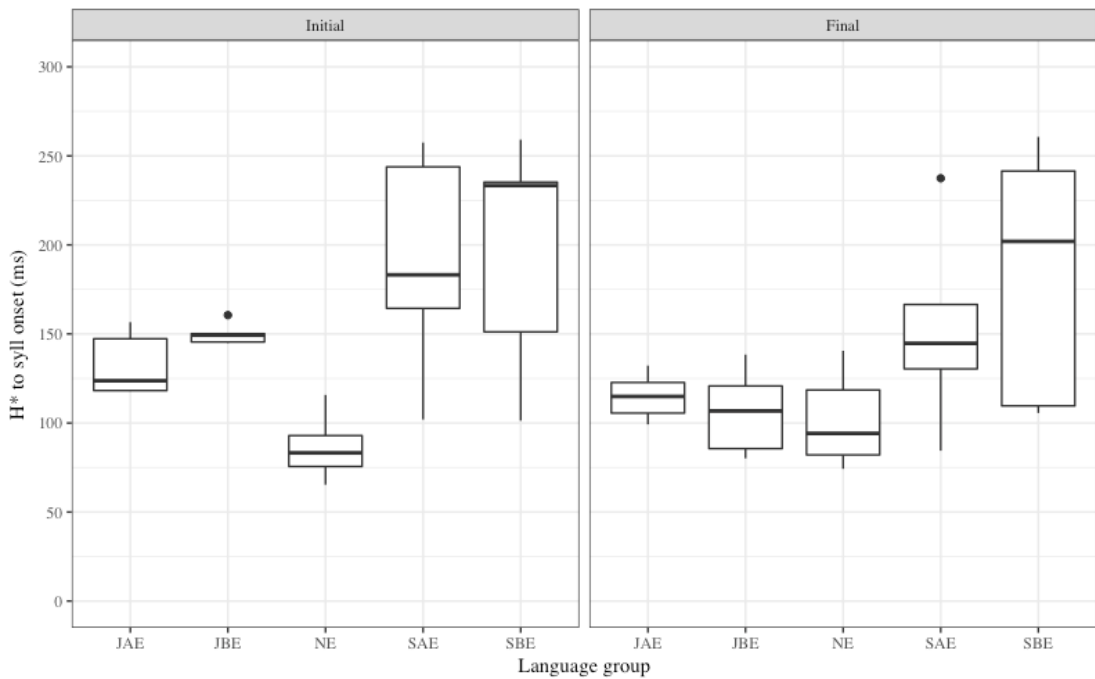


Figure 4 Boxplots for mean time (milliseconds) from H* peak to the syllable onset in each stress position for the five language groups. 95% confidence intervals used for comparing medians.

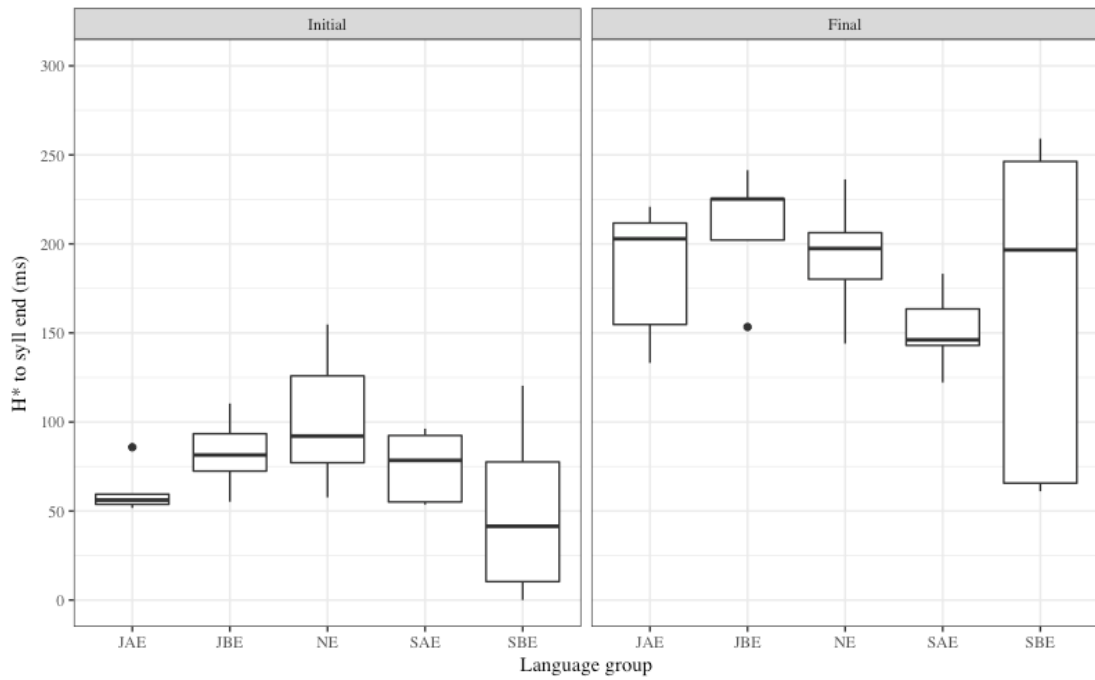


Figure 5 Boxplots for mean time (milliseconds) from H* peak to the syllable end in each stress position for the five language groups. 95% confidence intervals used for comparing medians.

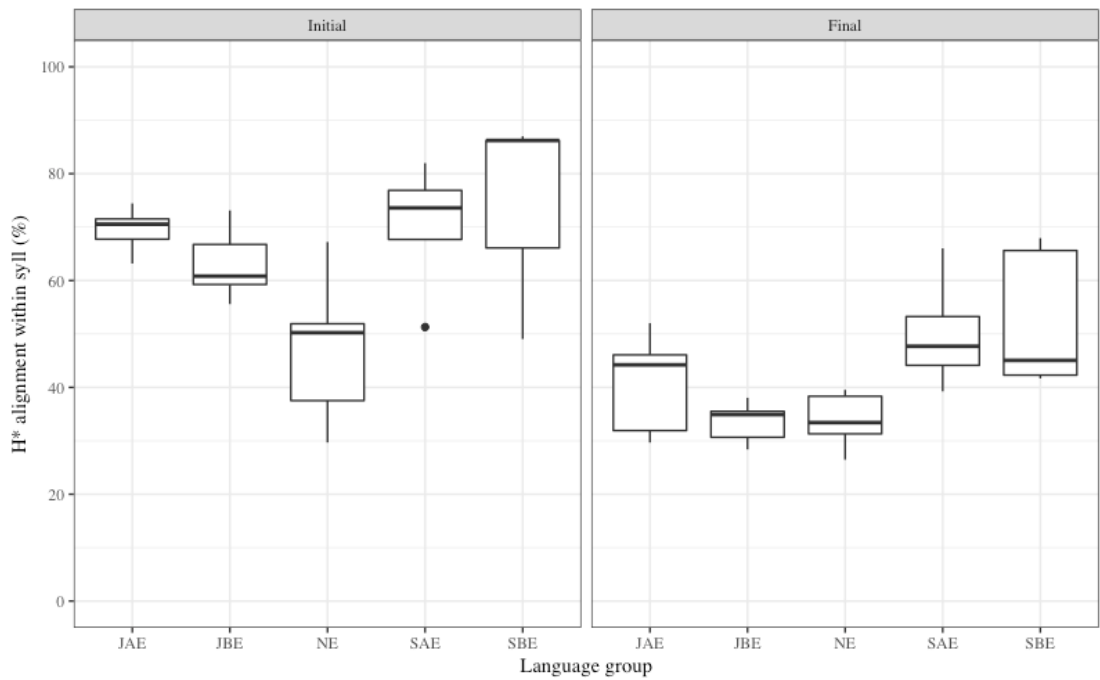


Figure 6 Boxplots for proportional H* alignment as a percentage of the total syllable duration by stress position for the five language groups. 95% confidence intervals used for comparing medians.